



HAL
open science

Des humains derrière l'intelligence artificielle. La sous-traitance du travail de la donnée entre la France et Madagascar

Clément Le Ludec

► To cite this version:

Clément Le Ludec. Des humains derrière l'intelligence artificielle. La sous-traitance du travail de la donnée entre la France et Madagascar. Sociologie. Institut Polytechnique de Paris, 2024. Français. NNT : 2024IPPAT001 . tel-04605206v2

HAL Id: tel-04605206

<https://hal.science/tel-04605206v2>

Submitted on 10 Jun 2024 (v2), last revised 25 Jun 2024 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



INSTITUT
POLYTECHNIQUE
DE PARIS

NNT : 2024IPPAT001

Thèse de doctorat



Des humains derrière l'intelligence artificielle. La sous-traitance du travail de la donnée entre la France et Madagascar

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à Télécom Paris

École doctorale n°626 de l'Institut Polytechnique de Paris (ED IPP)
Spécialité de doctorat : Sociologie

Thèse présentée et soutenue à Palaiseau, le 5 mars 2024, par

CLÉMENT LE LUDEC

Composition du Jury :

Isabelle Berrebi-Hoffman Directrice de recherche - Conservatoire national des arts et métiers	Présidente / rapporteure
Caroline Datchary Professeure des universités - Université Toulouse - Jean Jaurès	Rapporteure
Francesca Musiani Chargée de recherche CNRS / CIS	Examinatrice
Luke Stark Assistant professor - University of Western Ontario	Examineur
Julia Velkovska Chargée de recherche - Laboratoire SENSE, Orange Labs	Examinatrice
Paola Tubaro Directrice de recherche - ENSAE	Co-directrice de thèse

Remerciements

Il y a cinq ans, Paola Tubaro et Antonio Casilli m'ont embarqué avec eux dans cet ambitieux projet. Je les remercie vivement de m'avoir fait confiance et m'avoir accompagné durant le début de mon parcours de chercheur. Je tiens également à remercier les membres du jury qui ont bien voulu lire cette thèse et la commenter, Isabelle Berrebi-Hoffman, Caroline Datchary, Julia Velkovska, Luke Stark et Francesca Musiani.

Rien n'aurait été possible sans le travail administratif de l'ensemble du personnel administratif du Département SES de Télécom Paris. Je pense notamment à Karine et à Marie-Josée qui ont toujours été réactives et à l'écoute dans la préparation du voyage à Madagascar. Je pense également à Sylvie qui m'a permis d'avoir des bulles de sérénité tout au long de cette thèse.

J'ai également une pensée particulière pour les représentants CFDT de Télécom Paris qui ont su écouter et porter ma défense des conditions de travail doctorants. Au département SES, j'ai également pu compter sur le soutien attentif de Jean-Samuel Beuscart. Outre nos discussions, je le remercie d'avoir pris du temps pour créer du lien. Je remercie également Pierre-Yves Baudot d'avoir accepté d'être membre de mon comité de suivi de thèse. et d'avoir pris du temps lorsque cela était nécessaire.

Je n'oublie pas les étudiants que j'ai eus l'occasion de côtoyer durant ces quatre années : à l'UCO, à Télécom Paris ou à Sciences Po, ces enseignements ont été des respirations et des inspirations qui m'ont permis d'avancer. Je remercie également l'ensemble de l'équipe de recherche DIPLab, ainsi que tous les collègues doctorantes et doctorants du Discord "*Coordination des précaires de l'ESR*".

Merci à également à l'ensemble de mes collègues doctorantes et doctorants : Assia, Juana et Victor. Vous avez été de précieux collègues et je vous remercie pour votre écoute. Je suis très heureux d'avoir pu grandir humainement et académiquement avec vous. J'espère qu'il y aura encore plein de belles occasions de travailler ensemble. Un remerciement tout particulier à Arnaud et Mandie pour votre contribution à ce travail de recherche.

Je remercie également Maxime pour cette belle collaboration. Rien n'aurait été possible sans cette union des forces. Malgré l'intensité du terrain, il nous restera le souvenir de l'île Sainte-Marie, et je l'espère, encore de beaux articles à écrire ensemble.

Il me faut également évoquer ceux dont les voix constituent la colonne vertébrale de cette thèse. Je remercie également Julie pour son immense contribution. J'ai été ravi de collaborer autour de quelques bières pour échanger sur l'IA, le travail des données et Madagascar. J'ai également une pensée particulière pour Miary qui a été à l'origine de nombreuses entrées de terrain à Madagascar. Je n'aurais jamais pensé que son pays

nous réunirait lorsque nous étions encore étudiants à Lyon. Il a toujours été là, pour aider, pour donner son sentiment, pour traduire la langue et le contexte de son pays. Merci à Éric d'avoir pris le temps de répondre depuis 5 ans à mes nombreuses sollicitations. Avec Andry est votre contribution est considérable, car, sans vous, rien de tout cela n'aurait été possible. Je pense également à Renée, qui a été un point d'ancrage pour comprendre Madagascar et que je remercie pour sa bonne humeur et pour sa générosité.

Andry, mon premier souvenir de notre rencontre est la machette que tu cachais sous le siège de ta voiture, lors de notre premier départ pour une journée d'entretien. J'avoue avoir eu peur ce qui m'attendait à ce moment-là. Ta tâche n'était pas évidente : accompagner et aider deux Vazahas pour une enquête de sociologie. Tu as été chauffeur, traducteur, juriste, un peu sociologue et surtout au moment de repartir pour l'aéroport après une dernière soirée arrosée : un ami. Il me faut également remercier l'ensemble des travailleurs qui ont participé à cette enquête, car il faut du courage pour raconter ces morceaux de vie dans une langue qui n'est pas la votre.

Rien n'aurait été possible sans une équipe de relecteurs et de discutants. Leila, la meilleure des sociologues, merci pour nos cafés sur le canal de l'Ourcq. Ruben, un merci tout particulier, car ces dernières années, tu as été un mentor et surtout un ami. Marc-Alban, Benoit, Léonard, Matthias vous avez eu le courage de lire ces mots encore hachés et ces fautes d'orthographe avec une grande patience. Merci. Vous avez surtout été des amis sur qui j'ai pu compter durant ces longues années pendant lesquelles les sorties se sont parfois faites rares. Merci également à l'ensemble de mes amis qui ont trop souvent subi des discussions pas nécessairement passionnantes sur l'IA ou sur Madagascar : Pware, Annabelle, Céline, V, Marion, Bastien et tous les autres.

Aurore Le Ludec, ma soeur, tu as été là, comme toujours. Là, pour m'aider à rendre mes schémas potables de par tes talents d'artistes. A l'écoute surtout pour me redonner parfois la voie de l'équilibre. Très chers parents, vous m'avez toujours donné la confiance et le choix de faire ce que je croyais juste et bien pour moi. Je remercie Sylvie Le Ludec pour m'avoir appris à débattre, à discuter, mais aussi à être tenace et à ne jamais lâcher. Jean-François Le Ludec, je te remercie pour m'avoir donné le goût de développer des intérêts spécifiques et de les vivre avec passion. Vous m'avez donné la possibilité d'aller jusque-là.

Philippine, nous avons commencé à être ensemble peu avant cette thèse et rien ne pourrait décrire le bonheur que j'ai d'être avec toi. Au quotidien, ta présence lumineuse, ta drôlerie et tes précieux conseils ont été des "amers" au milieu de l'océan. Des plateformes à Madagascar, tu as été là, pour me relire, pour m'aider à faire le tri dans mes idées et pour me donner de la force.

Table des matières

Remerciements	1
Introduction générale	9
Partie 1 : Étudier le processus de production des systèmes d'IA pour en étudier les conséquences sur le travail	22
Chapitre 1 - Les deux "ordres" de la sociologie des algorithmes : l'IA et ses effets et l'IA comme expression de ceux qui la fabrique	22
1.1 Une critique focalisée sur la manière dont les systèmes d'IA sont mobilisés dans la société . . .	26
1.1.1 Les algorithmes au travail et leurs conséquences sur leurs utilisateurs	26
1.1.2 L'opacité des systèmes d'IA comme défi pour les sciences sociales	30
1.2 Une critique qui consiste à examiner les biais de l'IA comme expression des représentations de ceux qui la fabrique	33
1.2.1 La non-objectivité des systèmes algorithmiques et ses conséquences sur les utilisateurs	34
1.2.2 Les négociations des acteurs à l'œuvre derrière la fabrique des dispositifs de quantification	37
1.2.3 Algorithmes et travail numérique. Comment les nouveaux moyens de production numérique créent-ils de nouvelles tâches?	42
Chapitre 2 - Le "digital labor" comme cadre général pour comprendre l'externalisation du travail des données	46
2.1 La littérature sur les "travailleurs du clic" se focalise prioritairement sur deux modes d'organisation du travail : le travail gratuit et le travail des plateformes	46
2.1.1 Produire des données grâce à des consommateurs-travailleurs peu ou pas rémunérés .	46
2.1.2 La littérature sur le micro-travail se concentre sur les plateformes numériques de travail, principalement anglophones	51
2.2 De l'étude de l'organisation du travail de plateformes au tournant global des études sur le travail des données	56
2.2.1 Qualifier et quantifier la foule de micro-travailleurs	56
2.2.2 Les conditions de travail et l'organisation du travail sur les plateformes de micro-travail	59
2.2.3 Questionner le micro-travail comme vecteur de développement pour les pays du Sud global	67

Chapitre 3 - Les coulisses de l'accès à un terrain <i>multi-situé</i>	76
3.3.1 Un premier terrain "français" pour comprendre les pratiques d'externalisation du travail de l'IA	77
3.3.1 Un travail exploratoire qui a permis de cadrer notre objet en écartant certaines pistes présentes dans la littérature	78
3.1.2 Une collecte de données sur la chaîne de valeur de l'IA pour mieux comprendre les acteurs et leurs rôles	80
3.1.3 Un premier terrain pour comprendre l'organisation et les pratiques d'annotation des entreprises d'intelligence artificielle françaises	81
3.2 Le terrain en Afrique francophone : l'accès aux entreprises d'annotation et à leurs travailleurs	83
3.2.1 A l'autre bout de la chaîne de valeur : négocier un accès aux intermédiaires d'annotation et à leurs travailleurs	84
3.2.2 Sur le terrain. Quelles différences entre le distanciel et le présentiel ?	89
3.3 Exposé des techniques d'enquête et d'analyse	91
3.3.1 Méthodologies d'analyses qualitatives	91
3.3.2 Méthodologies d'analyses quantitatives	93

Partie 2 - Comment s'organise le processus de production de l'IA? Un agencement marchand qui repose sur la qualité des données, une organisation du travail qui s'appuie sur des *hiérarchies électroniques* **97**

Chapitre 4 - L'intégration logistique des travailleurs des données : la qualité de l'annotation comme limite du recours au "marché pur"	101
4.1 Décider du mode d'externalisation du travail des données. Une force de travail stable pour garantir la qualité des données et maîtriser la volatilité des modèles d'IA	107
4.1.1 Formaliser l'organisation du travail d'annotation dans les entreprises d'IA françaises : la qualité des données par la stabilité de la force de travail	108
4.1.2 Un impératif d'applicabilité du modèle qui diffère du monde académique : la confrontation avec le réel des modèles déployés en production	119
4.1.3 Maîtriser le processus d'annotation pour tenir compte de la volatilité des modèles d'IA	130
4.2 La mise en place de processus de communication entre annotateurs et employés des startup françaises pour gérer la qualité des données	144
4.2.1 Des outils d'arrangements pour la standardisation et la stabilisation des concepts de l'intelligence artificielle	147

4.2.2 Une forme d'organisation plus proche de la hiérarchie pour prendre en compte les boucles de rétroactions humaines sur les modèles	158
Chapitre 5 - La chaîne de valeur de l'intelligence artificielle : comprendre l'intégration des travailleurs dans la production de l'IA	166
5.1 Construire une typologie des activités de production de l'IA	170
5.1.1 Un continuum d'activités intensives en travail et en capital divisé entre les pays du Nord et ceux du Sud global. Modéliser d'un côté, produire des données et des matières premières de l'autre	171
5.1.2 La variété des organisations qui fabriquent les données pour l'IA : le cas de la chaîne de sous-traitance de Google	177
5.2 Approfondir l'hypothèse du "travail profond" à travers la comparaison des modèles d'externalisation : une organisation en apparence basée sur le marché qui a pourtant tous les traits de la "hiérarchie électronique"	181
5.2.1 Mturk comme outil d'annotation : une version "relationnelle" des plateformes de type marché "pur"	182
5.2.2 Les BPO intégrés aux plateformes : la forme la plus proche du "marché pur"	193
5.2.3 L'organisation du travail dans une entreprise d'impact sourcing	199
Conclusion du chapitre 2 - La mise en place d'une division internationale du travail de la donnée	206
Partie 3 - L'activité réelle des <i>petites mains</i> de l'IA : vers la solidification d'un groupe professionnel de travailleurs de l'IA	215
Chapitre 6 - La délocalisation de services pour pallier les limites de l'automatisation algorithmique	218
6.1 Le travail de l'optimisation algorithmique réalisé depuis Madagascar	223
6.1.1 Le travail algorithmique direct pour la numérisation des livres et des documents administratifs	224
6.1.2 Le travail algorithmique indirect des travailleurs malgaches au service des machines qui ordonnent le web	232
6.2 Les tâches complémentaires à l'automatisation algorithmique : l'essor des travailleurs des relations clientes dans les pays du Sud Global	237
6.2.1 Le travail de la voix en miroir du travail gratuit des consommateurs	238
6.2.2 Le travail du clavier : la relation client à l'ère du Web	246

Chapitre 7 - Les pratiques d'annotations : un travail de classification de données et de problématisation, négocié entre acteurs de la chaîne de valeur	256
7.1 Atteindre des standards de qualité par le management des travailleurs. Comment la formation et l'organisation du travail favorisent le développement d'une expertise dans le travail des données ?	262
7.1.1 Au-delà du travail du clic : l'externalisation d'un processus de production intégré . . .	262
7.1.2 Des rôles et des compétences diverses pour co-produire les annotations : une approche par les qualifications	267
7.1.3 Le travail prescrit de l'annotation de données : la formation des travailleurs	274
7.2 Une solidification des pratiques professionnelles qui passe par la formalisation des qualifications requises pour optimiser le fonctionnement des modèles	282
7.2.1 Découper, classer, saisir : le travail manuel des travailleurs des données	283
7.2.2 Comprendre, identifier, interpréter : des jugements experts orientés métiers	295
7.2.3 Classifier et problématiser : quand le travail des données devient un travail de l'IA . . .	308
Conclusion de la partie 3 - La professionnalisation du travail de l'IA : un travail de stabilisation des modèles, une explosion du travail lié aux systèmes d'IA	315
Partie 4 - A l'autre bout de la chaîne, la précarité des travailleurs des données	323
Chapitre 8 - La dynamique post-coloniale de "l'extractivisme des données" : l'attraction des investisseurs par l'existence de qualifications "coloniales"	325
8.1 Le développement du secteur repose sur un accès facilité aux facteurs de production pour les investisseurs étrangers	331
8.1.1 Le secteur se développe avant l'arrivée du numérique à Madagascar en s'appuyant sur le statut des entreprises franches	332
8.1.2 La qualification des travailleurs des données : des parcours de formation initiaux qui attestent de compétences suffisantes pour intégrer le secteur	343
8.2 Un processus d'assetisation coloniale pour transformer des qualifications coloniales en valeur de subsistance	347
8.2.1 Sous le regard des entrepreneurs de la sous-traitance. Analyser la tension entre captation et valorisation des qualifications et minimisation de la contribution des travailleurs	351
8.2.2 Les stratégies d'apprentissages des travailleurs de l'IA malgaches : une volonté d'échapper à leurs conditions par le (<i>virtual</i>) <i>migrant labor</i>	359

Chapitre 9 - Vers le développement d'une <i>global precarity chain</i> : l'industrie de l'IA vectrice d'une <i>intégration laborieuse</i> de la jeunesse urbaine malgache dans le marché du travail	369
9.1 Le secteur de l'externalisation comme unique piste d'entrée sur le marché du travail pour la jeunesse diplômée d'Antananarivo?	372
9.1.1 Une industrie qui repose largement sur la jeunesse diplômée de la classe moyenne urbaine	373
9.1.2 Un secteur qui permet une <i>stabilisation professionnelle et financière</i> des travailleurs malgaches	386
9.2 Une externalisation protéiforme. Des conditions de travail variables qui dépendent de l'appartenance au secteur formel ou informel?	397
9.2.1 Analyse objective des conditions de travail dans le secteur de l'externalisation : des variations importantes selon l'appartenance au secteur formel ou informel	398
9.2.2 Les conditions de travail des annotateurs	403
9.3 Des perspectives personnelles et professionnelles contrariées par la structure extractive du secteur	413
9.3.1 Un revenu de survie en décalage avec la valeur générée par l'industrie de l'IA : la "race" comme variable d'ajustement des marges dégagées par le secteur de l'IA	414
9.3.2 Perspectives personnelles et professionnelles futures des travailleurs. Faire avec des <i>emplois sans avenir</i>	425
Conclusion du chapitre 4 - Le travail d'annotateur : limiter la précarité en attendant mieux	435
Conclusion générale	444
Bibliographie	461
Annexes	491
A Analyses qualitatives	491
A.1 Description des enquêtes	491
A.1.1 Les employés de start-ups	491
A.1.2 Les employés d'entreprises d'externalisation	492
A.1.3 Les entreprises d'externalisation étudiées et les modalités de contrat	492
A.2 Codage des entretiens	492
B Passation du questionnaire	496
B.1 Exemple de négociation d'entrée de terrain avec des entreprises d'annotations	496

B.2	Brief envoyé aux annotateurs en vue de la réponse à l'enquête	497
B.3	Annonce de recrutement publiée sur Facebook	498
B.4	Le générateur de position avec les scores associés	499
B.5	Questionnaire	500
C	Analyse des données quantitatives	544
C.1	Analyses descriptives	544
C.2	Tri croisé pour l'analyse du capital social, économique et culturel des annotateurs	549
C.3	Analyse des conditions de travail	552
C.4	Analyse des revenus du foyer	556
C.5	Analyse des perspectives professionnelles des annotateurs	557
	Table des figures	558
	Liste des tableaux	560

Introduction générale

”Même nous on rigole, parce qu’on a vu une vidéo de [la start-up] sur YouTube, comme quoi c’était de l’intelligence artificielle, et on s’est sentis un peu insultés, parce qu’on est pas artificiels nous. C’est nous qui détectons les vols [...] C’était un peu bizarre de savoir ça. [...] Et il disait que tout était automatique, que ça anticipe des vols automatiques. Mais c’est pas vrai, c’est nous derrière. C’est un peu insultant, mais on peut rien y faire, nous ça nous permet de manger.” (Nity, Annotateur de données malgache - Entreprise sous-traitante A)

Nity est un jeune travailleur qui exerce au sein d’une entreprise spécialisée dans l’annotation de données qui opère depuis Madagascar pour deux start-ups d’intelligence artificielle (IA) françaises. Tous les jours, il visionne des images de supermarchés français et identifie les vols qui y sont commis. Lorsque nous discutons, cela fait un an et demi qu’il occupe ce poste, après une expérience comme photographe indépendant dans la capitale, Antananarivo. Dans cet extrait, il résume le problème : derrière le caractère automatique de la solution proposée par la jeune pousse technologique, il incombe en fait à des centaines de travailleurs de surveiller à distance nos commerces et d’éventuellement produire des données pour entraîner un modèle d’IA.

La start-up pour laquelle il travaille développe une solution ”automatique” appliquée à la vidéosurveillance des magasins qui fonctionne en détectant des gestes associés au vol de produits. Les algorithmes de la start-up traitent les images des caméras de surveillance, puis un système de messagerie envoie directement une alerte aux employés du magasin lorsqu’un vol est détecté. Loin d’être automatique, ce système délègue en fait la surveillance des magasins à plusieurs centaines de travailleurs malgaches et indonésiens. Ces derniers se retrouvent alors à identifier les vols en temps réel, mais également à annoter les images des vols (Figure 1.). Au quotidien, ces travailleurs visionnent des flux de vidéos en direct afin d’identifier les vols le plus rapidement possible pour que l’alerte soit donnée. Ils annotent également des images de vols en détaillant les éléments constitutifs de ces derniers : le client fait ses courses avec un sac à dos ou il a caché un produit dans son chapeau...

En d’autres termes, AIView externalise la production des données servant à entraîner ses algorithmes, mais également la surveillance en direct des magasins qu’elle équipe. Lors d’une discussion avec le président et fondateur d’une autre start-up d’IA, celui-ci indique : *”Là c’est du vrai BPO [Business Process Outsourcing.]”,* de l’externalisation de processus métier entier (*la surveillance en direct*), mêlé à une promesse d’*automatisation*. Cet exemple pose l’enjeu principal de cette thèse qui est donc de comprendre dans quelle mesure l’IA repose en fait sur l’externalisation des processus de production dans de nombreux secteurs économiques.

Nity fait partie de la centaine de travailleurs qui exécutent cette tâche et travaillent sur d'autres projets liés à l'intelligence artificielle. Ces travailleurs visionnent des images et des vidéos afin de les annoter pour produire les jeux de données qui servent à entraîner les algorithmes d'IA. Il peut s'agir d'images de voitures, de plateaux de cantines, de boîtiers de fibre optique et bien d'autres choses encore, selon les secteurs qui s'emparent des promesses de l'automatisation. En France, à l'autre bout de la chaîne, les caissières de cantine voient apparaître des bornes automatiques qui scannent les plateaux-repas. Des techniciens fibres se voient imposer une application qui détecte automatiquement la non-conformité des installations sur la base des photos qu'ils ont prises. Qu'y a-t-il de commun entre les travailleurs invisibles qui opèrent depuis Madagascar et les travailleurs français qui, à l'autre bout de la chaîne, achèvent de produire ces services? Chacun agit à différents moments et endroits de la chaîne de valeur pour produire un service.

À mesure que nous en dévoilons les différentes couches d'activités productives, l'imaginaire de l'IA déraile déjà. Il ne s'agit plus seulement de programmes, de robots, de machines qui auraient des effets néfastes sur ses utilisateurs, mais d'une infrastructure très humaine. Le retour à la sociologie du travail industriel et plus particulièrement à un article de William [Crossin \(1960\)](#) permet de distinguer ainsi une vision "techniciste" de l'automatisation, soit l'intégration de processus morcelés dans une chaîne de production automatisée qui peut se corriger elle-même et qui "*dispose de sa propre énergie*", d'une vision sociologique, qui se caractérise quant à elle par le degré de recours à la main de l'Homme.

Dans cette thèse, nous verrons que bien souvent, les services basés sur l'intelligence artificielle ne se corrigent pas d'eux-mêmes et que l'énergie dont ils disposent est en fait très humaine. Par exemple, la caisse de cantine automatique que nous évoquions précédemment ne détecte pas parfaitement les produits présents sur le plateau. Dans ce cas, c'est aux consommateurs de sélectionner le bon produit parmi plusieurs choix pour suppléer les "défaillances" de l'automation. Notre perspective se situe donc dans une seconde approche de l'automatisation qui passe par l'étude des processus de travail humain. Qui produit les services basés sur l'IA? Quels sont les différents acteurs de la chaîne de valeur qui concourent à la production de l'IA? Comment ces activités sont-elles divisées?

Notre partie 2 traitera de cette division des activités liées à l'IA. Dans ce chapitre, nous abordons plus directement le secteur de l'intelligence artificielle en montrant comment s'organise le processus de production de l'IA. Nous y abordons la variété des modalités d'externalisation du travail des données selon les modèles des entreprises d'IA. Nous verrons que l'externalisation des activités liées aux données passe par une plus forte intégration des travailleurs au processus de production. Ce mode d'externalisation, entre le recours au marché pur et le recours au modèle de l'entreprise laisse alors entrevoir la mise en place de véritables "*hiérarchies électroniques*" qui sont nécessaires pour gérer les jugements qui nourrissent les modèles et influent sur leur qualité. La plus grande intégration des travailleurs des données se traduit alors par la mise en place de standards de qualité et d'outils permettant de les réaliser.

L'imaginaire de l'IA est également particulièrement mobilisé pour donner un cap à cet ensemble de technologies que l'économiste Erik Brynjolfsson, Rock, et Syverson (2017) qualifie de "technologie d'intérêt général". Derrière ce terme, l'économiste envisage l'IA comme une technologie, qui comme l'électricité, n'atteindra sa pleine puissance de transformation économique et sociale qu'à mesure que les entreprises se restructureront pour l'intégrer dans leurs processus de production. Cette vision s'est traduite politiquement par un important encouragement financier au développement de l'IA et qui s'inscrit, par ailleurs, dans un mouvement général vers la numérisation des services publics et des entreprises Alauzen (2019). En France, la publication du rapport Villani (2018) et le lancement du plan d'investissement public "AI for humanity" par le président de la République Emmanuel Macron témoignent de cet engouement. Dans son discours au Collège de France pour la remise de ce rapport, le président de la République évoque l'intelligence artificielle en ces termes :

"Il y a chez Leibniz cette hypothèse que « Dieu calcule pour nous le meilleur monde possible » et il y a donc quelque chose d'une option presque prométhéenne qui nous permet de revisiter la conception du monde de Leibniz à travers l'intelligence artificielle qui nous donnerait la capacité de réaliser nous-mêmes ce calcul et à travers en effet des machines apprenantes de pouvoir parcourir beaucoup plus rapidement les chemins du malheur pour choisir le bon chemin beaucoup plus tôt et beaucoup plus rapidement."

Derrière l'idée de "calculer le monde", Emmanuel Macron fait référence à une des origines de l'apprentissage machine qui est à trouver dans la naissance de la statistique moderne et des probabilités. L'imaginaire de l'intelligence artificielle (AI) est foisonnant (Simut, 2017). Qu'il s'agisse du système informatique HAL 9000 de 2001, l'*Odyssée de l'espace* de l'agent Smith de *Matrix* ou bien de Samantha dans le film *Her*, notre vision de l'IA est travaillée par les représentations tirées de la culture populaire. Elles donnent corps à nos craintes et à nos espoirs, elles structurent les recherches des entreprises et des chercheurs reflétant souvent les idéologies dominantes. Il n'est donc pas étonnant de voir se développer tout un programme de recherche sur ce sujet, tant cette promesse technologique suscite de réactions¹. Dans le rapport Villani (2018), le député français va également dans ce sens en considérant que :

"Du fait de ses ambitions, qui en font un des programmes scientifiques les plus fascinants de notre époque, la discipline de l'IA s'est toujours développée de concert avec les imaginaires les plus délirants, les plus angoissants et les plus fantastiques, qui ont façonné les rapports qu'entretient le grand public avec l'IA mais également ceux des chercheurs eux-mêmes avec leur propre discipline. La (science) fiction, les fantasmes et les projections collectives ont accompagné l'essor de l'intelligence artificielle et guident parfois ses objectifs de long terme"

1. Voir cette [journée d'études](#) du 15 et 16 avril 2021 qui s'est déroulée à l'université de Strasbourg et qui porte sur "Que font les images de l'IA? Imaginaires, fictions et représentations."

Concernant les risques liés à l'IA, Emmanuel Macron adopte une posture d'équilibre entre ce qu'il appelle le progrès technique et le progrès humain, c'est-à-dire les impacts de la technologie sur l'économie et la société pour "réussir à faire à que cette utopie prométhéenne ne devienne pas une dystopie." Pourtant, de notre point de vue de sociologue, le risque n'est pas tant que notre monde ressemble à celui de *Terminator* ou bien à celui de *Matrix*, mais plutôt que ces technologies activent ou réactivent des logiques très anciennes d'exploitation. Pour répondre à ces craintes, on voit fleurir les discussions sur *l'éthique de l'IA* et les biais algorithmiques qui y sont associés, sans forcément qu'elles ne questionnent l'organisation du travail de l'IA et les risques que cela pose pour les travailleurs les plus précaires de la chaîne de production. Plus généralement, ce mouvement *éthique* s'inscrit dans l'idée qu'il faudrait rendre la technologie meilleure pour la société. Il s'agirait de faire en sorte que ces entreprises soient responsables, qu'elles agissent pour le bien commun et qu'elles contribuent à limiter leurs externalités négatives en collaborant avec les États pour, par exemple rendre les réseaux sociaux moins haineux et violents grâce à la modération de contenus.

L'IA cristallise donc un ensemble de "controverses sociotechniques" qui occultent parfois la réalité de son déploiement. Loin de ces fantasmes, la Commission européenne (2020) désigne l'IA comme "un ensemble de technologies qui combinent des données, des algorithmes et de la puissance de calcul". Pris à part, chacun de ces termes évoque une réalité matérielle. Les données sont collectées et stockées dans des *data centers* et contiennent une large part de notre vie personnelle (par exemple : des images, du texte...). Les algorithmes sont des programmes composés de règles permettant de produire une action ou de réaliser une prédiction. À notre époque, ces algorithmes sont autoapprenants, et doivent être entraînés sur de grandes quantités de données. Ce processus d'apprentissage requiert une importante puissance de calcul, autrement dit, le recours à des services *cloud*, souvent proposés par Google ou Amazon, dans des *data centers*.

Mais l'IA est avant tout un domaine technologique dont l'histoire remonte à la fin de la Seconde Guerre mondiale². Tout au long de son histoire, deux courants se sont opposés : entre les connexionnistes et les symbolistes, jusqu'à la victoire des approches reposant sur l'apprentissage profond (connexionisme) racontée par Cardon, Cointet, et Mazières (2018). Pour définir l'IA, il est essentiel de comprendre ce que cette consécration du connexionisme implique pour le fonctionnement des modèles actuels. De manière schématique, Cardon et al. (2018) résument les deux approches. Ils opposent la machine hypothético-déductive (symbolique) de la machine inductive (connexionniste). Au final, ce qui était conçu comme la partie "humaine" de la fabrication du programme ou du modèle, c'est-à-dire les règles qui régissent ces derniers, n'est

2. Lors des conférences de Macy (1946-1953) des ateliers de Dartmouth, s'opposent deux cadres théoriques. Dans le premier cadre, le terme cybernétique était utilisé pour décrire une "science générale du fonctionnement humain" afin de repenser la société. Au sein du groupe, plusieurs disciplines étaient représentées, comme la psychologie, l'anthropologie ou l'informatique. Le deuxième courant est composé principalement d'informaticiens, l'idée était d'essayer d'améliorer le fonctionnement du cerveau humain. L'hypothèse principale était que "chaque aspect de l'apprentissage ou de tout autre apprentissage ou de toute autre caractéristique de l'intelligence peut être décrit avec une telle précision qu'une machine peut être construite pour le simuler" (McCarthy et al., 1955).

plus ce qui est introduit dans le système, mais ce qui en résulte. Pour réaliser cette induction, il faut alors s'appuyer sur une masse conséquente de données à partir desquelles il sera possible de dériver les règles du calculateur.

Les auteurs nous expliquent aussi que les modèles connexionistes s'appuient sur les statistiques et la puissance de calcul pour élaborer des hypothèses de règles permettant de répondre à la problématique initiale du modèle, en s'appuyant sur des données d'entraînement. Ces hypothèses sont ensuite traitées et éliminées par le modèle qui retiendra la règle la plus pertinente. Toutefois, même s'ils sont dit "auto-apprenants", cela ne veut pas dire qu'il n'y a plus de théorie derrière les modèles, mais qu'elle se déplace dans les paramètres de la machine et dans les représentations des bases de données, comme le précise Cardon et al. (2018) : "l'horizon du calcul n'appartient pas à la calculatrice, mais au monde qui lui a donné des exemples "étiquetés". Les sorties - produites, symbolisées et biaisées par les humains - constituent aujourd'hui l'une des entrées les plus précieuses des machines connexionnistes (...)". En conséquence, la compréhension des modèles de "machine learning" passe par l'étude de la "composition des données d'entrée, [de] l'architecture choisie par les différents systèmes et [des] objectifs qui ont été donnés à la supervision de l'apprentissage". Autrement dit, c'est en étudiant le choix des paramètres finalement retenus, ainsi que des données d'entraînement, qu'il est possible de mieux comprendre ce que cachent ces algorithmes.

La promesse de l'intelligence artificielle réside également dans sa capacité à rendre calculables les phénomènes du monde réel. Pour reprendre les mots de Dominique Cardon (2015), il s'agit de « *chiffrer le monde* », de le « *rendre calculable en tout* ». À l'instar de la construction des statistiques modernes étudiées par Alain Desrosières (2010, 2013), il importe alors de comprendre comment se construit la mesure de ces phénomènes, de la définition des indicateurs à la collecte des données. Desrosières indique en effet que les *data* ne sont pas données, puisqu'elles sont le résultat d'un processus de production qui implique de penser des indicateurs et donc aussi des moyens de collecte de données. Selon lui, quantifier, c'est "se mettre d'accord et mesurer". Cela nous amène un autre ensemble de questions : comment se met-on d'accord sur ce que l'IA est censée calculer ? Qui est impliqué dans ce processus de *problématisation* des algorithmes ? Par conséquent, ne faudrait-il pas comprendre qui décide des "configurations" algorithmiques et qui sont les travailleurs impliqués dans ce processus *invisible* ?

Dans notre troisième partie, nous analysons les pratiques professionnelles de l'IA sur l'ensemble de la chaîne de valeur. En partant de l'analyse de plusieurs études de cas, nous montrons que ce travail est un processus collectif, distribué sur toute cette chaîne et fait l'objet de "négociations" entre les acteurs. Nous y analysons les activités liées au numérique et à l'automatisation réalisés par les travailleurs de Madagascar. Nous verrons également que les travailleurs effectuent différentes tâches "complémentaires", essentielles pour faire fonctionner la numérisation et l'automatisation. Nous montrons ainsi, que l'IA repose sur un travail de *classification* des données et de *problématisation* des modèles qui implique de nombreux acteurs et phases

pendant lesquels le travail humain de "calibrage" du modèle est essentiel et repose sur un travail qualifié. Ce faisant, nos données suggèrent une *solidification* des pratiques professionnelles par laquelle le travail des données se transforme finalement en un travail de l'IA à part entière.

Malgré les nombreuses critiques déjà abordées dans cette introduction, l'intelligence artificielle est néanmoins un projet scientifique, technique et économique en plein développement. Dans un rapport sur les investissements en matière d'IA, l'OCDE (2019) a ainsi mesuré la pénétration de cette technologie dans nos sociétés : *"Le volume des investissements a en effet doublé entre 2016 et 2017, pour atteindre 16 milliards USD en 2017. Les startups spécialisées dans l'IA ont attiré 12 % du capital-investissement mondial au cours du premier semestre de 2018, en nette progression par rapport à 2011, où elles n'en concentraient que 3 %"*. L'organisation internationale considère par ailleurs que, si la mise en place de cette technologie nécessite un certain nombre d'adaptations de la part des organisations, à mesure que *"les technologies et les modèles économiques gagnent en maturité, on s'achemine vers un déploiement à grande échelle de l'IA."*

Cinq années plus tard, le [rapport Artificial Intelligence Index 2022](#) de Stanford a confirmé l'accélération de la pénétration de l'IA dans la recherche et dans l'économie. Dans la recherche scientifique, le nombre de publications passe de plus de 160.000 en 2010 à plus de 330.000 en 2021. En ce qui concerne la recherche et développement, on constate une forte augmentation du dépôt de brevet, avec 30 fois plus de brevets déposés en 2021 par rapport à 2015. La Chine continue de renforcer sa place de leader sur cette technologie en publiant un *"nombre de publications dans des revues, conférences et dépôts de données sur l'IA - 63,2% supérieurs aux États-Unis"*. L'Europe est quant à elle deuxième au niveau des publications, des citations, des publications tirées de conférences, à chaque fois derrière l'Asie de l'Est et le Pacifique et devant l'Amérique du Nord et troisième au niveau des demandes de brevet déposés, derrière l'Amérique du Nord et les pays de l'Asie de l'Est et du Pacifique.

Le rapport note encore que l'investissement privé et public a été quasiment multiplié par 10 entre 2016 et 2021 pour atteindre 176,47 milliards de dollars. Par ailleurs, on observe également une augmentation très forte des investissements privés dans l'IA, avec pour 2021 *"environ 93,5 milliards de dollars, soit plus du double de l'investissement privé total en 2020."* Au sujet de la France, le rapport note qu'en comptabilisant l'ensemble des investissements privés entre 2013 et 2021, la France se situe au 8e rang avec 3,9\$ milliards d'investissements, contre 4,4 pour un pays comme l'Allemagne et 5,7 pour le Canada. Les États-Unis et la Chine sont très loin devant avec respectivement 149 et 61,9 milliards de dollars.

Au niveau européen, la France se situe dans le top 3 d'après le [dossier de presse](#) du ministère de l'Économie français annonçant la deuxième phase de sa stratégie nationale pour l'IA. Depuis la parution du rapport Villani en 2018, la France s'est en effet dotée d'une stratégie nationale dont la première phase portait sur l'accompagnement et le financement de la recherche en matière d'IA et dont la seconde vise à former et

à attirer des "talents". Cette dernière s'est traduite par une augmentation des financements publics dédiés à cette technologie, de 170 millions d'euros en 2018 à 1 180 millions en 2021. [cartographie](#) de l'association France Digitale identifie 502 startups spécialisées en IA en 2021, en hausse de 11% par rapport à 2020. Enfin, l'importance prise par le secteur de l'IA se traduit de plus en plus en termes d'emplois : *"13 459 personnes travaillent dans les start-ups de l'IA en 2021 (pour 70 000 emplois indirects générés), c'est-à-dire une augmentation de 35% par rapport à 2020 (10 008 employés) (...)"*. Si la France est loin des pays comme la Chine et les É.-U. en matière d'investissement et de production de recherche, elle fait tout de même partie des pays comme le Canada ou l'Allemagne dans lesquels le secteur se développe de plus en plus.

Le [Artificial Intelligence Index Report 2022](#) analyse également les domaines dans lesquels se concentrent le plus les investissements en matière d'IA. Ainsi, les données des investissements réalisés entre 2017 et 2021 montrent que la *"catégorie médicale et soin de santé a reçu le plus gros investissement privé au niveau mondial (28,9 milliards de dollars), suivi par la gestion, le traitement et le cloud des données (26,9 milliards de dollars), la fintech (24,9 milliards de dollars) et le commerce de détail (21,95 milliards de dollars)"*. En matière d'adoption par l'industrie, le rapport constate que la : *"plus grande adoption a eu lieu dans le développement de produits et/ou de services pour les hautes technologies/télécommunications (45%), suivi par les opérations de service pour les services financiers (40%), les opérations de service pour les hautes technologies/télécommunications (34%), et la fonction de risque pour les services financiers (32%)"*. Enfin, le rapport note que le taux d'intégration le plus élevé concerne *"La compréhension de textes en langage naturel pour l'industrie des hautes technologies/télécommunications (34%), suivie par l'automatisation des processus robotiques pour les services financiers et l'industrie automobile et de l'assemblage (33%) et la compréhension de textes en langage naturel pour les services financiers (32%)"*.

Ces deux derniers points tendent à montrer que, même si l'IA est considéré par certains comme une technologie qui vise à pénétrer l'ensemble des secteurs, elle reste pour le moment cantonnée au secteur du numérique. Par ailleurs, son adoption par l'industrie se fait, pour l'instant, principalement par l'usage de certaines technologies comme le *Natural Language Processing* (NLP) qui correspondent à des services basés sur du texte comme les chatbots ou des outils d'aide à la lecture de contrat dans les banques ou les assurances. L'usage de la *"robotic process automation"* permet quant à elle d'automatiser des services ou des machines permettant la production de biens. Par exemple, améliorer grâce à l'IA des robots produisant des voitures, ou dans les services, automatiser le traitement de factures. D'un point de vue qualitatif, l'OCDE rapporte encore qu'un certain nombre de secteurs critiques sont ou seront touchés par des évolutions liées à cette technologie. Il s'agit par exemple de l'application de l'IA au secteur de la santé, notamment pour l'aide au diagnostic dans la santé, de la supervision de la conduite dans les transports ou bien de tout un ensemble de techniques prédictives dans la justice comme dans l'analyse des risques de récidive ou dans le secteur financier pour la détection de la fraude. Citons également, le cas de la surveillance, secteur pour lequel Ste-

ven [Feldstein \(2019\)](#) parle d'une expansion globale du recours à l'IA avec 75 pays sur les 176 étudiés qui ont recours à cette technologie, notamment pour la police intelligente ou bien pour la reconnaissance faciale. Dans l'Union européenne (UE), cette surveillance se traduit notamment par des projets comme [Roborder](#)³, dont l'objectif est de mettre en place une solution de détection automatique de menace aux frontières à l'aide de drones autonomes. C'est en partie ce type d'usages de l'IA qui a conduit à une inflation des publications sur l'éthique de l'IA ([Jobin, Ienca, & Vayena, 2019](#)). Toutefois, on constate que la question du travail des données n'y est pas nécessairement abordée, du moins, pas de manière systématique.

En matière de recherche, le [Artificial Intelligence Index Report 2022](#) de Stanford confirme cette tendance de fond en soulignant la montée en puissance des recherches consacrées à l'éthique de l'IA⁴. Depuis plusieurs années, un corpus de rapports administratifs, d'entreprises ou d'organisations non gouvernementales s'est en effet constitué sous le poids des craintes vis-à-vis de l'IA, mais tiennent très peu compte des enjeux liés à la production de données. Pour donner du sens à ces *guidelines*, une littérature s'est également développée sur la production de la régulation de l'IA. L'objectif est d'identifier les acteurs qui produisent ces cadres de régulation les thèmes principaux qui y sont abordés. [Jobin et al. \(2019\)](#) ont analysé des dizaines de documents et retiennent que les thèmes principaux, sont : la transparence, la justice l'équité, la non-malfaisance, la responsabilité, la vie privée, la bienfaisance, la liberté et l'autonomie, la confiance, la soutenabilité et la dignité et la solidarité⁵. Mihály Héder [2020](#) identifie quant à lui deux enjeux propres à l'IA : le contrôle par les humains et la transparence. Les travaux de Anna Jobin et al. [2019](#) et de Jessica Fjeld et al. [2020](#) identifient plusieurs enjeux qui font indirectement référence au travail des données, respectivement la nécessité d'une "responsabilité humaine sur les systèmes d'IA" et la nécessité de disposer de "données de haute qualité". Toutefois, on constate que la question de l'annotation n'est pas mentionnée explicitement, ce que confirme Thilo Hagendorff [2020](#) en soulignant que le micro-travail est nécessaire au *machine learning* et s'accompagne de nombreux problèmes sociaux "qui ont tendance à être ignorés par la communauté de l'IA."

Au sein de ce corpus, un certain nombre de textes internationaux sans visée normative tente également d'encadrer le développement des politiques publiques nationales en matière d'IA. C'est le cas de deux textes, les "*Principles on Artificial Intelligence*" (OECD, 2019) et les "*Recommendations on the ethics of artificial in-*

3. <https://roborder.eu/>

4. "Le développement de l'éthique de l'IA est omniprésent : la recherche sur l'équité et la transparence dans l'IA a explosé depuis 2014, avec une multiplication par cinq des publications connexes dans les conférences liées à l'éthique. L'équité et la partialité des algorithmes sont passées d'un statut de recherche académique à celui de sujet de recherche grand public avec des implications de grande envergure."

5. Dans la même logique, deux autres travaux identifient des thèmes très similaires (Fjeld et al., 2020; Hagendorff, 2020; Héder, 2020). Ainsi, Jessica Fjeld et al. (2020) identifient les thèmes suivants : « privacy », « accountability », « safety and security », « transparency and explainability », « fairness and non-discrimination », « human control of technology », « professional responsibility and promotion of human values ». Quant à Thilo Hagendorff (2020), il identifie des problématiques bien couvertes : "privacy protection", "fairness, non-discrimination, justice", "accountability", "transparency, openness", "safety, cybersecurity", "common good, sustainability".

telligence” (UNESCO, 2020). L’OCDE mentionne très indirectement des enjeux liés au travail des données. Ainsi, l’OCDE rappelle que “les acteurs de l’IA devraient respecter l’État de droit, les droits de l’homme et les valeurs démocratiques, tout au long du cycle de vie du système d’IA.”. L’UNESCO aborde plus directement cette question, d’abord sous l’angle de la « data justice » (p.30), c’est-à-dire le fait que les jeux de données doivent représenter au mieux la diversité humaine en recommandant “*l’élaboration de normes pour l’annotation des ensembles de données devrait être encouragée, y compris la ventilation des données en fonction du genre et d’autres priorités*”, comme la représentation des minorités, puis sous l’angle de la formation des chercheurs à l’éthique et notamment à la « *façon dont ces ensembles sont annotés* » (p.36). La représentation des minorités est également présente dans le travail du « *data governance working group* » (data justice) du Global partnership on AI (GPAI).

Le travail de régulation de l’UE est, quant à lui, incontournable en ce qu’il constitue le plus abouti au monde à ce jour, conduisant certains chercheurs à parler d’*“effet bruxellois”* pour définir l’impact potentiel de la régulation européenne sur le secteur mondial de l’IA. L’AI Act, proposé le 22 avril 2021 est actuellement en discussion au Conseil de l’Europe et vise à stimuler et encadrer l’innovation en matière d’intelligence artificielle, en protégeant mieux les consommateurs. Dans un commentaire de la directive, [Veale et Borge-sius \(2021\)](#) détaillent les différents régimes de conformité retenus par la Commission européenne : « les systèmes aux risques inacceptables » (Titre II), « les systèmes à hauts risques » (Titre III), « les systèmes aux risques limités » (Titre IV), « les systèmes aux risques minimaux ». Concernant les données utilisées dans la conception des systèmes à haut risque, l’article 10 prévoit que les entreprises respectent un certain nombre de critères de qualité des données : représentativité, « *conception pertinente* », « opérations de traitement pertinentes pour la préparation des données, telles que l’annotation »... L’article 17, prévoit que les acteurs mettent en place un « système de gestion de la qualité ». Une des dispositions les plus décisives pour l’IA commerciale concerne l’obligation de détailler le fonctionnement des systèmes d’IA présentant un haut risque, à travers un mécanisme de conformité qui consiste à obliger les entreprises à publier une documentation technique comportant plusieurs points (Article 11.). Les points devant figurer dans cette documentation sont précisés en annexe. Ainsi, concernant les choix de conception des modèles, pourront être publiés : « *les spécifications de conception du système, à savoir la logique générale du système d’IA et des algorithmes; les principaux choix de conception, y compris le raisonnement et les hypothèses retenues (...) les principaux choix de classification* ». Concernant les données et le travail des données, plusieurs éléments devront figurer sur la documentation technique et notamment : « *les exigences relatives aux données en ce qui concerne (...) les méthodes et techniques d’entraînement et les jeux de données d’entraînement utilisés, y compris des informations sur la provenance de ces jeux de données, leur portée et leurs principales caractéristiques; la manière dont les données ont été obtenues et sélectionnées; les procédures d’étiquetage (...), les méthodes de nettoyage des données (...)* »

Ainsi, même si la question de travail des données est relativement absente des débats sur la régulation, l'AI Act prévoit tout de même que des informations concernant les opérations réalisées sur les données soient publiées et mises à jour régulièrement. De nombreuses questions restent ouvertes quant à l'application de concept parfois vague (par exemple : la « conception pertinente » des données) et la manière dont le processus de production des données et des modèles se déroule. La fabrication des jeux de données constitue, en creux, un enjeu technique également souligné par le [Artificial Intelligence Index Report 2022](#) de Stanford : *“Les meilleurs résultats obtenus dans le cadre des critères de référence techniques reposent de plus en plus sur l'utilisation de données d'entraînement supplémentaires pour établir de nouveaux résultats de pointe. En 2021, 9 systèmes d'IA de pointe sur les 10 critères de référence de ce rapport sont entraînés avec des données supplémentaires. Cette tendance favorise implicitement les acteurs du secteur privé ayant accès à de vastes ensembles de données”*. Notre travail contribue également à cet effort collectif en documentant les choix humains qui gouvernent le fonctionnement final des modèles d'IA.

Les acteurs privés et publics déploient des efforts importants afin de rester dans la course mondiale à la maîtrise et au déploiement de cette technologie. Comme indiqué précédemment, la France constitue quant à elle, un des pays majeurs de l'IA que ce soit du point de vue de la recherche ou de la création d'entreprises spécialisées. En parallèle, le développement de cette technologie s'est également traduit par un ensemble de controverses, qui sont en passe de se traduire en droit⁶. Concernant le travail des données, nous constatons qu'il s'agit d'un sujet relativement absent des débats bien qu'étant pourtant à l'intersection de plusieurs grands enjeux tels que la vie privée à travers l'externalisation de données parfois sensibles ou les biais liés au travail d'annotation. D'une certaine manière, c'est aussi cette vision qui invisibilise les travailleurs comme Ny Tocky qui contribuent à produire les modèles.

Outre les débats sur la régulation de l'IA, des efforts récents comme le *Digital Services Act* qui visent à améliorer la modération des contenus font ressurgir la place prépondérante des travailleurs du Sud Global qui entretiennent nos infrastructures numériques. Qu'y a-t-il en commun entre les débats sur les biais de l'intelligence artificielle (IA) et ceux sur la modération de contenus, comme en 2019 autour de la proposition de loi portée par la députée française Laetitia Avia de lutte contre les contenus haineux ? Dans le premier cas, les débats se cristallisent autour de la capacité des acteurs de l'IA à prendre en compte ces effets négatifs potentiels et à adoucir leur impact. Dans le second, une des dispositions particulièrement contestées de la loi Avia revient à encourager les plateformes à recourir à un filtrage automatisé des contenus publiés sur les réseaux sociaux, pour respecter l'obligation de retrait de ces contenus en 24h. Basé sur l'intelligence artificielle, et sans regard de la puissance publique sur le fonctionnement des algorithmes, ce principe pourrait conduire, de fait, à privatiser le pouvoir de censure. Surtout, le projet occulte toute réflexion sur la part

6. Comme le souligne d'ailleurs, le rapport [Artificial Intelligence Index Report 2022](#) de Stanford : une seule loi mentionnant l'IA est passée en 2016 contre 18 en 2021

toujours importante du travail humain, qui en complément des algorithmes vient partager la responsabilité de maîtriser la parole dans l'espace public numérique⁷.

La catastrophe sanitaire du Covid-19 a par exemple conduit des entreprises comme YouTube à se passer de modérateurs humains. Ce faisant, l'entreprise californienne a indiqué qu'elle s'attend à ce que des erreurs perturbent son service⁸. En dépit de la promesse d'une organisation technico-sociale débarrassée de "l'incompétence" humaine, les dispositifs d'IA reposent donc sur le travail humain. Lorsque ces modérateurs sont basés dans des pays comme les Philippines, comme le montre Sarah T. ROBERTS (2019), il est nécessaire de s'attarder sur une caractéristique majeure de nos vies numériques : les services que nous utilisons quotidiennement sont largement dépendants de chaînes de valeur internationales qui structurent leur fonctionnement. Il est alors nécessaire de se demander pourquoi et comment les entreprises occidentales mettent en place ces organisations du travail ?

Lors d'une expérience professionnelle antérieure à cette thèse, en tant que rapporteur au Conseil national du numérique, j'ai évoqué le cas de l'externalisation de la modération dans le cadre d'un atelier de "co-construction" consacré à la régulation des contenus illicites publiés en ligne. Associant des fonctionnaires, des associations et des chercheurs, l'objectif était d'envisager les régulations potentielles de cette problématique. Lorsque j'ai évoqué le cas de modérateurs embauchés dans de mauvaises conditions de travail dans certains pays en voie de développement, un participant m'a tout de suite coupé la parole en me disant : *"j'ai travaillé plusieurs années à la Banque Mondiale et je pense qu'il vaut mieux qu'ils aient accès à ce type d'emplois plutôt que de ne rien avoir du tout"*. Ce type de remarque illustre bien certains débats en cours sur les perspectives de développement des "nouveaux" métiers du numérique.

Dans ce cas, il est évident qu'on ne parle pas des métiers de consultant en informatique, en cybersécurité ou bien de chef de projet dans le développement web. Non, il est plutôt question de modération, de relations client, de saisie de données, ou bien d'un certain nombre de tâches d'annotations liés à l'intelligence artificielle. Une dernière question émerge alors : quels sont les effets de cette externalisation sur les pays qui hébergent ces travailleurs, et sur les travailleurs eux-mêmes ? Nous traiterons de cette dimension dans notre troisième partie qui porte sur les conséquences du travail de l'IA sur les travailleurs des pays en voie de développement. Nous montrons que l'IA est un secteur "extractif" qui repose sur l'extraction de valeurs de travailleurs faiblement rémunérés, malgré leur niveau de formation. Au travers de notre enquête, nous mettons ainsi au jour les chaînes de production mondialisées au sein desquelles se déploient de nombreux métiers qui produisent, déploient et maintiennent les infrastructures algorithmiques.

7. Initialement rejetée par le Conseil constitutionnel, cette disposition est prévue par le Digital Services Act européen (DSA), adopté en octobre 2022 et a déjà été transposée en droit français

8. Voir cet [article](#), qui indique notamment que : *"Because of the heavier reliance on AI, YouTube basically says we have to expect that some mistakes are going to be made."*

Comme me l'a raconté, Miary l'un de mes informateurs privilégiés durant mon enquête de terrain : "L'IA ne marche pas sans les job 2.0 des pays sous-développés". Au bout du compte, c'est l'ensemble du numérique français qui ne pourrait fonctionner sans ces "jobs 2.0". En témoigne l'étendue des activités que nous avons observées à Madagascar : call centers, modération de contenus, relation client par mail et bien sûr saisie et annotation de données. Nity appartient à ce vaste secteur de l'externalisation numérique, dont la production de données pour l'IA constitue la dernière manifestation.

L'enjeu de notre travail est alors de faire parler toutes ces voix oubliées et invisibles du numérique, de faire voir leur position sociale, d'identifier leur place dans des chaînes de production mondialisées et de montrer leur contribution à une des technologies majeures de ces dernières années. Ce sera l'objet de notre partie quatre qui articule la question du développement avec des questionnements sur les trajectoires personnelles et professionnelles des travailleurs des données et les bénéfices qu'ils tirent de cette activité pour s'interroger sur : à quoi rêvent les travailleurs des algorithmes ?

Plus globalement, ce débat illustre la part de plus en plus importante que prend l'intelligence artificielle dans nos vies. Le mythe de la régulation par l'automatisation, celui de la "fin du travail" constituent autant de thématiques qui conduisent à éviter de s'attarder sur les conditions réelles de production de l'intelligence artificielle. Ces dimensions si structurantes du numérique renvoient à une caractéristique classique de la technique, une : " *infrastructure invisible quand elle fonctionne, devient visible quand elle est défaillante*" (Star, 2010).

* * *

Ces orientations résonnent avec les grandes perspectives de la littérature en sciences sociales consacrée à l'intelligence artificielle et aux algorithmes, qui s'est principalement intéressée à l'étude des biais et de leurs conséquences sur les utilisateurs et l'étude des effets de l'IA sur le travail. Comme nous le verrons dans notre première partie, la littérature a pour le moment porté, soit sur l'analyse de la fabrication des modèles, principalement dans le monde scientifique et des effets sur leurs utilisateurs (*sociologie de l'IA*), soit sur l'analyse des conditions de travail des "petites mains", qui comme Ny Tocky participe à la phase d'annotation de données (*micro-work*).

D'après nous, pour analyser les effets de l'IA sur le travail, il est pourtant nécessaire de prendre en compte l'ensemble de la chaîne de production des services que l'IA transforme. Pour cela, il est nécessaire d'orienter le regard, non pas vers la fabrication des modèles, mais plutôt sur comment sont conçus les produits et des services basés sur l'IA et sur ce que ce processus de production implique en termes d'évolution du travail.

Précisément, notre thèse consacrée aux travailleurs des données de l'intelligence artificielle s'inscrit au carrefour de ces deux champs de recherche. Dans les deux prochaines sections, nous revenons sur les orienta-

tions, d'abord de la sociologie des algorithmes et de l'IA, puis du *digital labor* et nous proposons finalement de revenir sur la manière dont nous avons construit notre objet, "le travail des données de l'IA" autour de trois facettes : étudier le travail de l'IA, étudier les effets de l'IA sur le travail et étudier le l'IA par le travail.

En suivant les traces de plusieurs traditions de recherches sur le travail et la technologie, nous proposons de s'interroger sur la réalité du processus de production de l'IA. Comment les entreprises d'IA produisent-elles les systèmes basés sur l'intelligence artificielle? Qui sont travailleurs humains qui définissent le fonctionnement de ces algorithmes? *In fine*, cela revient à se demander qu'est-ce que le travail de l'IA, voire même qu'est-ce que l'IA?

En partant de cette perspective, notre thèse vise à étudier la structure de l'industrie de l'IA sous l'angle des dynamiques organisationnelles derrière l'externalisation des activités qui suppléent l'automatisation algorithmique. Nous montrons que les structures de délégation du travail (plateformes numériques de travail ou sous-traitants) masquent en réalité la continuité des formes hiérarchiques et leur application à cette nouvelle industrie (partie 2). La nécessité de définir et de maintenir un haut niveau de qualité des données sous peine d'altérer les modèles pousse les entreprises d'IA à intégrer les travailleurs de Madagascar à leurs processus de production. Cette façon de structurer la production est également rendue nécessaire par l'importance accordée à la formation des travailleurs et par la reconnaissance de leur place importante dans la production. En ce sens, nous montrons que le travail des données s'inscrit dans une tendance plus large à la numérisation et à l'approfondissement de l'externalisation d'activités numériques dans les pays en voie de développement. Concernant l'IA, l'analyse des activités des "petites mains" de l'IA, montre qu'ils exercent en fait un travail qualifié qui contribue à la "problématisation" des modèles d'IA (partie 3). Enfin, nous nous intéressons aux conditions de vie et de travail des travailleurs des données malgaches en articulant les effets de la structure du secteur et l'agentivité des travailleurs. Nous montrons l'émergence d'un extractivisme des données basé sur la colonialité de la chaîne de production, limitant ainsi leurs perspectives d'évolutions.

Partie 1 : Étudier le processus de production des systèmes d'IA pour en étudier les conséquences sur le travail

Dans cette thèse, nous travaillons sur les "systèmes d'IA" (SIA) pour reprendre la terminologie de l'AI Act. Nous examinons plus particulièrement le travail de conception des algorithmes et les interactions entre les entreprises qui conçoivent les modèles et les travailleurs qui produisent les données. Nos deux premiers chapitres sont consacrés à la revue de la littérature en sciences sociales qui traite d'une part des algorithmes et de l'IA (Chapitre 1) et d'autre part du micro-travail (Chapitre 2).

A la fin de cette partie, nous reviendrons plus en détail sur nos questions et hypothèses de recherche. Nous y présentons un cadre d'analyse qui articule l'étude des pratiques d'annotations avec les enjeux liés à l'externalisation de ces activités et notamment les conséquences sur le travail. Au-delà des débats sur la fin du travail et la déqualification des métiers du fait du développement de l'IA, nous proposons une analyse des effets du travail de l'IA sur les pays du Sud Global qui hébergent les activités liées aux données. Notre troisième chapitre est consacré à la présentation de la méthodologie suivie lors de notre enquête. Nous y détaillons les différentes phases de notre travail, ainsi que les différentes catégories d'acteurs étudiés dans cette thèse. Par ailleurs, nous proposons quelques éléments d'explication au sujet du choix des travailleurs des données de Madagascar. Une dernière partie de ce chapitre est consacrée aux difficultés d'accès à notre terrain d'étude et à un retour critique sur les modalités d'analyse de cette thèse.

Chapitre 1 - Les deux "ordres" de la sociologie des algorithmes : l'IA et ses effets et l'IA comme expression de ceux qui la fabriquent

Dans une revue de littérature à propos d'un sous-ensemble d'IA appelé les systèmes experts, "des systèmes informatiques d'aides à la décision" à la mode dans les années 1980, Steve Woolgar (1985) suggère une première approche dans l'étude des systèmes d'IA. Il y amorce une critique, qui sera reprise plus tard, à propos de la tendance des recherches sociologiques sur l'IA à se focaliser sur les effets de l'IA sur ses utilisateurs (Boden, 1987). Il considère ainsi que lorsque ces chercheurs parlent de social, il s'agit du "social (dans l'IA), qui a un rapport avec les effets de l'IA, mais pas avec sa genèse". Selon lui, il serait pourtant nécessaire d'intégrer les apports des nouvelles théories en sociologie des sciences en évitant toute dichotomie entre ce qui est de l'ordre de la science ou de la technique et ce qui est de l'ordre du social. Par conséquent, un des enjeux

soulignés par Woolgar est que la sociologie des machines doit se saisir des *”activités de recherche elle-même”* (autrement dit des activités de fabrication des modèles).

Pour combler ces lacunes, Steve Woolgar (1985) propose deux manières d’envisager l’étude de l’IA : le rapprochement des sociologues et des chercheurs en IA dans leurs méthodes et objets de recherche et la sociologie des pratiques de recherches en IA. Cette première catégorie d’analyse consisterait à développer une sociologie des phénomènes étudiés par l’IA qui entraînerait *in fine* le développement des interactions entre les deux disciplines. D’une certaine manière, le *”moment big data des sciences sociales”* tel qu’il est mentionné par Gilles Bastin et Paola Tubaro (2018), et qui désigne l’exploitation d’importantes masses de données notamment extraites des réseaux sociaux par les sciences sociales correspond en partie à cet enjeu. Tout comme l’ensemble des nouvelles perspectives ouvertes par les sciences sociales computationnelles qui empruntent parfois aux outils des concepteurs d’IA (Boelaert & Ollion, 2018; Chateauraynaud, 2019; Do, Ollion, & Shen, 2022). Une alternative à ces analyses passerait par l’étude ce que Steve Woolgar appelle les *”AI research practices”* et qu’il développe autour de quatre pistes : la sociologie des chercheurs, l’étude de la fabrication des SIA, la sociologie des machines intelligentes et l’étude des discours sur l’IA.

Une première piste consisterait ainsi à étudier la sociologie des chercheurs qui conçoivent ces systèmes, en identifiant leurs caractéristiques sociales susceptibles d’influencer le fonctionnement des modèles. L’étude de la fabrication des systèmes d’IA consiste à analyser les hypothèses et présupposés scientifiques sur lesquels repose le développement des produits basés sur l’IA. Cette sociologie de la *”caractérisation”* consisterait très concrètement à étudier comment sont fabriquées et utilisées ces machines. L’auteur considère que *”cette approche accorde la priorité aux humains en tant qu’agents constructeurs, et adopte implicitement la distinction clé entre les humains et les machines qui imprègne le discours sur l’IA.”* De notre point de vue, la distinction entre humains et machines qu’implique cette approche semble exagérée. Notre programme de recherche consiste précisément à considérer les systèmes basés sur l’IA comme une agrégation de travail humain et non comme des machines ni intelligentes et ni automatiques. Plutôt qu’une distinction, cette approche devrait donc plutôt passer par une *disparition* de la machine pour réintroduire ceux qui les fabriquent : les *”agents constructeurs”* quels qu’ils soient (incluant donc les annotateurs de données).

La sociologie des machines intelligentes renvoie au fait de placer la machine au centre des investigations en considérant *”les machines intelligentes comme des sujets d’étude”*. Ainsi, pour Woolgar, il serait possible *”d’imaginer la programmation réussie d’une machine pour produire des réponses à un questionnaire ou à des questions d’entretien”*. D’une certaine manière, les recherches contemporaines consistant à proposer des méthodes d’audit des modèles algorithmiques rentrent dans cette catégorie. C’est le cas de Diakopoulos (2015) qui propose d’*”auditer”* les résultats de plusieurs algorithmes dans une démarche journalistique. Plus récemment, la thèse de Henin (2021) vise à contribuer aux impératifs légaux d’explicabilité des algorithmes⁹

9. Notamment prévus par la loi République Numérique et par l’AI Act pour les systèmes à haut risques

en créant un système interactif d'explication des boîtes noires. Enfin, citons le travail de Mackenzie (2017) qui analyse l'IA à travers une auto-ethnographie de sa pratique d'apprentissage du *machine learning*. Enfin, une dernière possibilité consiste selon Steven Woolgar à proposer une sociologie des discours en étudiant la communauté des experts de ces machines, en analysant le lien entre les discours sur l'IA et les pratiques concrètes de la recherche sur l'IA.

Plus récemment, deux autres revues de littérature ont retenu notre attention. La première est proposée par K. Joyce et al. (2021) et reprend de nombreuses recherches anglo-saxonnes en sociologie des algorithmes. Dans cette publication, les auteurs distinguent deux catégories de publication : la première concerne *"les politiques des algorithmes, des données et du code"* et la seconde *"le façonnage social de l'IA dans la pratique"*.

Au sein de cette première catégorie de travaux, on retrouve des recherches sur la "non-neutralité" des données et donc des modèles. Pour ces recherches, une *"compréhension sociologique des données est importante étant donné qu'une utilisation non critique des données humaines dans les systèmes sociotechniques d'IA aura tendance à reproduire, voire à exacerber, les inégalités sociales préexistantes."* Une deuxième catégorie concerne la non-neutralité des algorithmes et montre qu'au-delà des discours sur leur prétendue objectivité, ces derniers renferment au contraire des valeurs reproduisant le contexte social et culturel dans lequel ils s'inscrivent. Plus récemment, cet axe de recherche s'est développé en réaction à l'expansion des techniques de *machine learning*, de manière à proposer un discours critique. C'est par exemple le cas du travail de Noble (2018), qui montre que le contexte culturel et la position des travailleurs de Google affectent les résultats du moteur de recherche.

La question de l'encastrement social de l'IA est également traitée sous l'angle de l'étude de la fabrication des modèles. Au sein de cette catégorie, les auteurs distinguent trois approches. La première concerne le contexte politique et organisationnel dans lequel l'IA émerge et fait notamment référence à l'ethnographie de laboratoire réalisé par Hoffman (2017), dans laquelle il compare deux laboratoires, l'un financé par le gouvernement, l'autre par le secteur privé et montre le lien entre la source du financement et les stratégies de recherches. La deuxième piste d'études a consisté à montrer les effets de l'introduction des systèmes d'IA sur ses utilisateurs. À partir de l'étude du déploiement d'un système d'IA au sein de la police Brayne et Christin (2021) a montré comment le pouvoir hiérarchique encourage, voire impose, l'utilisation des systèmes, malgré les résistances initiales de certaines catégories d'agents. Une dernière approche consiste à traiter la question des liens entre l'IA et le travail, que ce soit les nouvelles catégories de travail créées (par exemple : le travail des données) ou bien les nouveaux modes de distribution de certaines activités (par exemple : les plateformes numériques de travail).

Dans deux articles consacrés à la sociologie des algorithmes et de l'IA Fabian Muniesa (2019) et Jean-Sébastien Vayre (2021) amorcent quant à eux une réflexion sur l'analyse de l'intelligence artificielle en

sociologie et plus particulièrement en *"Science and Technology Studies"* (STS). Ils y évoquent deux "ordres" de la littérature critique de l'IA et des algorithmes en proposant de distinguer deux catégories de littérature critique en sciences sociales : une critique interne qui *"consiste à réfléchir à la manière dont ces dispositifs peuvent ou non être pensés, fabriqués et mobilisés [...] pour observer la société"*, et une critique externe qui examine quant à elle *"la manière dont les technologies d'intelligence artificielle, en tant que conception, expriment les représentations de tous ceux qui participent à leur fabrication"*. Le reproche principal adressé par Jean-Sébastien Vayre (2021) aux travaux formulant une critique interne est une certaine tendance à faire disparaître les humains qui permettent de faire fonctionner ces machines pour se concentrer uniquement sur leurs effets sur les acteurs, que ce soit des consommateurs ou des travailleurs. Ce faisant, ces recherches manquent parfois de chercher les causes réelles des effets négatifs qu'elles pointent¹⁰.

À travers cette introduction, on perçoit d'emblée les différentes lignes de démarcation du champ d'études qui s'est constitué autour des algorithmes et de l'IA. L'argument initial de Steve Woolgar consiste à critiquer le principe de dichotomie homme/machines pour considérer les systèmes d'IA comme des "acteurs sociaux", au même titre que les humains. En mobilisant une théorie proche de celle de l'acteur-réseau, Steve Woolgar tend à considérer les machines comme des entités "actantes" participant alors à construire ou à légitimer une situation sociale. Sur ce point, il est utile de rappeler que le principe du *machine learning* est d'apprendre sur la base de données censées représenter un phénomène en vue de fournir des propositions, des résultats correspondant à un problème qu'une organisation cherche à automatiser. Il n'y a pas de système fixe. Un système d'IA déployé continuera à apprendre. Il sera continuellement amélioré par des équipes de développement incorporant encore davantage du social que ce soit via les interactions avec des utilisateurs ou bien grâce à des travailleurs des données.

Nous considérons donc que la question n'est pas de savoir s'il est possible de considérer les machines comme des acteurs sociaux, mais plutôt de savoir d'où viennent les représentations qu'elles incorporent, en s'intéressant à leurs processus de fabrication. C'est précisément, l'enjeu de la critique externe telle qu'elle est présentée par Fabian Muniesa (2019) et Jean-Sébastien Vayre (2021). L'intérêt de cette critique de "second ordre" est qu'elle s'appuie sur la nécessité d'étudier la fabrication de l'IA au travers de ce que ce processus incorpore comme travail humain, et donc comme éléments sociaux, plutôt que de s'intéresser uniquement aux effets des systèmes d'IA sur les autres acteurs (utilisateurs, consommateurs). La suite de notre développement s'appuiera sur cette distinction en proposant une revue de la littérature des études se rattachant, d'une part, à la critique interne et à la critique externe d'autre part.

10. "Pour autant, ces travaux critiques vis-à-vis des effets sociaux des algorithmes ont, toujours d'après l'auteur, tendance à laisser penser que ces technologies fonctionnent de manière autonome. Le lecteur retrouve ici l'argument du « drame algorithmique » de Malte Ziewitz (2016) : « de dispositifs ancrés empiriquement et délicatement irrigués, les algorithmes sont progressivement devenus de puissantes entités flottantes » (Jaton, 2019, p. 321). Les objectifs scientifiques et politiques de Florian Jaton sont alors exposés avec beaucoup de style. Ils peuvent être résumés ainsi : « au lieu de se limiter à l'étude des effets des algorithmes, il est maintenant temps d'enquêter sur les causes de ces effets pour mieux les changer » (Jaton, 2019, p. 316)" (Vayre, 2021)

1.1 Une critique focalisée sur la manière dont les systèmes d'IA sont mobilisés dans la société

Ce premier corpus de littérature aborde la manière dont les systèmes d'IA (SIA) sont "mobilisés" dans la société. Dans cette perspective, plusieurs recherches appuient sur le manque de transparence de l'intelligence artificielle. Par exemple, le recours aux SIA pour l'action publique, de plus en plus encouragé, est controversé, car il concerne des services critiques pour leurs conséquences sur les utilisateurs (par exemple : ceux d'attribution des aides sociales). Que ce soit dans le secteur privé ou dans le secteur public, l'utilisation des SIA concerne en effet de plus en plus de fonctions dans les organisations. Dans le recrutement, des systèmes automatisés permettent par exemple de sélectionner les candidats les plus pertinents. Sur certaines plateformes numériques de travail, des algorithmes sont utilisés pour appairer l'offre et la demande de travail, par exemple en répartissant les courses entre les chauffeurs de la plateforme Uber. Dans le secteur de la surveillance, des SIA commerciaux promettent d'aider les professionnels de la sécurité ou les policiers à identifier des comportements répréhensibles. Face à cela, de multiples questions se posent. Quels sont les effets sur les utilisateurs qui interagissent avec ces systèmes? Plus particulièrement dans le secteur public, comment exercer un contrôle citoyen sur les SIA qui sont déployés?

Une partie de la recherche s'est focalisée sur la non-objectivité des algorithmes (*biais*) et ses effets sur les utilisateurs. D'autres études ont tenté d'analyser plus spécifiquement l'utilisation de ces systèmes dans le monde du travail et notamment au sein des plateformes numériques de travail (PNT). Dans ce contexte, c'est tout un narratif centré sur l'objectivité des systèmes algorithmiques qui contrôlent les travailleurs des plateformes qui est remis en question.

Cet enjeu se double d'une critique relative au caractère opaque des algorithmes. D'abord d'un point de vue technique, à travers le recours au *machine learning* qui rend ces systèmes évolutifs via l'autoapprentissage. Ensuite, du point de vue organisationnel, à travers l'encastrement de ces systèmes dans la société et la manière dont ils sont fabriqués. Enfin, du point de vue des effets sur les utilisateurs, une dernière catégorie de questions concerne la transformation du travail à l'ère de l'automatisation. L'enjeu est alors d'apporter des éléments empiriques concrets, d'une part à ce qu'implique l'IA en matière de travail humain et d'autre part sur ce que l'introduction de ces systèmes change réellement pour les travailleurs qui interagissent avec ces systèmes dans leur quotidien.

1.1.1 Les algorithmes au travail et leurs conséquences sur leurs utilisateurs

Une partie de la littérature s'est focalisée sur les effets de l'introduction de l'intelligence artificielle dans l'utilisation quotidienne de ces systèmes par des utilisateurs (Christin, 2020b; Eubanks, 2017). L'intérêt de ces recherches est d'avoir souligné les biais associés à l'intelligence artificielle et leurs effets sur le fonc-

tionnement des services basés cette technologie. L'intégration de la technologie dans le poste de travail soulève une contradiction entre la nécessité d'accroître l'efficacité et la productivité au travail et le désir d'augmenter le bien-être des salariés au travail (Moore, Upchurch, & Whittaker, 2018). Par exemple, l'introduction des logiciels de messagerie et de gestion de projet peut faciliter le télétravail en permettant aux managers de suivre le travail réalisé par leurs employés, *in fine* de les surveiller. Ainsi, on peut distinguer entre les recherches sur la surveillance au travail et les effets des systèmes automatisés sur les utilisateurs, en l'occurrence sur les travailleurs.

Dans une des premières enquêtes sur les opérateurs de données, Shoshanna Zuboff (1988), consacre un chapitre à l'analyse de la surveillance permise par l'introduction de l'informatique. Elle y détaille la manière dont est surveillé, non seulement le contenu du travail, mais également le comportement des opérateurs dans la réalisation de leurs tâches (p.319). Dans son enquête sur les centres de traitement de données au Brésil, Angelo A. S. Soares (1991) montre la présence de dispositifs pour surveiller le nombre de frappes sur le clavier, pour suivre la productivité des travailleurs et ainsi contrôler l'attitude de l'opérateur vis-à-vis de sa tâche (par exemple : est-ce qu'à un moment l'agent a diminué son rythme de travail?). D'après Zuboff, lorsque les managers utilisent l'informatique pour documenter les décisions qu'ils prennent, l'ordinateur devient lui-même superviseur, provoquant une distanciation entre le manager et l'opérateur (p.326).

Or, cette dimension devient d'autant plus sensible lorsque les travailleurs ont aussi accès à des données concernant leur performance au travail. C'est parfois le cas sur les plateformes de travail au sein desquelles les travailleurs ont accès à un certain nombre de métriques sur la qualité de leur travail. Par exemple, les chauffeurs Uber ou les hôtes Airbnb ont accès à la notation de leurs clients et les micro-travailleurs ont accès à un tableau de bord des tâches réalisées et de leurs scores de performance (A. Casilli et al., 2019b). Dès lors que l'organisation du travail repose sur un paiement à la tâche, comme c'est le cas sur la plupart des plateformes numériques de travail, cette mesure de la performance risque d'être d'autant plus défavorable pour les travailleurs. Ces derniers devant par exemple justifier du retard dans une commande ou d'une mauvaise note, parfois attribuée sans aucune raison.

L'utilisation de métriques peut être également être employée pour déterminer l'attribution des tâches ou bien pour gérer de la force de travail. On parle de *management algorithmique* (Rosenblat & Stark, 2016). Cette nouvelle façon de gérer ces derniers implique un suivi et une évaluation constants des performances, la mise en œuvre de décisions sans intervention humaine (les travailleurs communiquant directement avec le SIA), ainsi qu'une faible transparence, notamment sur les décisions qui les concernent directement (Möhlmann & Zalmanson, 2017).

Mateescu et Nguyen (2019) expliquent que le management algorithmique se déploie dans de plus en plus de secteurs, à partir des entreprises de la *gig economy*. Il se développerait ainsi dans les secteurs en pleine

mutation vers la production "juste à temps" comme la logistique ou la grande distribution alimentaire. Les auteurs identifient plusieurs enjeux du déploiement d'algorithmes ou de SIA sur le travail : la surveillance et le contrôle des travailleurs, le manque de transparence des algorithmes, les biais et la discrimination qui peuvent être induits par les systèmes algorithmiques et la loyauté des systèmes algorithmiques. Dans le secteur des plateformes de *freelancing* et de micro-travail, Wood, Graham, Lehdonvirta, et Hjorth (2019a) identifient les conséquences de ce contrôle algorithmique. Ainsi, en plus de l'isolement et des horaires de travail irréguliers qui résultent de l'organisation du travail platformisé, les auteurs montrent qu'il résulte du contrôle par les algorithmes un manque de pouvoir de négociation de la part des travailleurs¹¹.

Ils soulignent encore que la qualité du travail fourni par les plateformes est déterminée par les compétences et la réputation des travailleurs sur la plateforme. Lorsqu'ils ont une note faible ou qu'ils n'affichent pas les bons savoir-faire sur leurs profils, les travailleurs risquent d'autant plus de faire face à l'insécurité au travail et à des salaires bas. Néanmoins, on peut interroger la tendance des auteurs de l'article à surestimer le pouvoir des algorithmes, au détriment des choix des plateformes tels que la définition des conditions de travail ou de leurs conditions générales d'utilisation. Par exemple, le fait qu'un client ne soit pas obligé de payer une mission est un choix délibéré de la plateforme et non lié à la mise en place d'un algorithme. En d'autres termes, les algorithmes ne sont aussi et surtout que l'expression de décision managériale, qu'ils complètent pour faciliter le contrôle sur les travailleurs.

Derrière le management algorithmique se pose aussi la question de la circulation des données dans les organisations ou, autrement dit, qui a accès à quoi? Par exemple, les chauffeurs Uber ont accès à un certain nombre de métriques, mais l'application en enregistre en fait bien plus, comme le nombre de freinages et les plaintes des consommateurs au service-client. Le traitement de ces données peut conduire à des décisions radicales telles que désactiver un compte, ce qui revient en pratique à mettre fin à l'activité du chauffeur. L'organisation du travail repose alors, au moins, en partie, sur l'illusion de la maîtrise de leur travail par les travailleurs des plateformes : ils sont indépendants et ils ont des retours sur leur performance via l'application. En revanche, l'application collecte énormément de données supplémentaires dont l'utilisation n'est pas clairement indiquée aux travailleurs : est-ce que cela sert à créer des données pour entraîner une voiture autonome? Est-ce que cela sert à produire des notes "internes" à Uber donnant lieu à des désactivations de comptes?¹²

11. "L'autonomie résultant du contrôle algorithmique peut conduire au surmenage, au manque de sommeil et à l'épuisement en raison du faible pouvoir structurel des travailleurs vis-à-vis des clients. Ce faible pouvoir structurel est le résultat de systèmes d'évaluation et de classement basés sur des plateformes qui permettent une forme de contrôle capable de surmonter les barrières spatiales et temporelles que la non-proximité impose à l'efficacité de la surveillance et de la supervision directes du processus de travail." (Wood et al., 2019a)

12. Le livre de Judith Duportail (2020) sur l'application de rencontre Tinder raconte bien cette asymétrie d'information entre l'application et l'utilisateur, ces derniers étant notés par l'entreprise sur beaucoup plus de "variables" que ce qu'ils renseignent sur l'application. Sur la base de la manière d'écrire, Tinder créait par exemple une variable sur le niveau d'éducation qui rentrait en compte dans la note attribuée à l'utilisateur et sur la base de laquelle il se voyait proposer d'autres utilisateurs.

La question de la circulation des données est également problématique en ce qu'elle redistribue la responsabilité entre les acteurs. Shoshanna Zuboff (1988 : 297) montre ainsi que certains opérateurs ont accès à beaucoup plus de données que celles qui concernent leur propre travail (par exemple : les performances de leurs collègues). Si le travailleur repère une malfaçon, une erreur en regardant ces données, peut-il faire comme s'il ne l'avait pas vue? Non, indique Shoshanna Zuboff. De ce point de vue, l'introduction de ces métriques opère aussi un transfert de responsabilité des managers vers les opérateurs qui sont ainsi incités à se surveiller mutuellement. Dans le cadre de notre enquête sur les micro-travailleurs français (A. Casilli et al., 2019a), nous avons observé que le forum pouvait être utilisé par les salariés de la plateforme ou par certains micro-travailleurs pour indiquer des malfaçons dans la réalisation de certaines micro-tâches. Dans ce cas, certains travailleurs pouvaient se plaindre des autres micro-travailleurs, et le cas échéant donner des conseils sur la manière de faire. Le travailleur devient alors superviseur de lui-même, voire des autres travailleurs, alors même qu'il n'a pas forcément la maîtrise de l'ensemble de l'organisation du travail, créant de ce fait une décentralisation des décisions managériales qui ne sont désormais plus l'apanage de l'encadrement intermédiaire. En ce sens, le problème du management algorithmique ne réside pas uniquement dans le contrôle ou la surveillance des travailleurs, mais dans le fait que les décisions managériales deviennent plus diffuses et moins aisément identifiables, diluant les responsabilités. Ainsi, au-delà des conséquences négatives directes qu'ont ces systèmes de contrôle automatisés sur le travail, l'enjeu reste bien de comprendre la chaîne de décision qui fait fonctionner le système et non uniquement ses effets. On retrouve ici la critique de Florian (Jaton, 2019) sur la nécessité de comprendre la cause des effets des algorithmes sur les utilisateurs et pas non uniquement d'analyser les conséquences.

En dehors de la question des effets du management algorithmique, le travail d'Angèle Christin (2020b) met en exergue l'agentivité des professionnels du droit et de l'information (journalistes) dans la manière d'interagir avec des systèmes algorithmiques. Elle met en lumière les critiques, les résistances et la variété de ces dernières selon la profession concernée. L'intérêt principal de cet article est de proposer une analyse "en pratique" des algorithmes en montrant le décalage entre ce que l'algorithme est censé faire et la manière dont les travailleurs interagissent avec lui. D'une certaine manière, on retrouve, en creux des formes de résistance ou de "braconnage" identifiés par Rosenblat et Stark (2016) à propos de l'utilisation des forums de chauffeur Uber, par des travailleurs cherchant à connaître des astuces pour optimiser leur travail au regard du fonctionnement de l'algorithme.

Finalement, au-delà de l'étude de ce que changent les algorithmes dans le travail, il est nécessaire de comprendre les différentes catégories d'acteurs qui interagissent avec ces derniers et de quelles manières. Pourquoi tel ou tel choix de conception de l'algorithme d'appariement sur l'application de Tinder ou d'Uber? Comment les travailleurs-usagers tentent-ils de détourner ces dispositifs? Quelle est la chaîne de décisions qui conduit à la production d'un service basé sur un système automatisé? Autrement dit, au-delà des al-

algorithmes, dans quelle organisation ces derniers s'insèrent-ils et quelles sont les décisions qui se cachent derrière l'apparente objectivité algorithmique? ¹³

1.1.2 L'opacité des systèmes d'IA comme défi pour les sciences sociales

L'analyse des algorithmes s'est rapidement heurtée à leur opacité. Franck Pasquale (2015) utilise d'ailleurs le terme de "black box" pour désigner les algorithmes et la société dans laquelle ils sont déployés. Le juriste amorce une critique du développement des systèmes automatisés qui se base sur deux points clés : le système est opaque d'un point de vue technique et politique et il crée des déséquilibres entre, d'un côté, les citoyens et de l'autre les entreprises privées et les états. Partant de ce principe, il propose une analyse de plusieurs secteurs, notamment la sécurité, la finance et l'accès à l'information et propose des pistes de solutions.

Cette critique de l'opacité de la société est basée sur deux constats. Tout d'abord, le fait que de nombreuses décisions "qui jusqu'alors étaient basées sur la réflexion humaine se prennent à présent de manière automatique" (p.19). Ensuite, s'appuyant sur ce qu'il observe du recours croissant aux *big data*, il considère que de "grandes entreprises détiennent une connaissance sans précédent des moindres détails de nos vies, alors que nous savons peu, voire rien du tout, de l'usage qu'elles font de cette connaissance pour influencer leurs décisions - et par conséquent les nôtres" (p.21). Opacité donc, mais également inégalité dans la capacité des citoyens à accéder à leurs informations et à la manière dont ces dernières sont traitées par les grandes entreprises. Par ailleurs, il analyse les algorithmes en tant que *faisant partie* d'une société qui facilite le secret et renforce cette opacité. Il cite pour cela l'exemple du secret des affaires ou bien de l'*obfuscation*, technique par laquelle des entreprises accusées de mauvaises pratiques noient juges et citoyens sous des millions de documents, sans répondre aux accusations. Par conséquent, cette opacité est donc technique, mais aussi juridique et sociale.

Un autre aspect du livre de Franck Pasquale est de souligner l'affaïssement de la frontière entre le secteur public et privé en pointant du doigt les collaborations entre l'administration américaine et les grandes entreprises technologiques. Par exemple, avec les "*fusions centers*", des centres de données et d'analyse de données dont la raison d'être est de faciliter à l'accès aux données publiées sur les réseaux sociaux au gouvernement américain. Bien que la configuration ne soit pas exactement la même, l'exemple du "moissonnage" de données des réseaux sociaux pratiqué par le fisc français illustre à quel point les États s'appuient sur les entreprises technologiques pour assurer leur mission de service public ¹⁴.

Franck Pasquale explore également les effets concrets des algorithmes sur les utilisateurs, par exemple

13. "Les instances publiques et privées qui se servent aujourd'hui de prétextes algorithmiques pour gouverner les conduites humaines participent de cette énorme imposture. Il n'y a pas d'algorithme, il n'y a que la décision de quelqu'un d'autre." (A. A. Casilli, 2017c)

14. Voir cet [article de Next Impact](#)

lorsqu'ils sont utilisés pour l'attribution de prêts immobiliers ou pour le calcul de taux d'assurances. Il montre que ces dispositifs contribuent à renforcer les inégalités entre les entreprises et les citoyens à cause de l'asymétrie d'information créée par le recours aux algorithmes (Rosenblat & Stark, 2016). À travers ces inégalités, c'est bien un pouvoir invisible qui s'exerce sur les citoyens. Selon lui, *"le nouveau principe en vigueur est : moins les auteurs de spams, les pirates informatiques, les tricheurs (...) et le public en général en savent (au sujet des algorithmes), mieux c'est. La transparence a été remplacée par un secret en béton, à la fois réel et juridique. La question de la légitimation a été ajournée"* (p.283).

Pour remédier à ces problèmes, l'auteur propose deux pistes. Tout d'abord, d'ajouter un principe de respect des principes de loyauté et de non-discrimination dans les traitements algorithmiques. Ensuite, de permettre aux citoyens de mieux connaître ces systèmes en contraignant les organisations à des obligations de transparence des conditions de collecte et d'utilisation des données personnelles. Proposant une méthodologie d'audit des algorithmes, Diakopoulos (2015) rejoint les pistes de Franck Pasquale en y abordant également d'autres solutions éventuelles, comme la formation des citoyens et des experts à l'audit algorithmique et à davantage de transparence du code source des algorithmes.

La question de l'opacité a également fait l'objet d'un article de Jenna Burrell (2016), qui actualise le concept au regard des développements liés au *machine learning*. Elle y propose une typologie des différentes formes d'opacité qui touchent les systèmes d'IA. Elle reprend tout d'abord l'argument de Franck Pasquale en définissant l'opacité comme *"secret intentionnel des états et des entreprises"* à travers le secret des affaires notamment. Une des réponses à cette première forme d'opacité a été de rendre public le code source des algorithmes. C'est par exemple le cas en France, où, depuis le décret n°2005-1309 du 20 octobre 2005 - art. 90, tout citoyen a le droit de demander *« les informations permettant de connaître et de contester la logique qui sous-tend le traitement automatisé en cas de décision prise sur le fondement de celui-ci et produisant des effets juridiques à l'égard de l'intéressé »*. La loi du 7 octobre 2016 pour une République numérique puis la loi du 20 juin 2018 relative à la protection des données viennent proposer un cadre pour la communication des informations relatives aux systèmes automatisés. Elle dispose notamment que tout citoyen peut solliciter la publication du code source. Toutefois, ce principe d'ouverture peut se heurter au manque de littératie technique des citoyens, deuxième forme d'opacité soulignée par Jenna Burrell.

Une dernière forme d'opacité est plus spécifique aux modalités de fabrication des modèles basés sur le *machine learning* et désigne *"une opacité qui découle des caractéristiques des algorithmes d'apprentissage automatique et de l'échelle requise pour les appliquer utilement."* qui s'appuie sur l'exemple des algorithmes du moteur de recherche Google, qui intègre de nombreux composants. Par conséquent, même les ingénieurs qui doivent travailler sur ces modèles n'ont pas forcément conscience de toutes les caractéristiques de ce qu'ils produisent. Cette dimension crée également une forme d'opacité du point de vue de l'organisation du travail. Or, nous dit-elle, cet aspect est particulièrement prégnant dans le cas du *machine learning*, car

ils intègrent de nombreuses données et de nombreux paramètres. Cette caractéristique rend même illusoire pour les concepteurs de ses systèmes d'interpréter correctement les résultats des modèles ¹⁵.

Là encore, la publication du code source des modèles constitue une solution partielle, tout comme les propositions relevant des audits algorithmiques, car *"le raisonnement, le débogage ou l'amélioration de l'algorithme deviennent plus difficiles avec davantage de caractéristiques fournies en entrée, chacune modifiant subtilement et imperceptiblement la classification résultante."* L'intérêt de l'article de Jenna Burrell est d'élargir le concept d'opacité, notamment aux enjeux liés à l'échelle des algorithmes et de ce qu'implique leur fabrication dans le cas du *machine learning*. En creux et même si elle l'aborde plutôt sous l'angle technique, il est possible de retrouver partiellement un des arguments soutenant l'orientation de notre thèse : la compréhension de l'IA passe par l'étude du processus de production de ces systèmes.

Face à cette nécessité, on a vu l'émergence de discussions sur la manière de mettre en place des méthodes ethnographiques susceptibles de contribuer au dévoilement des algorithmes (Christin, 2020a; Seaver, 2017). Nick Seaver (2017) réinscrit ces objets techniques dans les pratiques qui les activent : *"Comme d'autres aspects de la culture, les algorithmes sont mis en œuvre par des pratiques qui ne tiennent pas compte d'une forte distinction entre les préoccupations techniques et non techniques, mais qui les mélangent plutôt"*. Il propose ainsi de dépasser la question de la technique, en considérant les algorithmes comme des objets instables, façonnés en permanence par leurs interactions avec les acteurs ¹⁶.

Sur cette base, lui comme Angèle Christin proposent des "tactiques" (ou stratégies d'enquête) basées sur les méthodes ethnographiques. En réponse à l'opacité qui naît de la complexité de ces systèmes, Nick Seaver propose par exemple de faire de l'ethnographie "multi-située" pour suivre les algorithmes là où ils se fabriquent et se déploient. Il préconise également d'avoir une attention très spécifique à la texture du terrain, c'est-à-dire au degré de résistance des personnes étudiées selon les contextes. Angèle Christin propose quant à elle trois types de méthodes : la *refraction algorithmique*, qui consiste à analyser les contextes de déploiement des algorithmes, la *comparaison algorithmique* qui consiste à pallier à l'opacité des algorithmes en comparant plusieurs cas d'études, et enfin, la *triangulation* qui consiste à utiliser les méthodes computationnelles pour comprendre le fonctionnement des algorithmes.

Enfin, Bechmann et Bowker (2019) décrivent les différentes étapes de fabrication d'un modèle d'IA. À l'instar d'autres articles en sciences sociales focalisés sur l'IA, ils ne soulignent pas clairement ce qui relève du travail humain et qui se cache derrière les algorithmes. Dans leur article, ils invitent d'ailleurs à dépasser

15. *"Avec des ressources informatiques plus importantes et de nombreux téraoctets de données à exploiter (aujourd'hui souvent collectées de manière opportuniste à partir des traces numériques des activités des utilisateurs), le nombre de caractéristiques possibles à inclure dans un classificateur augmente rapidement, bien au-delà de ce qui peut être facilement appréhendé par un être humain doué de raison."* (Burrell, 2016)

16. *"(...) c'est une façon utile d'aborder les algorithmes - non pas comme des objets stables avec lesquels on interagit de plusieurs points de vue, mais comme les conséquences multiples d'une variété de pratiques humaines."* (Seaver, 2017)

ser le cadre de la critique technique pour se focaliser sur les choix humains qui sont effectués pendant la conception. Cette littérature nous pousse finalement à délaisser l'analyse des choix techniques pour identifier la chaîne de décisions qui façonnent l'usage algorithmique, depuis le choix de la tâche à automatiser jusqu'à l'utilisation du système, en passant par les phases de conception incluant tout le travail des données. Comme l'expliquent [Bechmann et Bowker \(2019\)](#) : *"nous devons moins nous concentrer sur l'accès aux modèles et aux algorithmes en tant que constructions techniques en soi que sur la documentation des choix humains effectués dans le processus de travail entourant l'IA afin d'agir de manière durable par rapport aux valeurs démocratiques partagées, en évitant le masquage, les préjugés discriminatoires, et la discrimination des électeurs en particulier."*

* * *

Dans cette première section, nous sommes revenus sur la manière dont les algorithmes sont mobilisés dans la société et sur l'opacité de leur fonctionnement. Une importante littérature s'est développée sur l'impact des algorithmes au travail et notamment à travers les plateformes numériques de travail. Pour nous, ces recherches montrent le décalage entre les objectifs et l'usage des algorithmes illustrant la nécessité d'élargir le regard aux utilisateurs. La question de l'opacité a quant à elle fait la part belle à des solutions d'audits techniques et de transparence pour limiter les risques d'impacts négatifs. Pourtant, au-delà de la question des méthodes d'analyses, nous pensons qu'il s'agit surtout de "délaisser" l'audit technique par des approches qui permettent de multiplier les regards et les analyses en s'intéressant aux acteurs qui participent à la fabrication des modèles d'IA. Dans cette thèse, nous souhaitons approfondir l'hypothèse de l'*opacité organisationnelle* en montrant la manière dont se structure la fabrication des modèles d'IA et surtout en analysant les décisions de conception qui façonnent la forme finale du modèle.

1.2 Une critique qui consiste à examiner les biais de l'IA comme expression des représentations de ceux qui la fabrique

Un autre pan de la littérature consacrée aux algorithmes et aux SIA se focalise sur la question des représentations sociales et politiques inscrites dans ces systèmes. C'est par exemple le cas lorsque les biais de certains concepteurs de systèmes contribuent à en orienter le fonctionnement. Les questions qui se posent sont de deux types. Il s'agit premièrement d'identifier les effets de ces systèmes lors qu'ils pénètrent certains secteurs essentiels, tels que le recrutement ou l'attribution d'aides sociales; et deuxièmement, de tenter de comprendre comment et où résident ces biais dans le processus de production. Quels sont les biais des systèmes d'IA? Comment émergent-ils lors du processus de fabrication? Comment limiter les impacts de ces biais?

Cette thématique a eu un écho important dans la littérature à la fois en sciences sociales et en sciences

informatiques. En sociologie, l'enjeu principal a été de rendre visible les biais (liés à l'origine ou au genre...) présents dans certains modèles, et à comprendre les prénotions qui s'inscrivent dans des systèmes prétendument objectifs. En *computer science*, certaines recherches ont tenté de tenir compte de ces enjeux en proposant des approches susceptibles de dé-biaiser les modèles. La question de l'explicabilité des SIA est devenue de plus en plus prégnante à mesure que la recherche en *machine learning* produisait des systèmes utilisant de plus en plus de données, incorporant même plusieurs techniques d'IA dans des modèles aux millions de paramètres : on parle de "Large-Scale AI models"¹⁷.

C'est aussi au travers de l'explicabilité que s'est posée la question des biais et *in fine* de la non-neutralité des modèles. À l'intersection des deux disciplines, s'est ainsi développé le concept de *data justice* qui s'appuie sur des recherches ayant montré la faible prise en compte de certaines catégories de population minorisées dans le processus de fabrication des modèles. Derrière ce concept, il s'agit de renforcer la création de jeu de données représentatives de l'ensemble de la population. C'est le cas de certains systèmes de reconnaissance faciale dont il a été prouvé qu'ils ne détectent pas correctement les personnes noires.

Dans une première section, nous présenterons ces recherches, puis nous attarderons sur le pan de la littérature à l'intersection des STS et de la sociologie de la quantification qui a particulièrement bien montré les processus de négociation à l'oeuvre derrière de nombreux systèmes de quantification, algorithmiques ou non comme le Produit Intérieur Brut et les "jeux d'acteurs" qui façonnent ces dispositifs. Nous terminerons cette partie en faisant le lien entre le micro-travail et la question des biais si présente dans ce champ.

1.2.1 La non-objectivité des systèmes algorithmiques et ses conséquences sur les utilisateurs

Dans un [rapport](#), le GPAI analyse plusieurs systèmes d'IA considérés comme nocifs, car ils tiennent insuffisamment compte des populations minorisées. C'est dans l'optique de dénoncer et résoudre ce type de biais qu'a émergé le concept de *data justice* qui propose de "rassembler une critique des dynamiques de pouvoir qui sous-tendent une accélération de la tendance à la datafication avec un engagement normatif envers les principes de la justice sociale - un engagement pour la réalisation d'une société qui soit équitable, juste et capable de s'attaquer aux causes profondes de l'injustice."¹⁸ Elle repose sur le constat d'un développement inéquitable du big data et de l'IA, ces technologies laissant de côté l'idéal d'une société plus égalitaire, en se construisant sur le fondement de constructions sociales ancrées profondément dans l'histoire. On peut par exemple citer les biais raciaux dans le cas de systèmes de reconnaissance faciale.

Parmi les nombreux exemples cités par le GPAI, certains d'entre eux permettent de mieux comprendre l'éten-

17. C'est le cas de deux modèles largement médiatisés : GPT3 et Dalle-2 développés par l'entreprise OpenAI, fondée notamment par Elon Musk. Le premier est un modèle linguistique qui permet de faire de la traduction et de générer des textes complexes comme des articles de presse. Le second permet de générer des images sur la base de textes

18. Traduction de l'auteur

due du problème. Le "gaydar" automatique est un premier exemple de système, qui a été popularisé dans la littérature à travers le travail de Wang et Kosinski (2017), un chercheur américain en psychologie qui a tenté de créer un système de détection de l'homosexualité basé sur les traits du visage. Souhaitant initialement questionner les risques de ces systèmes en matière de vie privée, son article a finalement déclenché d'importantes controverses qui se sont concentrées sur sa méthodologie et le cadrage théorique de son travail. Nicolas Baya-Laffite, Beaudé, et Garrigues (2018) revient notamment sur le fait que le chercheur a mobilisé des théories de l'homosexualité, inspiré de l'évolutionnisme social, sujettes à de fortes controverses scientifiques, afin de justifier les résultats de son algorithme de classification, comme le fait que son système associe des joues plus larges de femmes à l'homosexualité. De plus, son modèle n'utilisait que des photos de personnes caucasiennes. Malgré ces failles considérables, le principe de détection automatique de l'orientation sexuelle a été repris et est vendu par des entreprises comme Amazon ou IBM, avec des risques évidents pour les personnes concernées.

L'usage des technologies de reconnaissance faciale aux frontières constitue un deuxième cas. Déployé dans les aéroports français sous le nom "Paraphe", ce type d'usage vise à classer les personnes autorisées à passer ou non une frontière. Toutefois, il a été prouvé que ces systèmes reconnaissent de manière moins performante les personnes de couleur et les personnes transgenre. Enfin, l'usage de SIA dans la police, à travers le déploiement de modèles de prédiction des violences notamment aux États-Unis (Benbouzid, 2018), soulève aussi des questions. Parmi ces dernières, figure le fait qu'ils se basent sur l'historique de données des crimes aux États-Unis qui intègre souvent la période de ségrégation raciale. En conséquence, ces systèmes ont tendance à attribuer des scores de risque plus importants pour les personnes de couleur. L'association ProPublica a ainsi testé un algorithme utilisé dans l'attribution de scores de risques assignés à des personnes arrêtées dans le comté de Broward en Floride. Cet audit montre que le système est très peu précis, puisque seulement 20% des personnes prédites comme susceptibles de commettre un crime violent l'ont fait. En plus de ce manque de précision, le système avait tendance à beaucoup plus identifier les personnes noires comme de futurs criminels, presque deux fois plus que des personnes blanches. Ces dernières se voyant attribuer un score beaucoup plus faible en moyenne.

Au fil de la multiplication des controverses, le concept de "data justice" a été directement repris par des autorités publiques, à l'instar de la Charte internationale de l'UNESCO. Il y est ainsi précisé que les jeux de données doivent représenter au mieux la diversité humaine en recommandant "l'élaboration de normes pour l'annotation des ensembles de données devrait être encouragée, y compris la ventilation des données en fonction du genre et d'autres priorités". Il y est également indiqué l'importance de la formation des chercheurs en IA à l'éthique et plus spécifiquement à la « façon dont ces ensembles sont annotés » (p.36). C'est également dans cette perspective que la Commission européenne, dans sa proposition de régulation relative aux SIA, propose un mécanisme de déclaration de la fabrication et du fonctionnement des SIA pour les systèmes à

haut risque.

Au-delà des effets des biais, la littérature en sciences sociales consacrée aux algorithmes s'est beaucoup intéressée aux valeurs des concepteurs ou au contexte culturel inscrits et reproduits par les systèmes automatisés. Différents travaux ont remis en question l'objectivité de ces systèmes, en montrant que les systèmes de recommandations de sites comme Twitter reflètent des décisions basées sur les échelles de valeurs de ses utilisateurs. Plus récemment, une recherche conduite par des chercheurs de Twitter (Huszár et al., 2022) a montré que le réseau social a tendance à montrer des contenus plutôt considérés comme à droite du spectre politique, confirmant les résultats de chercheurs comme (Schradié, 2019)¹⁹.

C'est aussi l'enjeu des travaux de Noble (2018), qui ont montré que le moteur de recherche de Google met en avant des résultats pouvant être qualifiés de sexistes ou racistes. Elle prend en exemple le cas de l'algorithme d'auto-complétion qui présentait systématiquement le terme "guenon" à côté de "noirs". Noble a montré qu'à la fois le contexte culturel et les positions des employés de Google ont affecté leurs décisions sur ce qui devait être classé comme raciste ou sexiste par le moteur de recherche. Elle rappelle ainsi que la technologie peut reproduire des différences de traitement entre différentes catégories de population. Bien loin de l'objectivité de l'algorithme de recherche de Google, les inégalités sont intégrées et amplifiées par le moteur de recherche. Cette dimension est également évoquée par K. A. Joyce, Darfler, George, Ludwig, et Unsworth (2018), qui constatent que les ingénieurs en *machine learning* ne sont pas suffisamment formés aux enjeux sociaux et éthiques de la technologie.

De la même manière, Virginia Eubanks compare plusieurs systèmes automatisés utilisés dans différents services publics aux États-Unis. À travers une étude ethnographique, Eubanks décortique la façon dont ces derniers sont fabriqués et les effets qu'ils produisent sur les utilisateurs. Elle les qualifie d'instruments de gouvernements, dans la mesure où les algorithmes ont des impacts sur la manière dont les aides sociales sont distribuées (EUBANKS, 2018), ou sur la manière dont les enseignants obtiennent un poste (O'Neil, 2016). Les impacts des systèmes automatisés ont donné lieu à de nombreuses controverses sur les biais algorithmiques. L'idée que l'objectivité d'un algorithme pouvait donner lieu à de meilleures décisions que celles prises par des êtres humains s'est confrontée, dans la réalité, à deux écueils. D'une part, les performances des algorithmes sont parfois limitées par le choix des variables et des données utilisées, d'autre part leur développement repose à chaque étape sur du travail humain (EUBANKS (2018)²⁰.

19. Une partie des explications résideraient dans les capacités (notamment de coordination) et les formes de la mobilisation des groupes de droite et d'extrême-droite qui fonctionneraient parfaitement avec le mode de tri de l'information des réseaux sociaux

20. Elle détaille ainsi l'influence humaine lors des différentes étapes de création d'un algorithme dans le cadre d'un algorithme d'évaluation de la maltraitance infantile "Mais le biais humain est également une caractéristique intrinsèque du modèle de risque prédictif. Les variables de résultats sont des approximations des dommages subis par les enfants; elles ne reflètent pas la négligence et la maltraitance réelles. Le choix des variables de substitution, et même le choix d'utiliser des variables de substitution, est le fruit d'une appréciation humaine. Les variables prédictives sont tirées d'un univers limité de données qui ne comprend que des informations

En résumé, le concept de *data justice* englobe deux dimensions complémentaires. La première met en lumière les effets négatifs des SIA sur les usagers lorsque ces derniers ces systèmes sont conçus de manière biaisée. La seconde met l'accent sur les solutions, notamment en termes de formation des ingénieurs et de représentativité des jeux de données. La littérature en sciences sociales a permis d'identifier certaines sources de ces biais, notamment ceux associés aux concepteurs de systèmes automatisés.

Ce faisant, ces recherches ont posé un jalon important : les SIA sont constitués de décisions humaines. Toutefois, cette littérature se focalise surtout sur le travail de conception des ingénieurs qui développent ces systèmes et ne permet donc pas de comprendre l'ensemble des décisions prises durant le développement, y compris durant le déploiement de ces systèmes, malgré l'importance de cette phase pour les modèles de *machine learning*. Les "interventions manuelles" derrière ces systèmes, y compris par des travailleurs situés dans les pays en voie de développement, demeurent dans l'ensemble peu étudiées par cette littérature.

1.2.2 Les négociations des acteurs à l'œuvre derrière la fabrique des dispositifs de quantification

L'IA constitue un ensemble de techniques qui s'appuient sur l'informatique et sur les statistiques. Or, un pan de la recherche en sciences sociales s'est intéressé à la genèse des chiffres et des métriques, en bref des outils de quantification. Comment construit-on la mesure du produit intérieur brut (PIB)? Comment sont fabriqués les indicateurs devant servir à représenter un phénomène social ou économique? Pensée comme fait social, la quantification est définie par Henneguelle et Jatteau (2021) comme "*l'étude des activités sociales de construction, d'élaboration et d'utilisation des chiffres, quels qu'ils soient*". Dès lors que les chiffres sont utilisés comme preuve scientifique ou comme outil d'aide à la décision pour les gouvernements, plusieurs enjeux émergent, car ils deviennent dès lors des objets sociaux à part entière. On peut distinguer plusieurs séries de questionnements, telles que ceux portant sur la construction de ces indicateurs : comment se met-on d'accord sur la mesure d'un phénomène? Qui participe à la définition des indicateurs? Une deuxième série concerne la circulation et l'utilisation de ces chiffres : comment les citoyens s'en saisissent-ils et les utilisent ou non? Comment les médias rendent-ils compte de ces indicateurs? Comment les pouvoirs publics les utilisent-ils pour justifier de certaines politiques? Parce qu'ils s'appuient sur des méthodes de recueil de données et visent également à rendre calculables des phénomènes, les dispositifs algorithmiques s'inscrivent dans la lignée de ces objets sociaux, amenant à voir dans le développement de la littérature sur ces objets un élargissement du programme initial de la sociologie de la quantification.

Dans un ouvrage fondateur, Alain Desrosières (2010) montre que le processus de quantification passe par plusieurs opérations de classification et de codage. Autrement dit, "*quantifier, c'est convenir puis mesurer*", ce qui met en lumière les négociations qui conduisent les différents acteurs producteurs de chiffres à se mettre

sur les ressources publiques (...) Les données de validation du modèle sont un enregistrement des décisions prises par des travailleurs sociaux, des enquêteurs et des juges humains, portant toutes les traces de leur humanité." (EUBANKS, 2018)

d'accord sur le phénomène à étudier, comme la définition du chômage ou de la croissance économique. C'est précisément dans ce qu'il mesure (et ce qu'il ne mesure pas) que le PIB a été critiqué, par exemple en oubliant la dimension écologique ou le bonheur des sociétés. Desrosières formule ainsi une règle d'autant plus pertinente à l'heure du big data et de l'IA : *"Les données ne sont pas données."* Ainsi, il refuse de traiter les chiffres comme une matière première neutre et disponible pour différents usages ; bien au contraire, il considère que leur utilisation n'est pertinente que si elle s'accompagne d'une réflexion sur les conditions sociales et techniques de leur production. Dans un ouvrage ultérieur, Desrosières incite dès lors à : *"explorer les formes de négociation, d'équivalence, de controverse et de routinisation impliquées dans l'invention, l'utilisation et la disparition éventuelle de ces nouvelles manières de dire, de faire et de gouverner"* (Desrosières, 2013). Ce projet de sociologie des quantifications vise désormais à décoder les révolutions numériques en intégrant le big data ou l'IA. Par exemple, Emmanuel Didier dirige un programme de recherche axé sur l'étude de la trajectoire des nombres dans la société et comment cette circulation transforme les rapports de force et la société.

Fabian Muniesa (2003) a, quant à lui, étudié les "configurations algorithmiques" à travers le prisme de l'automatisation de la cotation des prix à la Bourse de Paris dans les années 1990. Par ce travail ethnographique, il a analysé la circulation et la modification des conceptions à l'œuvre derrière ce dispositif technique, tout au long du processus de conception, en identifiant trois notions : *l'explicitation*, la *configuration algorithmique* et *l'effet de justesse*. Concernant l'explicitation, Muniesa affirme que chaque : *"nouvel élément du dispositif représente un compromis entre plusieurs points de vue, entre plusieurs intérêts. Chacune des parties concernées peut ainsi se reconnaître bien, peu, mal ou pas du tout dans la forme que prend le dispositif marchand à un moment donné."* Reprenant un argument classique de la sociologie économique, il développe l'idée qu'un "marché" constitue à la fois un ensemble de règles et une configuration sociale. Le concept de "configuration algorithmique" est construit sur cette base, englobant ces deux dimensions. Pour lui, ces *"règles ne sont pas abstraites : elles sont localisées. Le concept de configuration algorithmique insiste bien sur la matérialité de l'espace de calcul, sur son emplacement au sens littéral (...) ces règles sont surtout des règles de production de signification."* (Muniesa, 2003).

Les effets de justesse rendent compte de la légitimité de la configuration algorithmique : *"le prix (tel que calculé algorithmiquement) doit être collectivement acceptable"*. À travers les algorithmes d'automatisation de la fixation des prix, s'exprime une vision du marché qui tend certes à se rapprocher de l'idéal walrasien, mais qui tient compte du point de vue des acteurs sur le mode de la controverse pour aboutir à la solution la plus consensuelle possible afin que le mécanisme soit le plus largement accepté. La finance a d'ailleurs été un champ d'études fertile pour les algorithmes, car ce secteur a mis en œuvre ce type de technologie très tôt : dès les années 1970 aux États-Unis et dans les années 1980 et 1990 en Europe. De plus, l'innovation dans ce domaine a été importante. Par exemple, les algorithmes de *trading à haute fréquence* sont désor-

mais massivement employés dans la finance, comme le montre [D. A. MacKenzie \(2021\)](#). En étudiant ces algorithmes, ce dernier montre ainsi qu'ils façonnent l'univers dans lequel ils opèrent. En effet, les places de marché se sont adaptées à ces nouveaux outils de diverses manières, en créant les systèmes d'information qui permettent aux entreprises d'avoir accès aux données nécessaires à la prédiction ou encore en mettant en place la connectivité nécessaire au fonctionnement de ces algorithmes²¹. Le cas des algorithmes de fixation des prix a aussi été étudié dans d'autres secteurs. Jean [Finez \(2014\)](#) s'est par exemple intéressé aux algorithmes de tarification à la Société Nationale des Chemins de Fer (SNCF). Il montre l'évolution de la méthode de tarification, d'abord fondée sur le nombre de kilomètres, puis sur les variations de la demande (autrement appelé le *yield management*). Dans ce cas, SNCF fait le pari que cette évolution est cohérente avec les aspirations des voyageurs, qui seraient de parfait *homo economicus* capable d'agir rationnellement pour obtenir les prix les plus cohérents avec ce qu'ils sont prêts à payer. D'autre part, cette transformation correspond au passage d'un modèle très dirigiste, mais transparent puisque le coût pour le client est basé sur le coût d'exploitation, à un modèle où les variables d'ajustement du prix ne sont pas connues du client (boîte noire).

Que dit cette orientation de la recherche sur l'avènement de la "société des algorithmes" [Burrell et Fourcade \(2021\)](#)? Ce champ d'études autour des *quantifications*, situé à l'intersection de la sociologie économique, des STS et de l'économie des conventions, s'est emparé d'objets techniques à visée calculatoires ou quantificateurs (statistiques ou de fixation des prix) qui sont désormais de plus en plus informatisés. Bien que certaines de ces études précèdent la vague du big data et de l'intelligence artificielle, elles offrent un regard très pertinent sur la mise en œuvre de ces "dispositifs socio-techniques", montrant qu'ils reposent sur un travail de négociation et de convention entre les acteurs.

Cette perspective permet également de révéler des visions du monde sous-jacentes, parfois fondées sur des modèles ou des théories économiques : le modèle walrasien de la Bourse de Paris étudié par Fabien Muniesa ou encore la théorie du prix d'utilité appliquée à la SNCF étudiée par Jean Finez. Les perspectives de recherche sont donc nombreuses. Tout d'abord, il y a une réflexion sur le regard que l'on peut porter sur ces objets complexes, le regard de l'historien, qui permet de comprendre l'historique des jeux d'acteurs et les raisons de la mise en place d'un outil, mais aussi le regard de l'ethnographe, qui cherche à comprendre quelle représentation du monde se cache derrière la technique. On retrouve également l'idée d'une "collectivisation" de la décision algorithmique, avec des enjeux de pouvoir s'inscrivant dans les modalités de fonctionnement de l'algorithme. En effet, ce dernier n'est jamais la décision d'une seule personne, ou le

21. "Là aussi, il existe des poches de structure prédictive, dont certaines sont le résultat de processus politico-économiques. La question de savoir qui ou quoi a accès aux données nécessaires pour identifier les poches potentiellement rentables relève de l'économie politique, tout comme la capacité à réaliser ces profits (dans certains domaines, cette capacité est limitée par l'exercice flagrant du pouvoir de marché). Lorsque l'accès aux données, les contraintes et les possibilités d'échange changent, l'équilibre des pouvoirs entre les acteurs institutionnels est menacé". [D. A. MacKenzie, 2021](#)

fruit d'un développeur. Elle est systématiquement le résultat d'un tâtonnement collectif, qui laissera plus ou moins de place à certains acteurs et dont l'algorithme sera la matérialisation.

Dans le champ des études consacrées aux données et à leur production, plusieurs recherches se sont saisies de la question de leur construction et des représentations inscrites dans les bases de données. Rob [Kitchin \(2014\)](#) traite à la fois de la question de la production des données : qui produit les données et quelles en sont les implications ? Mais aussi de leur utilisation : quelles sont les conséquences sur l'analyse de la société et même sur les outils qui permettent de trier les individus ? Quelles sont les conséquences sur la vie privée ? En définitive, la question est de savoir ce qu'il y a de politique dans des données et quelles en sont les conséquences pour la société.

Du côté de la production de données, les études sur l'*open data* et sur la production de ces données font échos au travail d'Alain Desrosières en indiquant que : "*les données brutes sont un oxymore*" [Gitelman \(2013\)](#). En fait, comme le montre notamment Jérôme [Denis et Gardey \(2018\)](#) et Samuel [Goëta \(2016\)](#) à propos des données administratives et de leur ouverture pour une mise à disposition du public (*open data*), ces données n'existent pas naturellement, elles sont systématiquement construites à des fins précises. Dans son livre, Jérôme Denis analyse les activités de fabrication des données à travers deux études de cas : le service administratif d'une grande banque française et le développement d'une application de calcul d'itinéraires cyclables basée sur des données produites par les municipalités françaises. Dans le premier cas, il montre que le travail des données s'appuie sur des opératrices qui manipulent et trient des dossiers administratifs et qui doivent tenir compte des ratés (par exemple, les pièces manquantes) tout en exécutant le travail le plus rapidement possible, introduisant leur propre logique dans ce travail de classification. Dans le second cas, il montre la difficulté dans l'agrégation et la standardisation de données qui proviennent de différentes villes françaises. De son côté, Samuel Goëta montre que la "mise en circulation" des données dans le cas de l'*open data* passe par un circuit de validation et de transformation permettant de créer ces données, c'est-à-dire de les rendre disponibles à tout le monde. Ces données sont alors transformées, passant d'éléments administratifs à des bases de données exploitables par d'autres.

Le travail de [Jaton et Bowker \(2021\)](#) fait office de référence dans la littérature en sciences sociales sur les négociations qui ont lieu lors du processus de création des algorithmes. Jaton a réalisé une enquête ethnographique dans un laboratoire d'IA spécialisé dans la "saliency detection", une branche de l'IA destinée à produire des modèles de traitement d'image ([Jaton, 2017](#)). Dans cet article, le détail des différentes étapes suivies par les chercheurs permet de mettre en lumière les problématisations successives qui permettent de concevoir le modèle algorithmique. D'abord, la redéfinition du "saliency problem" se donne pour objectif de créer un modèle capable de détecter et de segmenter le contour des visages alors que les modèles existants ne détectaient que des objets (*problématisation*). Cette nouvelle *problématisation* nécessite la construction d'une nouvelle "*ground truth*", une base de données d'images labélisées "*qui permet à la fois de façonner*

efficacement les algorithmes et d'évaluer leurs performances.”, différente de ce qui existait jusqu'alors dans la littérature. La construction de cette base passe elle-même par plusieurs étapes : agrégation d'images de visages trouvés sur Internet et annotation par le biais d'un service de *crowdsourcing* comme Amazon Mechanical Turk. L'utilisation de ce service permet aux chercheurs étudiés par Jatón d'envisager la création d'un modèle de détection de multiples visages : *”La collecte des nombreuses caractéristiques saillantes étiquetées manuellement deviendrait plus facile à gérer pour le groupe qu'elle ne l'était (auparavant) et une extension de la notion de saillance à de multiples caractéristiques deviendrait réalisable*”. Finalement, après avoir développé leur propre outil d'annotation dédié à la détection de visages, les membres du laboratoire ont pu payer 30 travailleurs pour réaliser cette tâche. Ils ont ensuite dessiné eux-mêmes les contours des visages au pixel près, sur la base du travail déjà réalisé par les travailleurs.

Le principal apport de Jatón est alors de redéfinir ce qu'on entend par le travail de modélisation, pour lequel il fait référence au concept de problématisation, c'est-à-dire *”une succession de pratiques collectives qui visent à définir les termes d'un problème à résoudre*”. Ce faisant, ses travaux montrent la nécessité de scruter le processus par lequel les modèles sont définis pour répondre à un problème précis, qui n'est jamais complètement stabilisé puisqu'il évolue tout au long des étapes de conception²². Sans forcément aborder de manière approfondie la question du travail des données, il montre l'importance de ce travail dans la constitution d'une *ground truth* essentielle à l'entraînement et à l'évaluation des modèles. Dès lors, il considère que les algorithmes sont des modèles de recherche d'informations, qui fonctionnent à partir de bases de références, produites par des travailleurs des données. Si ce travail est aussi essentiel à la problématisation, il devient donc particulièrement important de comprendre comment celui-ci est effectué²³.

Au croisement des STS et de la sociologie de la quantification, ces recherches abordent les négociations et les conventions qu'on retrouve derrière les dispositifs algorithmiques. Ce faisant, elles dévoilent les processus de production qui se cachent derrière la prétendue *”objectivité”* de ces dispositifs. Que ce soit pour la collecte et l'ouverture des données ou bien pour la création d'un modèle d'IA, tout revient à se demander qui sont les acteurs qui prennent les décisions, pour quelles raisons et à quel moment du processus de production. Derrière des dispositifs qui visent à rendre le monde *”calculable en tout”*, pour reprendre les mots de Dominique Cardon, c'est toute une chaîne de décisions humaines qui participe à la *problématisation*. L'intérêt de ces recherches est donc de participer à la critique de ces objets, en dévoilant les causes des

22. *”L'algorithme du groupe mettrait l'accent sur les situations et les pratiques spécifiques qui ont conduit à la définition des termes du problème que l'algorithme est censé résoudre. Comment le problème a-t-il été défini? Comment l'ensemble des données a-t-il été collecté?”* (Jatón, 2017)

23. *”Le fait de considérer les algorithmes comme des entités de recherche peut mettre l'accent sur les bases de données référentielles qui définissent ce que les algorithmes tentent de rechercher et de reproduire. Quelle vérité de base a défini les termes du problème que cet algorithme tente de résoudre? Comment cette base de données de vérité fondamentale a-t-elle été constituée? Quand et par qui? En mettant en évidence les moments et les lieux où les résultats à extraire ont été - ou sont - constitués dans des bases de données de vérités fondamentales, ce regard analytique sur les algorithmes peut suggérer de nouvelles façons d'interagir avec les algorithmes et ceux qui les conçoivent.”* (Jatón, 2017)

effets dispositifs de quantification sur les acteurs qui interagissent avec eux. Toutefois, on peut regretter qu'à l'exception des recherches de Jérôme Denis (2018) et de Samuel Goeta (2016), peu de travaux se soient focalisés sur les pratiques de production des données.

1.2.3 Algorithmes et travail numérique. Comment les nouveaux moyens de production numérique créent-ils de nouvelles tâches ?

La littérature en sociologie de l'IA a finalement très peu rendu compte de la place centrale du micro-travail dans le processus de production de l'intelligence artificielle. Dans le champ des études consacrées au travail de plateforme et notamment à celles de micro-tâches comme Amazon Mechanical Turk, différentes études ont pourtant pu établir un lien direct entre ces activités de production de données et le développement de l'IA. Ce faisant, elles élargissent le champ des acteurs à prendre en compte dans la compréhension du processus de production de l'IA. Avant de revenir plus en détail sur cette littérature, revenons sur l'exemple du moteur de recherche de Google.

Une première manière de critiquer les discours sur l'objectivité du moteur de recherche est venue de la littérature sur les biais de l'IA et des moteurs de recherche et notamment le travail de Safiya Noble (voir supra). D'autres enquêtes ont par ailleurs montré que les résultats affichés par le moteur de recherche avaient régulièrement été influencés par des interventions manuelles. Ainsi, l'entreprise interfère elle-même dans la prétendue "neutralité" de son moteur de recherche²⁴, par différentes interventions :

- des changements dans l'algorithme pour favoriser les plus grandes entreprises
- les ingénieurs de Google font des ajustements pour les informations proposées en plus de la liste des résultats, comme les encarts d'information
- la mise à jour par Google de sites bloqués, en plus de la liste de sites bloqués par la loi
- le filtrage des résultats de l'auto-complétion pour éviter les sujets les plus jugés trop polémiques
- des contractants sont recrutés sur des plateformes comme Appen pour évaluer la qualité des résultats du moteur de recherche.

Derrière le fonctionnement de services numériques aussi complexes, se cachent de nombreuses interventions humaines qui sont réalisées depuis différentes positions de l'entreprise. Ainsi, certaines interventions sont le fait des ingénieurs, d'autres sont liées à des enjeux de régulation ou de conformité avec la réglementation en vigueur, par exemple lorsque des sites sont interdits par la loi et doivent de ce fait être déréférencés. D'autres encore sont nécessaires pour vérifier le bon fonctionnement du site et sont réalisées par des travailleurs à qui Google sous-traite la vérification de la pertinence des résultats via la plateforme numérique de travail Appen, qui propose des services liés aux données à des entreprises clientes en s'appuyant sur des

24. West, K. G., Sam Schechner, Robert McMillan and John. (2019, novembre 15). How Google Interferes With Its Search Algorithms and Changes Your Results. Wall Street Journal. <https://www.wsj.com/articles/how-google-interferes-with-its-search-algorithms-and-changes-your-results-11573823753>

travailleurs indépendants du monde entier.

La recherche commence à établir un lien clair entre le développement de l'intelligence artificielle et le recours à une importante foule de travailleurs et travailleuses. Le travail de Lily Irani (2015) sur Amazon Mechanical Turk (AMT), la plateforme de micro-travail la plus connue, montre que le recours à une force de travail invisible permet de simuler une intelligence artificielle. On parle alors d'« *artificial artificial intelligence* » tel que l'indique le slogan d'Amazon Mechanical Turk. Par exemple, le lien entre micro-travail et intelligence artificielle est particulièrement saillant dans le secteur de l'automobile, pour le développement de la voiture autonome (F. A. Schmidt, 2019; Tubaro & Casilli, 2019). En effet, pour que les SIA des véhicules apprennent à interpréter correctement les signaux dans leur environnement, par exemple la présence de piétons dans la rue, il faut les « entraîner » avec de grandes quantités d'exemples d'images où les piétons, les autres voitures, les panneaux de signalisation routière, etc., sont clairement identifiés. Les plateformes de micro-travail peuvent créer ces exemples en faisant annoter par des humains des images enregistrées par les capteurs et les caméras des véhicules autonomes, afin d'identifier les différents éléments.

Paola Tubaro, Casilli, et Coville (2020a) ont identifié trois façons dont les travailleurs humains sont impliqués dans la construction de l'IA d'aujourd'hui : "AI preparation", "AI verification", "AI impersonation". Premièrement, ils doivent construire de grandes bases de données d'apprentissage de l'IA. Le travail humain est nécessaire pour étiqueter les images, transcrire les mots et interpréter les bribes de conversation enregistrées par les assistants vocaux (formation des données). Les travailleurs peuvent également effectuer des tâches de vérification de l'IA, car les technologies d'intelligence artificielle et d'apprentissage automatique existantes nécessitent ce qu'on pourrait qualifier de "contrôle qualité". Ainsi, ils vérifient que les moteurs de recherche donnent effectivement les résultats attendus, ou que les logiciels de traduction reconnaissent correctement les textes. Enfin, certains travailleurs peuvent imiter les comportements de l'IA. En effet, de nombreux systèmes d'IA ne sont pas encore capables d'effectuer des tâches de manière totalement autonomes.

Le travail de Benjamin Shestakofsky (2017) a également contribué à faire le lien entre le besoin de travail humain, souvent délocalisé hors des pays où l'IA est produite et commercialisée, et le développement de logiciels. À partir d'un travail ethnographique réalisé au sein d'une entreprise technologique qui propose un service d'intermédiation entre vendeurs et offreurs de services basé sur un algorithme d'appariement, il analyse trois configurations machine-travail humain. Lors de la phase de développement de la start-up (*scale*), il constate que les travailleurs délocalisés exécutent des tâches routinières, comme la recherche de données sur le web pour créer une base de données de clients potentiels. Les travailleurs se substituent au service algorithmique, en sélectionnant eux-mêmes les travailleurs éligibles par rapport à la demande. Dans une deuxième phase, l'entreprise est confrontée aux problèmes de compréhension des utilisateurs (*human lag*), problème résolu en s'appuyant sur le travail émotionnel de conseillers clients, également délocalisés.

Dans une dernière phase de stabilisation de la start-up, ce sont les améliorations du modèle de "matching" et les changements dans le service qui nécessitent continuellement de s'appuyer sur du travail humain, cette fois pour le perfectionnement de l'algorithme.

Plus récemment, Kate Crawford (2021) a analysé la matérialité de l'intelligence artificielle. Critiquant deux présupposés communs concernant l'IA, à savoir le fait qu'il s'agit de systèmes non-humains et le fait que l'intelligence puisse exister indépendamment du contexte social, culturel et historique, elle tente de comprendre comment l'intelligence est construite. Tout d'abord, elle montre l'ancrage de cette technologie dans un ensemble de processus extractifs, ressources naturelles (minerais), travail humain (extraction cognitive), l'histoire des savoirs et le travail de classification des concepts.

L'enjeu est de considérer que le travail de l'IA ne se réduit ni au travail des développeurs ni au travail même des annotateurs, mais de prendre en compte l'ensemble de ces dimensions ainsi que le contexte dans lequel ce travail est mené. Face aux discours invisibilisant les activités nécessaires à la production des systèmes d'IA, Crawford cherche à mettre en lumière le recours aux industries extractives (mines, données), souvent situées dans les pays en voie de développement, sur lequel ils sont en réalité fondés. Le lien entre l'automatisation, le développement des SIA et le travail a fait l'objet d'importants débats au cours des dernières années, avec, en premier lieu, l'hypothèse de la fin du travail. Les recherches présentées dans cette section montrent pourtant que ces activités d'automatisation nécessitent du travail humain, à différents niveaux de la production. Qu'on parle des mineurs qui récoltent les métaux rares présents dans le matériel ou bien des travailleurs qui produisent et annotent les données, voire même qui remplacent les algorithmes, les SIA ont besoin de travail humain pour fonctionner.

* * *

Dans ce chapitre, nous avons analysé un corpus de publications classées dans la sociologie des algorithmes et de l'IA. Il nous faut faire trois constats. D'abord, il nous semble d'une partie des recherches mobilisées tombent dans le "drame algorithmique" cité par Florian Jatton (2019) en analysant ces objets comme "entités flottantes" désencastrées de la réalité sociale. S'il convient de reconnaître une réelle richesse de la littérature sur les effets des algorithmes (les biais, le management algorithmique), on retrouve peu d'analyses des causes de ces effets. Il faut toutefois noter la part des recherches consacrées à la figure des concepteurs : ces développeurs sont alors réinscrits dans leur milieu professionnel (*la tech*) et dans leur situation sociale (des hommes, principalement blancs). Notre constatons également que peu de recherches analysent le processus de production de l'IA dans son ensemble en restituant l'ensemble des acteurs impliqués; bien souvent, une partie seulement de la chaîne de production est analysée. Ainsi, le travail de Kate Crawford mis à part, peu de recherches réussissent à traquer, ou du moins à brosser une vue d'ensemble, de toutes les

activités nécessaires à la production et au déploiement de SIA ²⁵. Par ailleurs, de nombreuses études se sont focalisées sur des modèles développés en laboratoire ou par les grands acteurs du numérique, notamment anglophones.

Pour remédier à cela, il a fallu mobiliser une littérature qui porte sur la circulation des données et la fabrication des dispositifs de quantification pour trouver un point d'ancrage théorique permettant d'affirmer un constat théorique simple. Les modèles d'IA sont le fruit de l'articulation de nombreuses décisions qui, prises séparément comme dans certaines recherches, surévaluent l'influence de l'acteur étudié (le concepteur, l'utilisateur...), sans nécessairement pour étayer le mode de fonctionnement du modèle. Dans ces recherches, les conditions de travail des travailleurs qui contribuent à produire l'IA, particulièrement depuis les pays en voie de développement, demeurent largement invisibles. Pour restituer cette dimension, une autre catégorie de littérature peut être mobilisée, qui, si elle n'a pas comme objet premier l'étude de la fabrication l'IA proprement dite, a comme objectif d'analyser la contribution des travailleurs "invisibles" qui créent les données utilisées pour cette technologie.

25. Il faut préciser qu'en plus son travail manque de profondeur empirique

Chapitre 2 - Le *”digital labor”* comme cadre général pour comprendre l’externalisation du tra-vail des données

La recherche en sciences humaines et sociales s’est largement intéressée à l’émergence de trois nouveaux types d’organisation du travail : le développement du *travail gratuit* a conduit à davantage de recherches autour du *digital labor* ou de *l’hétéromation* (Cardon & Casilli, 2015; A. A. Casilli, 2019; H. Ekbia & Nardi, 2017; Scholz, 2012). Ces notions permettent de désigner d’une part la captation de la valeur produite par les utilisateurs d’Internet par les entreprises du numérique (*travail gratuit*), d’autre part les risques associés au travail platformisé pour les travailleurs (*platformisation*). Enfin, l’étude du travail rémunéré de production de données a donné lieu à des enquêtes sur l’impact de cette activité sur le développement des pays du Sud et a finalement ancré le terme dans les études sur le développement international (*impact sourcing*).

Dans cette seconde section, nous traitons donc de l’émergence de ces nouvelles formes d’organisation du travail. Dans un premier temps, nous verrons que la littérature consacrée aux *”travailleurs du clic”* se focalise prioritairement sur le travail de production réalisé gratuitement par les usagers des services Internet et sur le travail rémunéré via des plateformes. Dans un second temps, nous analyserons plus en détail la littérature sur le micro-travail, un terme générique pour désigner le travail du clic. Nous verrons que la thématique des conditions de travail sur les plateformes a été abondamment traitée, au détriment d’autres questions liées à cet objet.

2.1 La littérature sur les *”travailleurs du clic”* se focalise prioritairement sur deux modes d’organisation du travail : le travail gratuit et le travail des plateformes

Dans cette première sous-partie, nous verrons que le *”digital labor”* a permis de mettre en lumière trois organisations du travail : le *crowdsourcing*, le *human computation* et le micro-travail, qui ont fait émerger les problématiques de travail gratuit et de la platformisation, et, par conséquent, d’externalisation des tâches. L’enjeu de cette partie est de discuter chacun de ces modèles et la manière dont ils ont été conceptualisés dans la littérature, en insistant sur le fait que le travail numérique crée du travail précaire, qu’il soit réalisé de manière gratuite ou sur des plateformes.

2.1.1 Produire des données grâce à des consommateurs-travailleurs peu ou pas rémunérés

Dans un article, Berg (2019) analyse le développement de formes d’emplois qualifiées de *”périphériques”* dans une tentative de tenir compte du développement du travail *”atypique”*. Ce type d’emploi est défini par contraste avec la forme canonique de la relation de travail, à savoir l’emploi salarié à temps plein et pour une durée indéterminée, au profit d’un unique employeur et bénéficiant des garanties du droit du

travail et des conventions collectives. Cette croissance de l'emploi non-standard passe par le recours à des contrats dits atypiques (travail temporaire, travail intérimaire) et, plus récemment, par la plateformes de l'économie. En fonction du système réglementaire de chaque pays, les entreprises numériques ont alors recours à diverses formes de statuts d'emploi non salarié, comme le statut de "micro-entrepreneur" en France ou les contrats "zéro heure" au Royaume-Uni (Régny et al., 2020).

Dans cette nouvelle configuration, le *digital labor* s'est attaché à décrire la forme la plus périphérique, à savoir le travail non rémunéré (*free labor*). Cette notion s'attache à faire entrer dans le périmètre du travail certaines tâches, malgré l'absence de rémunération qui fait pourtant partie des critères traditionnels permettant de distinguer ce qui relève ou non de la catégorie travail. Ainsi, le travail gratuit permet à l'origine, et sous l'impulsion de sociologues féministes du travail, de considérer des activités assumées par des femmes, telles que les soins ou les tâches domestiques. Avec la numérisation, le *digital labor* fait entrer dans cette catégorie des tâches réalisées sans contrepartie monétaire par les internautes.

Dès 2002, Penny Gurstein identifie le développement de formes d'emplois atypiques par le télétravail, ou travail virtuel, qu'elle définit comme une forme de travail flexible, intermédiée par les technologies de l'information et de la communication (Gurstein, 2002). L'enjeu pour Gurstein est de savoir si le télétravail est une alternative aux autres formes de travail plus formelles (p. 20) notamment pour les femmes, dans une perspective de genre, à partir d'une série d'entretiens réalisés aux États-Unis et au Canada. Le recours plus fréquent à ces nouvelles manières de travailler serait une conséquence de l'internationalisation de l'économie, en poussant les entreprises à réduire leurs coûts.

Dans le secteur des services, ce processus s'est doublé d'une transformation technologique que Gurstein décrit comme une automatisation des bureaux (*office automation*), soit l'utilisation des premières technologies numériques pour la réalisation des tâches des employés de bureau, telles que faire des calculs ou taper un texte. Enfin, de nombreuses entreprises ont eu de plus en plus recours à une force de travail externalisée pour contribuer à ces évolutions technologiques. L'autrice étudie notamment des femmes qui, depuis chez elles, produisent des données ou numérisent des documents (par exemple : pour des cabinets médicaux). D'après Gurstein, le développement du télétravail serait le résultat de ces trois dynamiques : réduction des coûts, numérisation et externalisation. Au croisement du changement technologique, de la mondialisation de la force de travail et du développement de l'emploi atypique, le travail de production de données par les télétravailleurs des pays développés rompt avec l'utopie d'un travail à la maison censé être libérateur pour les femmes. En effet, la "nature disparate du travail contractuel et du travail à la pièce, associée aux obligations domestiques de cette population majoritairement féminine, rend pratiquement impossible pour les travailleurs de s'organiser efficacement, ce qui les rend vulnérables à l'exploitation et prêts à accepter des conditions de travail inférieures aux normes." (p.197). Dès le début des années 2000, les travaux de Gurstein permettent donc de faire le lien entre travail des données, développement du numérique et atypisation des formes de

travail, reposant tantôt sur des délocalisations, tantôt sur du travail réalisé depuis chez soi plutôt par des femmes. Dans ce dernier cas, c'est la possibilité de recourir à des statuts d'emplois moins protecteurs, donc moins coûteux, pousse les entreprises à mettre en place cette organisation du travail.

Le *digital labor* naît formellement au début des années 2010, à travers la première conférence traçant les contours de ce qui deviendra un programme de recherche (Scholz, 2012). Avec la naissance de l'entreprise Uber, et l'émergence du concept d'"ubérisation", les participants constatent qu'"on a vu apparaître des environnements de travail sur le web, sans protection pour les travailleurs, y compris pour les emplois les plus précaires" (Scholz, 2012) (p. 1). Pourtant, loin de se résumer à l'ubérisation, le digital labor désigne tout autant les activités numériques - ou intermédiées par celui-ci - précaires que l'extraction de valeur marchande des "activités numériques quotidiennes des usagers de plateformes sociales, d'objets connectés ou d'applications mobiles" (Cardon & Casilli, 2015) (p. 13). Ainsi, le digital labor s'intéresse aux activités reposant sur : l'exploitation de certains contenus produits par les utilisateurs de services numériques; l'exploitation des données personnelles des utilisateurs; la précarisation de certaines activités (autrement dit l'ubérisation ou la plateformesation).

Dès ses origines, le digital labor accorde une place prépondérante à la question du travail gratuit. Andrew Ross y voit un exemple emblématique dans le rachat par AmÉrica Online (AOL) du Huffington Post, un média en ligne qui propose des articles publiés par des contributeurs bénévoles²⁶. Lors du rachat du média pour 315 millions de dollars, un groupe de contributeurs réguliers, mais non rémunérés, du Huffington Post lance une procédure judiciaire pour réclamer leur part de la valeur quantifiée par le processus de rachat. Dans le même ouvrage, Tizia Terranova et Abigail de Kosnik prolongent l'analyse en montrant qu'une grande partie de la valeur créée par l'économie numérique s'appuie sur la contribution volontaire de "travailleurs" bénévoles. La première étudie le cas des modérateurs bénévoles responsables de gérer les salles de discussion proposée par l'opérateur américain AOL. La seconde élargit la perspective à l'ensemble des contenus générés par les "fans", lorsqu'ils commentent une vidéo sur YouTube ou qu'ils commentent la publication de leur artiste préféré, ou bien même lorsqu'ils produisent des contenus de blogs consacrés à leurs passions. Au-delà des contenus publiés par ces internautes dévoués, c'est en fait désormais bien connu que l'ensemble des données générées par nos interactions sociales quotidiennes en ligne produit de la valeur pour les entreprises du numérique. En 2013, les réflexions sur le travail gratuit propre au développement du numérique deviennent au cœur de propositions sur la fiscalité des entreprises qualifiées de "géants" du numérique, afin de tenir compte de la valeur extraite des données produites gratuitement par les consommateurs. Le rapport de la mission d'expertise sur la fiscalité de l'économie numérique propose par exemple une fiscalité

26. Cet exemple est repris par Maud Simonet (2018), dans son livre sur le travail gratuit. Elle y analyse le développement de dispositif de réinsertion par l'emploi qui conduisent certains bénéficiaires d'aides sociales de réaliser des activités d'intérêt général (par exemple : nettoyer un parc), d'activités numériques comme l'écriture d'articles par les contributeurs du Huffington Post et les tâches liées au travail des consommateurs (par exemple : réserver son billet de train dans une borne)

spécifique à ces acteurs de manière à *”faire des données une matière imposable”* soit, d’une certaine manière, à valoriser le travail gratuit.

H. R. Ekbia et Nardi (2017) font émerger le terme d’*”heteromation”* pour désigner *”l’extraction de la valeur économique d’un travail à faible coût ou gratuit dans les réseaux médiatisés par ordinateur”*. Selon eux, de nombreux systèmes numériques sont conçus pour extraire un travail non ou peu rémunéré. Le système *”Captcha”*, par lequel il est demandé aux utilisateurs de confirmer qu’ils ne sont pas des robots, en est l’exemple par excellence. En effet, le Captcha repose sur la publication de la thèse de Luis von Ahn en 2005 sur le *”crowdworking”* (le travail de la foule), qui introduit le concept de calcul humain (*”human-computation”*), une technique de coopération homme-machine dans laquelle des individus résolvent des problèmes numériques à la place des programmes informatiques. Les Captchas en sont l’application directe. Ils sont notamment utilisés par Google pour compléter la numérisation automatique des ouvrages de sa librairie en ligne, Google Books, et deviennent l’archétype du travail non rémunéré réalisé par les consommateurs de services numériques, qui deviennent ainsi usagers-producteurs.

Dans la continuité de ces travaux, Antonio Casilli (2019) présente une autre vision des plateformes classées selon leur capacité à organiser la captation de la valeur produite par leurs utilisateurs-travailleurs. Selon lui, les plateformes numériques seraient des *”organisations hybrides, à l’intersection d’un marché et de l’entreprise”* brouillant les frontières entre l’intérieur et l’extérieur de la firme : les chauffeurs Uber sont par exemple qualifiés, par l’entreprise américaine, d’indépendants, la plateforme se bornant à assumer un rôle d’intermédiation proche du *”marché”*. Cependant, ce récit est largement remis en cause par les travailleurs eux-mêmes qui, au fil des procédures en requalification, montrent le rôle d’organisation hiérarchique des relations de travail par la plateforme, ce qui la rapproche du modèle *”entreprise”*. Ces hybrides marché-entreprise extraient la valeur des données collectées de trois manières, selon Antonio A. A. Casilli (2019) : la valeur de qualification, produite par les utilisateurs lorsqu’ils *”aiment”* des contenus ; la valeur de monétisation, soit la vente de données à des fins publicitaires ; la valeur d’automation, c’est-à-dire l’utilisation de données pour créer d’autres services. Ces modes s’appuient indirectement sur deux types de populations présentes sur Internet : les travailleurs des plateformes et les utilisateurs. Le fonctionnement de la plateforme Uber en est à nouveau l’illustration : les utilisateurs comme les chauffeurs produisent sans cesse des données, sous forme de notes ou de données de géolocalisation, essentielles au service. Par ailleurs, l’enregistrement constant de certaines données liées à la conduite, comme la vitesse ou le freinage alimentent un réservoir de données qui doivent permettre la production et l’amélioration d’autres services, tels que l’entraînement de modèles dans l’objectif de la voiture autonome.

En 2019, des travaux tels que le livre de Couldry et Mejias (2019b) ont comparé cette activité extractive de valeur, sans reconnaissance de ceux qui la produisent, à une forme de colonialisme des données (*”data colonialism”*). Les auteurs le définissent comme *”un ordre émergent pour s’appropriier et extraire des ressources*

sociales à des fins lucratives grâce aux données”. Comme dans d’autres recherches sur le capitalisme numérique, ils se fondent principalement sur le processus d’appropriation et de valorisation des données produites par les utilisateurs des réseaux sociaux. On peut néanmoins se demander s’il est vraiment possible de mettre sur le même plan l’extraction de valeur des données produites par les consommateurs des pays du Nord et l’extraction de valeur tirée des chaînes de sous-traitance internationale pour la production de l’IA. Est-ce que le terme de ”colonialisme” s’applique de la même manière? Que ce soit dans le travail de [Couldry et Mejias \(2019a\)](#) ou dans le reste de la littérature citée dans cette partie, on retrouve peu de mentions de la structure des coûts du capitalisme numérique, et donc la structure de l’appropriation économique des grands acteurs du numérique. De ce point de vue, on peut s’interroger : peut-on mettre au même point des travailleurs malgaches très peu payés qui passent 40 heures par semaine à ”raffiner” des données avec des influenceurs ou des utilisateurs d’Instagram qui postent des photos.

Le travail externalisé constitue un coût en ressources humaines alors que le ”travail” des usagers des réseaux sociaux permet aux *big tech* de gagner directement de l’argent à travers la publicité. Par ailleurs, peut-on mettre au même niveau l’expérience de travail de ces deux catégories de travailleurs? Les uns, situés à l’autre bout de la planète exerceraient plutôt pour survivre, tandis que les seconds bénéficieraient de la sécurité économique des pays du Nord. Finalement, l’intérêt de cette littérature est de mieux comprendre la domination des grands acteurs du numérique. Cette dernière s’exerce à travers une infrastructure technique leur permettant de capter de la valeur et s’appuie sur de multiples formes de travail précaire, considérées comme loin du ”cœur” de l’entreprise.

2.1.2 La littérature sur le micro-travail se concentre sur les plateformes numériques de travail, principalement anglophones

Dans le secteur de la production et de l'annotation de données, c'est le travail réalisé via les plateformes sociales comme Facebook et via les plateformes numériques de travail comme Amazon Mechanical Turk qui ont reçu une attention particulière. Julian Posada et Antonio Casilli considèrent ainsi que : *"le modèle de plateforme constitue le principal paradigme organisationnel des grandes entreprises qui développent aujourd'hui l'intelligence artificielle"* (2019). Le micro-travail est toutefois un terme qui n'a pas été uniquement associé à Mturk. Il a en effet été inventé par Leila Chirayath Janah pour présenter les activités de Samasource (aujourd'hui Sama), sa plateforme numérique de travail. Inspirée de la pratique du micro-crédit, celle-ci consiste à proposer des tâches numériques à des travailleurs situés dans des pays en voie de développement en échange d'une rémunération et d'un programme de formation au numérique (*impact sourcing*). Pour traiter de ce concept, dans avons choisi de proposer une analyse bibliométrique des articles en sciences sociales traitant du micro-travail.

En effet, pour le moment, peu de travaux ont en effet tenté de situer cet objet d'étude de manière systématique. Nous avons donc décidé de travailler sur ce concept en nous demandant : quelles sont les caractéristiques et les lacunes de la littérature sur le micro-travail? Dans les deux prochaines sections, nous nous appuyons sur ce travail quasi empirique dont l'objectif est d'analyser les tendances et les lacunes de la littérature en sciences sociales sur le micro-travail. Cette recherche a été menée sur la base de trois hypothèses : la majorité des recherches portent sur les pays anglophones; de nombreuses publications traitent des conditions de travail et la plupart des publications se concentrent sur la plateforme du travail.

Notre méthodologie se base sur le protocole PRISMA de collecte et de sélection des articles. Nous avons collecté 1279 articles sur 8 bases de données d'articles scientifiques. La collecte a été arrêtée au 31 juillet 2020. À l'issue du processus d'exclusion, nous avons finalement retenu 148 articles pour une lecture et une analyse complète. Nous avons trié manuellement les articles en fonction de plusieurs catégories que nous avons construites afin de répondre à nos questions de recherche. Une première variable a été de définir à quel type de travail les auteurs font référence (i.e. freelancing en ligne, micro-travail...). Une autre variable est la "face" du marché qui est couvert par un document donné (i.e. la plateforme, les travailleurs ou les clients). Nous avons également créé manuellement une variable "thème principal", qui permet de classer les articles en fonction de leur objectif principal, c'est-à-dire s'ils traitent de questions de développement ou de conditions de travail. La dernière variable consiste à examiner les pays étudiés dans chaque document. Il peut s'agir d'un seul pays (par exemple, les États-Unis), de deux pays (les États-Unis et l'Inde) ou de plusieurs. L'analyse de cette base de données a permis de dégager des grandes tendances de la recherche sur le micro-travail.

Nous notons d’abord que cet objet d’étude a pris de plus en plus d’importance depuis 2010. Plusieurs interprétations peuvent être avancées. Tout d’abord, cette augmentation correspond à une prise en compte plus générale des activités liées à la ”gig economy” dans le débat public (i.e *ubérisation*). Le développement de l’intelligence artificielle dans le monde industriel et une volonté réglementaire de la part des pouvoirs publics peuvent également expliquer le développement de recherches sur le sujet.

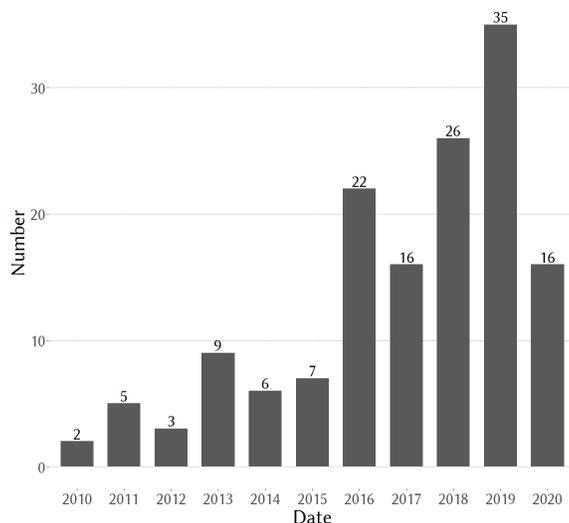


FIGURE 1 – Nombre de publications sur le micro-travail par an. L’année 2020 n’est pas complète, notre collecte de données s’étant arrêtée en juillet 2020

Sur le graphique ci-dessous, on observe que très peu de publications concernent d’autres modes d’externalisation du micro-travail que le recours à des plateformes numériques de travail. Ainsi, les publications se focalisent principalement sur les ”crowdworkers”, ”micro-work” et ”crowdsourcing” et donc principalement sur l’étude des plateformes numériques de travail. Les études qui traitent de formes plus traditionnelles d’externalisation sont quant à elles peu présentes. Très peu d’études s’attardent sur le ”Business Process Outsourcing” (BPO), c’est-à-dire une externalisation d’une entreprise cliente vers une entreprise sous-traitante. Dans la perspective de notre terrain, ce premier résultat a un impact assez significatif dans la mesure, où suivre les perspectives déjà étudiées nous conduirait à mener un terrain d’enquête sur les plateformes.

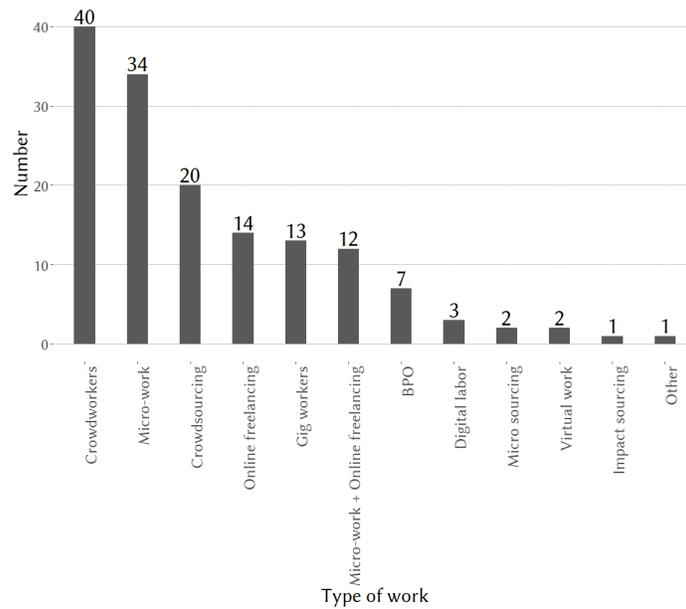


FIGURE 2 – Nombre de publications par concept mobilisé

Nos résultats montrent que la recherche s’est principalement focalisée sur quelques plateformes et notamment la plus connue d’entre elles, Amazon Mechanical Turk qui est surtout connue pour être utilisée par des entreprises américaines et par des travailleurs américains et indiens.

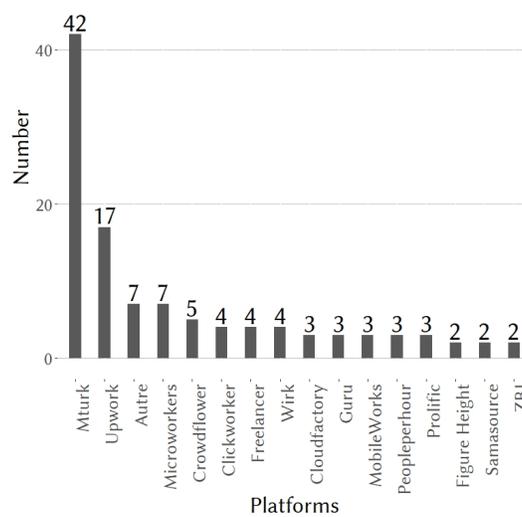


FIGURE 3 – Nombre de publications par plateformes étudiées

De notre point de vue, cette tendance de la recherche sur le micro-travail mérite d'être questionnée tant elle met particulièrement en avant, la sphère anglophone et le modèle de travail promu par Amazon. Or, il existe tout une variété des plateformes mobilisées dans cette activité, ainsi que d'autres modes d'externalisation. Par exemple, Paola Tubaro (2021) a proposé un cadre d'analyse pour distinguer deux types de plates-formes de micro-travail : les plates-formes comme intermédiaires de marché avec des liens économiques "minces" et des travailleurs très autonomes "dans une forme de désencastrement" et les plates-formes qui "participent à des structures complexes d'externalisation interentreprises" qui impliquent un "encastrement profonde" pour les travailleurs (Appen). L'intérêt de ce résultat est d'interroger ultérieurement la pertinence de cette orientation de la littérature. En effet, loin d'être un "objet unique", la plateforme est un ensemble d'entreprises qui mettent en place diverses formes d'organisation du travail. Un autre résultat important de cette revue de littérature est que les travailleurs et les plateformes sont les objets les plus étudiés dans la littérature alors que les clients finaux sont très peu discutés.

Une première raison qui explique cette caractéristique de la recherche sur le sujet tient de l'inscription de l'objet plateforme dans une crainte du "futur du travail" notamment à travers les différentes publications de l'Organisation Internationale du Travail (Berg, Furrer, Harmon, Rani, & Silberman, 2018; Organization, 2021). Par ailleurs, cette focalisation sur les travailleurs peut également s'expliquer par l'idée que les plateformes pourraient être des modes d'organisation pertinents dans des pays en voie de développement qui souffre de taux de chômage importants ou qui présente des taux élevés en matière de travail informel (Galpaya, Perampalam, & Senanayake, 2018; Kuek, Paradi-Guilford, Fayomi, Imaizumi, & Ipeirotis, 2015; Malik, Nicholson, & Heeks, 2017; Paradi-Guilford, 2013).

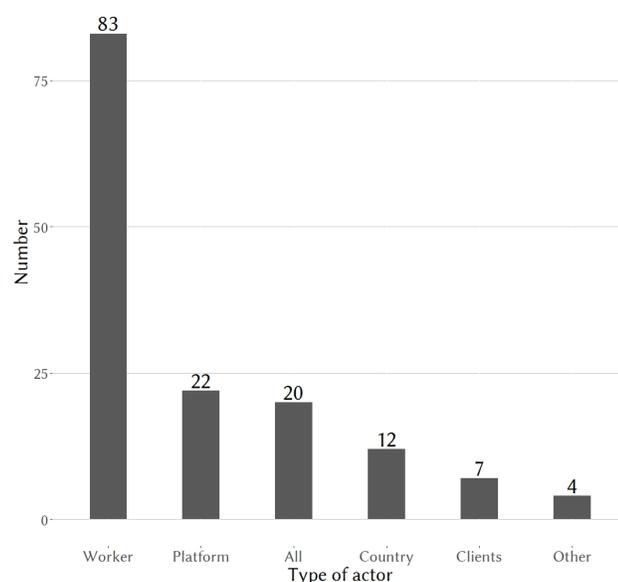


FIGURE 4 – Nombre de publications par faces de marché étudiées

Nous avons également classé les publications selon les pays étudiés dans le cadre différentes études. Très logiquement, par rapport aux éléments déjà discutés précédemment, on constate une forte prédominance des É.-U. et de pays anglophones des Suds comme l'Inde, les Philippines ou le Kenya (Figure 5.) Cette caractéristique de la recherche sur le micro-travail est problématique tant elle méconnaît d'autres contextes culturels, sociaux et linguistiques qui seraient autant de chaînes de valeur d'externalisation du travail de la donnée (et notamment le monde hispanophone et anglophone). Elle méconnaît également des mécanismes de sous-traitance en

cascade et de relocalisation qui peuvent se mettre en place à mesure que l'industrie de l'IA évolue et que la division internationale du travail se reconfigure (Anwar & Graham, 2020).

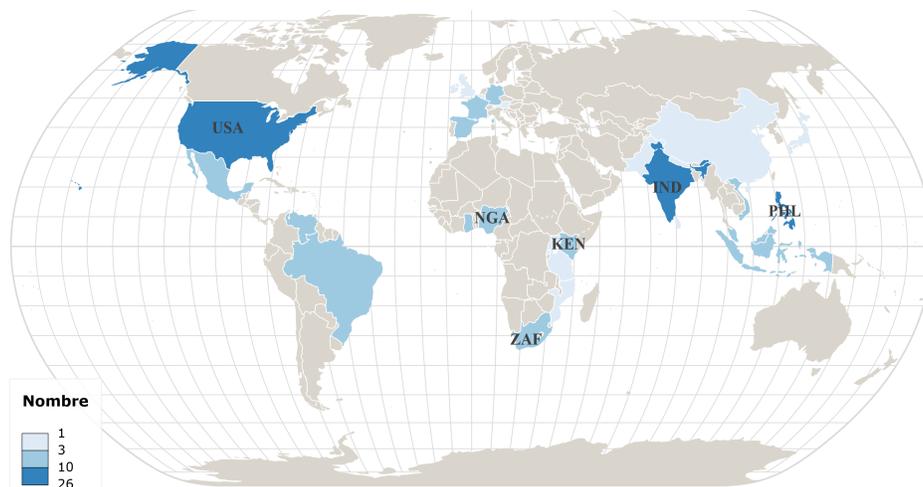


FIGURE 5 – Carte des pays enquêtés dans les études sur le micro-travail. Lecture : plus un pays a été étudié dans la littérature sur le *micro-travail*, plus il est affiché en sombre sur la carte.

Au travers de cette première analyse de la littérature consacrée au micro-travail, on constate plusieurs tendances. Tout d'abord, celle-ci se focalise principalement sur les plateformes numériques de travail et notamment Amazon Mechanical Turk. Certes, le service créé en 2006 constitue un des premiers et un des plus médiatisés, mais des articles plus récents tendent à montrer l'émergence de nouvelles formes d'organisation du travail, on pense notamment au concept de "deep labor" de Tubard (2021).

* * *

Dans une première section, nous avons proposé une analyse critique de la littérature qui porte sur le *digital labor* et le *colonialisme de données (data colonialism)* Couldry et Mejias (2019a). Ce faisant, nous avons réinscrit ces deux concepts dans une perspective comparative avec notre objet d'études qui porte au contraire sur des travailleurs rémunérés situés à Madagascar. Pour nous, si ces perspectives de recherches permettent d'analyser les processus de production de valeur des grandes entreprises du numérique, elles sont peu opérationnelles dans l'analyse des relations commerciales entre d'autres acteurs de l'IA et leurs prestataires.

Nous avons ensuite proposé une première analyse de la littérature scientifique sur le *micro-travail* en sciences humaines. On peut finalement se demander si ce mode d'organisation du travail est adapté à des modèles d'IA commerciaux qui ont vocation à être utilisés quotidiennement, contrairement aux modèles issus de la recherche scientifique. Ainsi, si la littérature en sciences informatiques questionne de plus en plus la construction des jeux de données et l'implication des travailleurs des données dans celle-ci (Paullada, Raji,

Bender, Denton, & Hanna, 2020), il manque encore d'études empiriques portant sur des cas réels sur non sur des exemples tirés de la recherche scientifique.

2.2 De l'étude de l'organisation du travail de plateformes au tournant global des études sur le travail des données

Les recherches sur le micro-travail se sont beaucoup focalisées sur les conditions de travail des travailleurs des données. Cette orientation de la recherche correspond à plusieurs enjeux. Un premier enjeu est éthique et concerne le niveau de paiement de travailleurs d'Amazon Mechanical Turk recrutés pour des projets de recherche²⁷. Un deuxième enjeu est lié à l'inscription du sujet dans une littérature plus globale sur la plateformes et l'ubérisation et sur les conséquences de ces organisations du travail sur les travailleurs.

Une autre thématique de recherche développée dans la littérature concerne plutôt la quantification et la qualification des micro-travailleurs. L'enjeu est de comprendre qui sont ces travailleurs et combien sont-ils. Ces deux enjeux correspondent alors à la nécessité d'évaluer les risques de biais dans le recours à des plateformes de micro-travail, notamment lors de la passation de questionnaire (notamment sur Mturk).

Une dernière thématique concerne les enjeux du micro-travail en termes de développement pour les pays à faibles revenus. Le développement de plateformes à visée sociale, ainsi que le constat de chaînes d'externalisations entre pays du Nord et pays du Sud (par exemple : entre les É.-U. et l'Inde) pointe vers ce type d'analyse. Par exemple, dans deux rapports, la Banque Mondiale considère que le recours au micro-travail est pertinent en matière de développement de certaines régions du monde. L'Organisation Internationale du Travail 2018 se montre plus nuancée en indiquant que : *“Le micro-travail procure certes des avantages financiers immédiats, mais il offre peu de perspectives en termes de progression de carrière et de mobilité économique”*. Nous verrons qu'une large partie de la littérature scientifique porte alors sur le potentiel d'attraction des pays en voie de développement pour ces activités.

2.2.1 Qualifier et quantifier la foule de micro-travailleurs

L'étude de la composition de la force de travail des sites de micro-travail a constitué un enjeu central de la littérature, notamment poussé par les utilisations de ces sites pour la passation de questionnaire. Cet enjeu méthodologique a soutenu des recherches dont l'objectif était de *qualifier* cette *foule*. La question de la quantification des micro-travailleurs répond quant à elle à deux enjeux bien distincts. Un premier enjeu déjà abordé est purement scientifique puisqu'il s'agit de connaître le "réservoir" de travailleurs susceptibles de participer à des expériences ou de collecter des données. Par ailleurs, les chercheurs en sciences sociales,

27. Il est alors nécessaire d'étudier les risques éventuels de biais en cas de paiement trop important

les juristes et les régulateurs tentent d'estimer le phénomène du travail des plateformes (*platform work*) et du travail en ligne (*virtual work*) (Pesole, Fernandez-Macias, Brancati, et Herrera (2019)).

Panos Ipeirotis (2010) a été un des premiers à conduire une enquête sur les caractéristiques sociodémographiques des travailleurs de Mturks (*turkers*). Il constate que 50% des travailleurs sont originaires des États-Unis et 40% de l'Inde. Les turkers sont plus jeunes que la population générale (54% d'entre eux ont 21 et 35 ans), ce sont des femmes (70%) notamment aux É.-U. et ils ont des revenus plus faibles que la population générale, malgré un niveau d'éducation plus élevé. La plupart des recherches souligne que cette population est composée en majorité de personnes entre 18 et 40 ans, relativement plus diplômées que la population générale, quelque soit le pays du travailleur (Berg et al., 2018; Posch, Bleier, Flöck, & Strohmaier, 2018). Des recherches futures observeront très peu de différences sur ces caractéristiques, malgré la création de nouvelles plateformes opérant dans de nouveaux pays en plus des États-Unis et de l'Inde.

Concernant le genre des travailleurs des données, plusieurs résultats se contredisent (Tubaro, Coville, Le Ludec, et Casilli (2022)), même s'il est possible d'observer une tendance générale : il y a plus de femmes réalisant du travail du clic dans les pays développés comme la France ou les É.-U., par rapport aux pays en voie de développement comme le Venezuela ou aux nouveaux pays industrialisés comme l'Inde. Dans le cas des États-Unis, plusieurs recherches trouvent une proportion similaire de femmes travailleuses des données d'à peu près 55% (Difallah, Filatova, & Ipeirotis, 2018; Ipeirotis, 2010) à 60% (Posch et al., 2018)²⁸. En France, la proportion est de 55% (A. Casilli et al., 2019b) sur une plateforme uniquement disponible dans ce pays. À l'inverse, une étude européenne menée sur quatre plateformes de micro-tâches a révélé une proportion plus élevée (60%) d'hommes que de femmes (Forde et al. (2017)). Couvrant soixante-quinze pays et cinq plateformes internationales, l'OIT a dénombré une femme sur trois travailleurs au total, une femme sur cinq travailleurs dans les pays en voie de développement (Berg et al. (2018)).

Une première recherche montre que les travailleurs de Mturk sont entre cent et deux cents mille et que cette population est stable (Difallah et al. (2018)). Une seconde recherche centrée sur la France et tentant d'agréger plusieurs plateformes montre que jusqu'à 260 000 personnes sont inscrites sur des plateformes de micro-travail. Parmi elles, quelque 50 000 sont des travailleurs "actifs" qui effectuent des micro-tâches au moins une fois par mois. Un troisième halo de travailleurs est plus restrictif et ramène ce chiffre à 15 000 travailleurs "très actifs" (Tubaro, Le Ludec, et Casilli (2020)). Au niveau européen, le travail de (Pesole et al., 2019) propose une estimation du nombre de travailleurs en pourcentage de la population générale. Ce nombre est situé entre 6 et 12% de la population selon les pays, dont 7% pour la France et 9,7% dans l'ensemble des pays de l'UE. (Kässi, Lehdonvirta, & Stephany, 2021) propose une approche mixte pour estimer le nombre total de travailleurs indépendants actifs sur des plateformes dans le monde, un objectif qui diffère substantiellement des études précédemment citées, mais dont le nombre englobe les travailleurs

28. Respectivement sur les plateformes Mturk et Appen

des données. Dans cette enquête, les chercheurs d'Oxford estiment à 163 millions le nombre de *travailleurs indépendants de plateformes* inscrits, à 14 millions ceux qui ont déjà obtenu du travail au moins une fois, et à 3,3 millions ceux qui ont réalisé au moins 10 projets ou gagné 1000 dollars.

Une dernière catégorie de recherche consiste à tenter d'identifier les flux entre d'un côté les demandeurs de travail (entreprises clientes) et les offreurs (travailleurs). Matthias Hirth, Hoßfeld, et Tran-Gia (2011) a étudié la démographie des microtravailleurs du site *microworkers* et montre que les employeurs viennent principalement des États-Unis (28%) et les travailleurs d'Indonésie (18%). Ils montrent que "les pays peu développés représentent environ le même nombre de travailleurs que les pays très développés" tandis que les pays à revenu moyen comme l'Indonésie représentent la source principale de main-d'œuvre, ce qui peut s'expliquer par un accès facilité à Internet par rapport à d'autres pays peu développés. Le travail de Mark Graham et Vili Lehdonvirta confirme cette tendance, en montrant, à partir des données d'une plateforme de *freelancing*, que la plupart des acheteurs sont situés dans les pays développés, par exemple aux États-Unis, tandis que les fournisseurs sont situés dans les pays en voie de développement, notamment en Inde (Graham, Hjorth, & Lehdonvirta, 2017a; Lehdonvirta, Barnard, Graham, & Hjorth, 2014). Ils montrent que lorsque les sous-traitants sont dans le même groupe linguistique que le client, ils ont légèrement plus de chances de conclure un contrat pour un travail complexe que les sous-traitants qui n'appartiennent pas à ce groupe linguistique.

En se basant sur les données de la plateforme de travailleurs indépendants *Upwork*, un rapport de la Banque Mondiale (2015) observe que les pays présentant le plus de travailleurs sont les É.-U. (23.9%), l'Inde (21.5%), les Philippines (18.6%), le Pakistan (5.8%) et le Royaume-Uni (4.2%) (Kuek, Paradi-Guilford, Fayomi, Imaizumi, Ipeirotis, Pina, & Singh, 2015). Pour ce qui est du micro-travail, les auteurs citent une interview réalisée avec un des représentants de la plateforme Crowdsourcing qui indique qu'une "grande part de leurs travailleurs sont aux É.-U., aux Philippines, en Inde et au Royaume-Uni".²⁹ Le rapport de l'OIT précédemment cité (2018) s'appuie sur les données d'un questionnaire passé sur plusieurs plateformes de micro-travail. Certaines plateformes ont une force de travail composée de travailleurs résidents dans le pays d'origine de la plateforme. C'est le cas pour AMT dont 47% des travailleurs viennent des É.-U., et 52 % d'Inde³⁰. C'est aussi le cas pour la plateforme allemande Clickworker qui compte une majorité de travailleurs européens dont 39 % de travailleurs allemands, 10 % de travailleurs au Royaume-Uni, 9 % en Italie et 5 % en Espagne. Prolific, est une plateforme fondée au Royaume-Uni qui attire essentiellement des travailleurs résidents de ce pays (47%) et des travailleurs basés aux É.-U. (39%). Pour deux autres plateformes, cette caractéristique est moins marquée. Sur Microworkers, on compte une majorité de résidents américains 30% en 2017, mais l'échan-

29. Ils indiquent également que certains clients ciblent les travailleurs selon leurs compétences pour des tâches en particulier. Par exemple, l'Ukraine pour les tâches de développement logiciel, le Népal pour des tâches de production de données et le Kenya pour des tâches de transcription audio (p.30)

30. Ces résultats sont à nuancer, car la proportion est respectivement de 75 et 16% d'après Difallah et al. (2018)

tillon reste plus diversifié avec la présence non négligeable de travailleurs venant : d'Inde, de Serbie, du Népal, d'Indonésie et le Nigéria (p.132). C'est encore plus le cas pour Crowdfunder avec un échantillon très diversifié et quelques pays plus représentés comme le Venezuela, la Serbie et la Bosnie.

Finalement, ces recherches ont permis d'établir un profil-type des micro-travailleurs travaillant sur des plateformes. Il s'agit d'une population qui vient des pays des Suds, jeunes, masculines (sauf dans les pays du Nord) et plutôt éduquées. Globalement, ces recherches se focalisent assez peu sur les explications du recours à ce type de population. Par exemple, peu de recherches ont montré pourquoi les travailleurs de certains pays sont impliqués plus que d'autres. En plaçant la focale sur les travailleurs et non sur les autres acteurs de la chaîne comme les entreprises clientes, ces recherches n'apportent pas d'éléments concernant les facteurs pour la mise en place de ce type d'externalisation. Enfin, ces recherches ne permettent pas de décrire ce que ces travailleurs au quotidien et donc comment ils contribuent à l'économie numérique.

2.2.2 Les conditions de travail et l'organisation du travail sur les plateformes de micro-travail

Une des principales thématiques de la recherche sur le micro-travail concerne cette fois les conditions et l'organisation du travail sur les plateformes de micro-travail et de *freelancing*. En plus de travaux généraux sur les conditions de travail, on note que plusieurs recherches spécifiques se sont focalisées sur le salaire des micro-travailleurs, ainsi que sur leurs relations sociales.

John J. Horton (2010) a d'abord tenté de déterminer si les utilisateurs d'Amazon Mechanical Turk étaient traités de manière équitable. Pour ce faire, il a mené une expérience afin de comparer la perception de l'équité des employeurs hors ligne par rapport aux demandeurs Mturk. Il a montré que *"les travailleurs Mturk considèrent que leurs chances d'être traités équitablement en ligne sont aussi bonnes ou meilleures que celles qu'ils peuvent obtenir hors ligne"*.

Dans une étude ethnographique des travailleurs d'Amazon Mechanical Turk en Inde, Gupta, Martin, Hanrahan, et O'Neill (2014) décrivent comment les micro-travailleurs organisent leurs activités sur Mturk en fonction de plusieurs éléments : leur quotidien, de l'accès aux infrastructures numériques, de leur localisation et leur expertise. Ils ont constaté que l'infrastructure technologique (vitesse et fiabilité d'Internet) et la connaissance de l'anglais ont un impact sur le type d'emplois et les revenus potentiels que les micro-travailleurs sont en mesure d'obtenir. Ils critiquent également l'argument selon lequel le micro-travail est une activité de loisir en montrant que le micro-travail est une activité qui participe substantiellement aux revenus des travailleurs.

À travers l'étude de cas ethnographique de Lisa, une travailleuse indépendante qui exerce son activité à distance pour une entreprise indienne, Koslowski (2016) montre *"que le travail virtuel s'éloigne des lieux clairement définis et se négocie plutôt dans un espace informationnel-relationnel"*. En tant qu'indépendante, elle

doit continuellement travailler avec de nouvelles personnes et utiliser des outils numériques pour maintenir le contact. Comme travailleuse "invisible" et "précaire", Lisa doit faire face au manque de confiance à propos de son travail. Pour pallier cette situation, elle met en place des stratégies lui permettant de prouver ce qu'elle fait (par exemple : en travaillant le week-end). Lisa est également censée utiliser certains outils pour interagir avec ses collègues. Les auteurs observent qu'elle fait partie d'une "communauté de pratique" en étant intégrée dans les processus de production de l'entreprise. Cet exemple est particulièrement utile pour comprendre que les travailleurs virtuels peuvent être plus ou moins intégrés dans une entreprise, mais qu'ils ne sont en aucun cas des "cavaliers solitaires" travaillant en dehors de l'organisation.

En s'appuyant sur la littérature, Webster (2016) a analysé trois caractéristiques du micro-travail : l'atomisation des travailleurs, la précarité au travail et l'absence d'organisation collective. Elle explique d'abord que le micro-travail doit être analysé à travers le prisme d'une tendance générale à l'abrogation du "lieu de travail traditionnel" qui a commencé avec les réflexions libérales sur l'organisation du travail et s'est poursuivie avec l'émergence d'Internet puis des plateformes de travail. En effet, elles permettent aux entreprises d'externaliser des services à moindre coûts (par exemple : les centres d'appels). Désormais, les entreprises utilisent un intermédiaire numérique (plateforme) pour proposer un travail médiatisé par l'écran et par le marché (dans le cas des plateformes comme Mturk). Sur le second point, elle explique à quel point le marché mondial du travail est compétitif pour les employeurs, ce qui se traduit par des revenus irréguliers pour les travailleurs. En outre, ils ne bénéficient d'aucune protection sociale. Enfin, elle explique que l'individualisation de ce travail, le manque d'espaces physiques communs et la tendance des travailleurs à intérioriser leurs problèmes rendent toute organisation collective difficile. C'est également en substance la conclusion du chapitre d'ouvrage de CHERRY (2016) qui soulève un paradoxe du *crowdsourcing*. Alors que cette activité pourrait être une grande opportunité pour de nombreux travailleurs traditionnellement exclus du marché du travail, le *crowdwork* "semble s'engager dans une 'course vers le bas' avec des normes de travail en déclin"³¹.

En s'appuyant sur deux études qualitatives sur des travailleurs américains et indiens de Mturk, Martin, O'Neill, Gupta, et Hanrahan (2016) discutent de la manière dont Amazon Mechanical Turk, une "place de marché mondial", n'est pas si désencastrée des spécialités nationales. En comparant les travailleurs américains et indiens, ils exposent les différences de vécu de ces derniers lorsqu'ils réalisent des tâches. Les travailleurs indiens doivent ainsi être plus attentifs aux fluctuations monétaires entre la roupie et le dollar. Ils rencontrent d'autres obstacles, comme ceux liés à l'accès à Internet qui les empêchent de faire "le même travail à la même vitesse". L'auteur discute également du "travail d'interprétation" réalisé par les turkers indiens pour comprendre le fonctionnement de Mturk et s'adapter aux besoins des demandeurs, la plupart du temps de manière collective via le forum.

31. À travers la notion de "responsabilité sociale des entreprises", les auteurs décrivent trois "meilleures pratiques" : des salaires équitables, un besoin de transparence des tâches et un "due process" qui signifie que l'évaluation des travailleurs "devrait être basée sur des critères équitables".

Dans une étude ethnographique, Sengul-Jones (2017) décrit des mécanismes similaires en analysant la manière dont les travailleurs indépendants définissent leur identité virtuelle en fonction des règles des plateformes et des normes sur le genre et la race. Ces normes peuvent parfois conduire à des cas de discrimination lors du choix des travailleurs. Par conséquent, ces derniers doivent partager de nombreuses informations sur eux-mêmes comme des photos, mais aussi communiquer sur leur travail (par exemple : sur Twitter ou sur un blog). Ces deux étapes, créer un profil et être actif sur les médias sociaux, sont cruciales pour obtenir plus de travail, mais elles exposent également les travailleurs à la surveillance des clients qui peuvent alors consulter ces informations supplémentaires. En utilisant les données d'une plateforme de freelancing, Niels Beerepoot et Bart Lambregts (2016) montrent d'ailleurs que les salaires des travailleurs américains et britanniques sont deux fois plus importants que les salaires indiens et philippins.

Posch, Bleier, Flöck, et Strohmaier (2017); Posch et al. (2018) ont mené une étude pour évaluer les motivations des micro-travailleurs en mesurant les raisons pour lesquelles ces derniers font des efforts dans l'accomplissement des tâches. Ils examinent ainsi l'absence de motivation (amotivation), la régulation externe matérielle (salaire), la régulation externe sociale (punitions, notes), la régulation introjectée (sentiment de valeur) et la régulation identifiée ("*alignement de l'action sur les objectifs personnels de l'individu*") et la motivation intrinsèque (plaisir de faire l'action). Ils constatent que la régulation externe matérielle était le facteur le plus important, ainsi que la motivation intrinsèque, ce qui renvoie à une tension déjà observée dans la littérature entre le micro-travail en tant qu'activité rémunératrice et donc comme *travail* et le micro-travail en tant qu'activité "*motivée par l'intérêt*" et donc comme *jeu*. Ils ont comparé les résultats des États-Unis avec ceux d'autres pays à faible revenu (le Brésil, le Mexique, l'Inde, les Philippines et le Venezuela) et des pays à revenu élevé (l'Allemagne et l'Espagne). Les travailleurs des pays à faible revenu sont plus amotivés que les travailleurs allemands et espagnols. En outre, les travailleurs de tous les pays, à l'exception de l'Allemagne, "*ont déclaré une régulation externe matérielle significativement plus élevée que les travailleurs américains, ce qui signifie que les travailleurs de ces pays sont plus motivés par les récompenses matérielles telles que l'argent gagné.*" Dans une enquête qualitative, Ling Jiang, Christian Wagner et Bonnie Nardie (2015) examinent comment les micro-travailleurs perçoivent leur activité. 86% des répondants mentionnent les récompenses monétaires comme un avantage au fait de travailler pour Mturk. Les travailleurs citent aussi des bénéfices supplémentaires "*l'amélioration de soi*" (*formation*). Les micro-travailleurs comparent leurs gains aux salaires du travail conventionnel en considérant cette activité comme "*sous-payée*". Les auteurs comparent également les travailleurs indiens et américains, et constatent que les revenus du micro-travail "*peuvent servir de source principale de revenus, en particulier pour les travailleurs des régions à faible revenu*".

D'Cruz, Mulder, Noronha, Beerepoot, et Magala (2019) décrivent les conditions de travail des freelances indiens travaillant sur la plateforme internationale appelée Upwork. Les auteurs concluent que les freelances

indiens ont *”une opinion positive de l'économie de plateforme”*, et ce malgré certains cas de discriminations subis par ces derniers de la part des clients. Cette vision positive du *”crowdworking”* est d'ailleurs dépendante de l'engagement des travailleurs : plus ils sont actifs, plus ils auront tendance à valoriser le travail de plateformes. Cependant, ils soulignent le fait que les travailleurs des plateformes sont issus de la classe moyenne qui est *”conventionnellement associée à l'emploi standard”*. Le travail de plateformes s'ajoute à des d'activités parfois réalisées de manière informelle qui ne garantissent pas la protection des travailleurs.

En s'appuyant sur les données du rapport de l'OIT de 2017, [Rani et Furrer \(2019\)](#) étudient l'expérience professionnelle sur les plateformes comme garantie d'accéder à des tâches sur les plateformes de micro-travail. Pour ce faire, elles ont d'abord classé les travailleurs en fonction de leur expérience sur les plateformes :

- 44% des travailleurs avaient commencé le crowdwork depuis moins d'un an (*”nouveaux entrants”*)
- 27% ont travaillé pendant un ou deux ans (*”travailleurs intermédiaires”*)
- 19% étaient actifs depuis trois ou quatre ans (*”travailleurs expérimentés”*)
- 10% ont commencé à faire du crowdwork depuis plus de cinq ans (*”travailleurs très expérimentés”*)

Ensuite, ils analysent les différences de revenus pour ces catégories et constatent que l'ensemble des travailleurs consacrent un quart de leur temps à des activités non rémunérées comme la recherche de tâches). Ils observent *”une amélioration substantielle”* des revenus au cours de la première année de travail sur les plateformes, les revenus moyens passant de 3,74 à 4,92 dollars américains par heure de travail rémunéré. Mais cette amélioration n'est pas mécanique et dépend de la plateforme ainsi que de la localisation des travailleurs : plus ils sont dans les pays des Suds, moins le revenu augmente. Concernant l'accès à une protection sociale, cette dernière s'améliore uniquement du fait d'une plus grande ancienneté sur les plateformes.

[Gray et Suri \(2019\)](#) étudient plusieurs plateformes : Amazon Mechanical Turk (micro-travail), Leadgenius (vente) et Amara (traduction). Les auteurs contextualisent d'abord historiquement le concept de *”dernier kilomètre de l'automatisation”*. Ce concept définit le fait que de nombreuses activités industrielles et de services sont censées être automatisées, mais qu'elles reposent en fait sur du travail humain. Ils décrivent ensuite les conditions de travail en mettant en lumière certains paradoxes problématiques spécifiques tels que : l'hypervigilance maquillée en flexibilité, l'*”isolement et [le] manque d'orientation présenté comme de l'autonomie”* au bénéfice des travailleurs.

En utilisant les résultats d'une étude à méthodes mixtes, [Wood et al. \(2019a\)](#) ont tenté d'élaborer une double approche du concept d'*”encastrement”* de Polany : l'encastrement dans les réseaux sociaux et l'encastrement normatif. La première approche fait référence aux communications interpersonnelles des travailleurs des plateformes et de leurs clients. Les auteurs ont constaté qu'*”en l'absence d'institutions économiques formelles”*, les travailleurs s'appuient sur leurs réseaux interpersonnels pour trouver des tâches et créer de la confiance avec les clients. La deuxième approche fait référence aux normes culturelles et juridiques qui

s'appliquent sur un territoire, autrement dit à la réglementation du travail. Dans le cas du micro-travail, l'absence de réglementation gouvernementale et les stratégies des plateformes pour se présenter comme des entreprises technologiques conduisent à un "désencastement normatif". Comme le disent les auteurs : *" malgré le fait que le travail reste encadré dans les réseaux interpersonnels des travailleurs, il est en même temps désincarné des normes culturelles et juridiques qui limiteraient sa marchandisation."*

À l'intersection des questions de développement et de qualité de l'emploi, [Anwar et Graham \(2019\)](#) ont mené une étude qualitative sur les conditions de vie et de travail des travailleurs africains d'Upwork³². Ils analysent quatre dimensions du travail numérique : la liberté, la flexibilité, la précarité et la vulnérabilité. Ils constatent que les travailleurs de la plateforme peuvent gagner de meilleurs salaires que les agents des centres d'appels qui sont salariés, mais qu'ils doivent payer leurs ordinateurs et leurs connexions à Internet, ce qui est très coûteux dans beaucoup de pays du continent africain. Un autre aspect apprécié par les travailleurs est la flexibilité. Cependant, cet aspect du travail sur plateforme peut également conduire à la solitude au travail. De plus, cette organisation du travail médiatisée par le marché engendre aussi un manque de pouvoir de négociation pour les travailleurs, surtout en l'absence de syndicats. [Moreschi, Pereira, et Cozman \(2020\)](#) ont quant à eux étudié les travailleurs brésiliens de Mechanical Turk. Ils montrent les mauvaises conditions de travail des turkers brésiliens s'accompagnent de faibles revenus malgré une charge de travail élevée et l'exposition à du stress et à de l'anxiété. Par ailleurs, l'absence de toute réglementation fait que les travailleurs sont parfois exposés à des contenus violents ou pornographiques sans avoir accès à des services de soutien notamment psychologique. Ils ont également observé comment les travailleurs s'organisent par le biais de groupes WhatsApp sur lesquels ils partagent leurs difficultés.

[Hara et al. \(2018\)](#) ont collecté des données sur les activités des travailleurs sur Amazon Mechanical Turk³³. Ce faisant, ils fournissent un aperçu très intéressant sur les salaires. Globalement, ils estiment que *"96% des travailleurs sur AMT gagnent moins que le salaire minimum fédéral américain"*. Le salaire horaire médian est de deux dollars par heure, ce qui est plus bas que l'estimation proposée par [Hara et al. \(2018\)](#) de \$3.13/heure en moyenne. Concernant le salaire, une méta-analyse de la littérature de [Hornuf et Vrankar \(2022\)](#) montre que le salaire horaire des micro-travailleurs est de 4,43 \$ de l'heure. Les auteurs constatent également que *"les demandeurs paient 11,58 dollars de l'heure en moyenne"*, la plupart des demandeurs actifs postant des tâches à bas salaire. Ils ont également fourni une analyse des sujets de tâches montrant quel type de tâches sont effectuées sur AMT comme création de contenu (320.220 HIT), l'interprétation et analyse (72.932 HIT) ou les enquêtes (47.192 HIT). En outre, ils ont constaté que le thème des tâches est corrélé avec un paiement plus ou moins élevé. La transcription de données est par exemple peu rémunérée alors que l'évaluation de vidéos l'est davantage. Ils étudient également les trois sources de travail non rémunérées : "recherche de

32. Une plateforme permettant à des clients d'acheter des tâches aussi différentes que du travail des données ou du développement web

33. Après d'un échantillon de 2676 travailleurs

tâches”, ”rejet de tâches” et ”retour de tâches”. Cette deuxième dimension de leur travail montre la variété des tâches réalisées sur les plateformes de micro-travail et qui dépassent le cadre de l’IA. D’ailleurs les tâches liées à l’IA ne sont pas forcément les mieux rémunérées.

La question du travail non rémunéré a fait l’objet de plusieurs analyses. Cet élément du de l’activité plateformisée implique des tâches aussi larges que la communication avec les clients, la recherche de tâches ou bien le fait de réaliser des tâches rejetées par le client ou des tâches non complétées. Par exemple, selon la plateforme, jusqu’à 15% de l’ensemble des tâches sont rejetées (Berg et al., 2018). La quantité estimée de travail non rémunéré dans le crowdworking varie également en fonction du type de crowdwork. Les chiffres varient également selon les études³⁴, mais dans leur méta-analyse, Hornuf et Vrankar (2022) montrent qu’entre 23 et 26% du travail n’est pas rémunéré. Ce point est confirmé par Howson, Johnston, et al. (2022) qui indiquent que ”le temps de travail non rémunéré n’est pas l’exception à la règle, mais plutôt une partie essentielle du modèle de plateforme et un indicateur du degré d’asymétrie du pouvoir géographique.” des chaînes de valeur des données.

En France, l’étude ethnographique de Lagerie (2019) analyse le phénomène du crowdsourcing à travers la perspective de la ”commodification du temps”. Pour ce faire, ils étudient comment les travailleurs investissent leur temps sur une plateforme de crowdsourcing française, montrant une variété d’investissements temporels, de la ”marchandisation du temps perdu” (pendant un voyage) à la ”marchandisation du temps libre” (micro-travail en regardant une émission de télévision) et la ”marchandisation de la disponibilité” (micro-travail pendant les activités domestiques). Ils ont également soulevé le paradoxe de l’utilisation du forum interne par les microtravailleurs : certains participent activement à la ”communauté” et ”perdent” ainsi du temps, mais pour tous les microtravailleurs, c’est aussi un moyen d’obtenir des conseils sur les tâches et/ou les demandeurs. Ils ont également montré que les micro-travailleurs ont le sentiment d’être exploités (”temps exploité”), ce qui renvoie à des tâches mal rémunérées du fait d’une sous-estimation du temps nécessaire à leurs réalisations.

Si le temps passé à communiquer sur les tâches entre micro-travailleurs a été considéré comme du travail rémunéré, il s’agit aussi d’une thématique qui a été analysée sous l’angle des relations sociales au travail. Yin, Gray, Suri, et Vaughan (2016) ont évalué si les crowdworkers collaborent ou non et pourquoi. Ils ont constaté que les crowdworkers collaborent ”avec les membres de leur réseau” (via un forum en ligne, avec des amis) pour trois types d’activités : ”gérer les frais administratifs associés au crowdwork”, ”trouver des tâches lucratives et des employeurs réputés”, ”recréer les liens sociaux et le soutien souvent associés aux environnements

34. D’après Hornuf et Vrankar (2022) : ”Rani et al. (2021) rapportent que 38% de tout le travail effectué sur les plateformes de freelancing et 33% du travail effectué dans le cadre du microtasking ne sont pas compensés. Dans une autre étude, Rani et Furrer (2021) estiment que 23% de tout le travail n’est pas rémunéré, tandis que Wood et al. (2019a) rapportent que 39% du travail effectué par les crowdworkers n’est pas rémunéré.”

de travail traditionnels". Dans une autre publication, ils constatent que 13,4% des travailleurs sont connectés et communiquent principalement par le biais d'un forum en ligne, ce qui signifie au contraire qu'une grande majorité est complètement indépendante.

Christine Gerber a mené une enquête qualitative (entretiens et observation) sur, "les pratiques technologiques et organisationnelles concrètes derrière les stratégies de construction de communauté des plateformes" et sur "comment les crowdworkers interagissent avec ces stratégies". L'article identifie deux types d'approches managériales pour façonner le comportement des travailleurs. La première, "loose community" vise à "activer l'auto-assistance et l'auto-régulation de la foule" qui fournit des espaces de communication libres de toute gestion. La seconde appelée "communauté contrôlée" qui encadre techniquement et par la modération l'interaction des travailleurs afin de susciter des comportements spécifiques. Dans la deuxième partie de cet article, l'auteur suggère qu'il y a plus d'interaction sur les espaces de communication de la "communauté libre". Elle montre aussi que les espaces de communication officiels sont utilisés par les modérateurs des plateformes pour traiter les plaintes.

Un des enjeux majeurs des discussions sur les relations sociales dans le micro-travail concerne les capacités de négociation des travailleurs. Par exemple, pour répondre au problème de "l'invisibilité des travailleurs sur Amazon Mechanical Turk", Irani et Silberman (2013) proposent de mettre en plus un outil appelé le *Turkpticon*. Cet outil se présente comme une extension de navigateur sur laquelle les travailleurs peuvent fournir des informations sur les demandeurs. L'objectif est de permettre aux travailleurs de noter les demandeurs afin de contrebalancer les asymétries de pouvoir. Lehdonvirta (2016) a analysé 30 entretiens réalisés avec des microtravailleurs des Philippines, du Népal et des États-Unis pour étudier comment les caractéristiques des plateformes créent une absence de lieu et une dispersion des collectifs qui ont un impact sur la formation de l'identité des travailleurs et sur le processus de régulation du travail par l'organisation collective. Ses conclusions s'articulent autour de deux thèmes clés : la production technologique de l'absence de lieu et la "dispersion des rôles et des identités dans le marché du travail du micro-travail". En plus de provenir de divers contextes socio-économiques, la formation de l'identité des micro-travailleurs en ligne est également "façonnée" par les recommandations des moteurs de recherche entraînant un manque d'identité collective ce qui affecte les possibilités de négociations collectives.

Sur le même sujet, Jamie Woodcock (2018) a décrit le paysage du "travail numérique" en soulignant la diversité des activités (de la programmation de logiciels au micro-travail) et des modes d'organisation de ces activités. Il prend ensuite l'exemple de l'absence de processus de syndicalisation dans le secteur de la programmation de logiciels (par exemple : dans l'industrie des jeux vidéo) pour appeler à l'organisation du travail dans d'autres formes de travail numérique, en faisant valoir que tous ces processus de travail sont liés. Décrivant l'expérience de "faircrowdwork.org", un site web d'évaluation des conditions de travail dédié aux travailleurs portés par le syndicat allemand IG Metal, Harmon et Silberman (2018) montrent

comment cet outil produit de manière interdisciplinaire pourrait être un outil intéressant pour discuter, négocier et régler le "futur du travail". Wood, Graham, Lehdonvirta, et Hjorth (2018) ont constaté que l'organisation collective en ligne des freelances repose sur des groupes de médias sociaux plutôt que sur des forums. En effet, 58% des répondants déclarent communiquer numériquement avec d'autres travailleurs en ligne alors que 78% des répondants déclarent ne communiquer que rarement ou jamais en face à face. En outre, ils n'ont trouvé aucune preuve de l'existence d'un processus de syndicalisation en cours. En effet, un ensemble de valeurs entrepreneuriales telles que l'autonomie est très présente dans l'identité des freelances, ce qui rend plus difficile le développement des syndicats.

Margaryan (2019) compare l'utilisation des activités d'apprentissage sur le lieu de travail et les stratégies d'apprentissage autorégulatrices entreprises par les microtravailleurs et les indépendants. Elle constate que l'apprentissage semble se produire dans les deux activités. En outre, bien que ces activités soient fragmentées, ces stratégies d'apprentissage sont collectives et pas seulement individuelles. En outre, les deux types de travailleurs indépendants développent des stratégies d'apprentissage autorégulatrices en "investissant du temps" et des "ressources financières dans l'amélioration de leurs compétences".

On l'a vu, de nombreuses recherches ont analysé les conditions de travail des micro-travailleurs. Cette littérature a permis de souligner à quel point le numérique et l'intelligence artificielle s'appuient sur des travailleurs souvent précaires qui exercent la plupart du temps depuis chez eux, avec leurs propres matériels. Initialement, ce travail est d'ailleurs présenté comme un "complément de revenu" qui serait même perçu par les travailleurs eux-mêmes comme un "jeu". D'autres analyses ont depuis montré l'importance de ces revenus pour les travailleurs de pays à bas revenus. Il s'agit également d'un travail substantiel pour les industries à haute valeur ajoutée auquel il contribue, ainsi que par le contenu parfois en fait plus complexe qu'il n'y paraît.

En définitive, nous posons comme hypothèse que si le salaire est micro, le travail en lui-même ne l'est pas. Ces recherches ont également contribué à critiquer l'émergence des plateformes numériques de travail et pointant du doigt certaines caractéristiques problématiques de celles de micro-travail et de freelancing. Par exemple, la possibilité pour le client de rejeter une tâche. On peut toutefois regretter que trop peu de publications replacent les travailleurs dans le processus de production d'ensemble de services numériques en détaillant de manière plus fine ce que les travailleurs font au quotidien. Finalement, si la plateforme est largement traitée dans sa relation avec les travailleurs, ce sont les clients finaux de ces tâches qui sont relativement peu documentés et qui restent donc très mal connus.

2.2.3 Questionner le micro-travail comme vecteur de développement pour les pays du Sud global

La question des impacts de micro-travail dans les pays en voie de développement a été largement discutée dans la littérature. Dès l'origine, cette manière d'organiser le travail est inspirée du micro-crédit. En échange de la production de tâches numériques, des travailleurs situés dans des pays en voie de développement bénéficient d'un programme de formation au numérique et à l'entrepreneuriat (*impact sourcing*). L'objectif est donc de proposer des activités d'insertion par le travail indépendant en ligne tout en aidant les travailleurs à créer leurs propres projets professionnels de long terme. Ce principe a donné des idées aux acteurs du développement qui se sont intéressés aux facteurs facilitant l'installation de ce type d'activités, ainsi qu'à ses vertus et ses aspects négatifs. Des acteurs académiques ont également remis en question les effets de ces modèles de développement basés sur des plateformes de travail.

Au niveau transnational, un rapport de la Banque Mondiale (Kuek, Paradi-Guilford, Fayomi, Imaizumi, Ipeirotis, Pina, & Singh, 2015), présente les opportunités permises par la sous-traitance en ligne, notamment pour résoudre le problème du chômage des jeunes dans les pays en voie de développement. Un précédent rapport de la Banque mondiale identifie également les bénéfices potentiels en matière de développement pour ce type de travail pour les territoires palestiniens (Paradi-Guilford, 2013). Ces deux documents considèrent que le micro-travail a le potentiel de réduire le chômage, mais qu'il reste des questions sociales importantes à résoudre, notamment les faibles rémunérations des micro-travailleurs et l'absence de sécurité d'emploi. Dans le rapport consacré au développement du micro-travail dans les territoires palestiniens, les auteurs estiment encore que les territoires palestiniens pourraient atteindre entre 21 000 et 54 000 équivalents temps partiel.

En plus du rapport de la Banque Mondiale précédemment cité, l'OIT a également souligné l'impact positif de ce type d'activités (Berg et al., 2018; Kuek, Paradi-Guilford, Fayomi, Imaizumi, Ipeirotis, Pina, & Singh, 2015). Ainsi, les revenus issus de ce travail sont affectés en priorité à des dépenses de première nécessité comme se nourrir. Ils permettent encore à certains travailleurs d'accéder à un logement ou de se former. Les travailleurs des plateformes peuvent par exemple développer des compétences techniques et non techniques (*soft skills*), d'améliorer leurs perspectives de carrière ou de se construire un réseau professionnel. Cependant, selon ces recherches, il est nécessaire d'accorder une attention particulière aux effets négatifs potentiels, notamment en matière de conditions de travail, de volatilité du flux de travail, de risques psychosociaux et de dérégulation du marché du travail, afin d'attirer ce type de travail. Du point de vue de la rémunération, le rapport indique que certains travailleurs trouvent qu'elle est faible et inéquitable, alors même qu'elle constitue un élément substantiel de leurs revenus. L'accès à la protection sociale est rendu difficile par le statut d'indépendant³⁵.

35. Les auteurs du rapport mettent d'ailleurs en évidence un lien de corrélation entre dépendances au travail des plateformes et les difficultés d'accès à la protection sociale. Ce résultat tend à montrer que les plateformes de travail ne sont pas des vecteurs

D'autres recherches analysent les facteurs d'installation de ces activités dans les pays en voie de développement. S. R. S. Aris et al. (2013) étudient par exemple la capacité de la Malaisie à accueillir ces activités. Ils montrent que la mise en place de "systèmes de réputation" (notes et commentaires sur les travailleurs) est obligatoire pour garantir le lien de confiance entre les travailleurs et les donneurs d'ordres. Par ailleurs, la facilité d'utilisation de la technologie par les travailleurs et la formation aux technologies sont nécessaires. N.H. Arshad et al. (2013) ont réalisé une analyse qui porte sur l'opportunité du développement d'activités de micro-sourcing en Malaisie³⁶. Du point de vue des entreprises clients, la présence de "directives gouvernementales existantes sur le travail à domicile" en Malaisie, également applicables au micro-sourcing permettent d'encadrer ces activités et favorisent donc l'installation d'entreprises de micro-travail. Cependant, le manque de sensibilisation à ces activités et les doutes sur la qualité et la fiabilité des travailleurs constituent des faiblesses pour la Malaisie. En ce qui concerne les travailleurs, il y a une réserve de main-d'œuvre potentielle importante, notamment parce que de nombreux ménages à faibles revenus ont un accès direct ou indirect aux ordinateurs et à Internet. Enfin, le faible niveau de compétences, notamment en anglais, constitue une sérieuse faiblesse. Du côté des plateformes proposant des tâches, le manque d'accès aux systèmes de paiement est également problématique pour recruter des travailleurs³⁷.

Helani Galpaya, Suthaharan Perampalam et Laleema Senanayake (2018) ont travaillé sur le potentiel de l'externalisation en ligne au Sri Lanka, qui est déjà une destination privilégiée pour les activités de type BPO. Par le biais d'un questionnaire et de groupes de discussion, ils analysent la sensibilisation des actifs de la population sri-lankaise pour accueillir ces activités. Les auteurs indiquent que 25% de la population est consciente des possibilités de travail en ligne et que 9% seraient prêtes à s'y engager. Les travailleurs valorisent la flexibilité du micro-travail, car cela leur permet de maintenir une autre activité et d'éviter de perdre tous leurs revenus, s'il n'y a pas de tâches sur les plateformes. La question des "avis" (*système de réputation*) est soulignée comme un signal de qualité indispensable. Par ailleurs, le système de paiement en ligne est défaillant puisque Paypal est toujours interdit dans le pays. L'étude menée par Fareesa Malik, Brian Nicholson et Richard Heeks (2017) s'intéresse aux implications et aux obstacles au développement de l'externalisation en ligne au Pakistan. Ils montrent que les travailleurs sont motivés par la nécessité de gagner leur vie, tant les autres possibilités d'emploi sont contrastées en raison de la situation de la région étudiée, en proie à des conflits, des catastrophes naturelles et une topographie montagneuse qui empêche l'implantation d'autres industries. Les autres obstacles posés par ce type d'activité sont non seulement physiques (infrastructures électriques et de réseau, système de paiement), mais aussi humains (division

d'accès au marché de l'emploi formel.

36. Le micro-sourcing se définit comme l'externalisation d'une tâche hyper-spécialisée à un travailleur, ce qui englobe donc les activités de freelancing et le micro-travail

37. Jabu Mtsweni et Legand Burge (2014) étudient également les avantages potentiels du micro-travail mobile pour les pays en développement. Ils notent également que l'accès aux ordinateurs et aux smartphones et le manque d'infrastructures de paiement sont des obstacles importants à l'émergence du micro-travail dans les pays en développement.

générée des rôles, vision culturelle de ce qui constitue un travail significatif). Ils montrent que les plateformes font des concessions, notamment salariales en faveur des travailleurs, du fait de l'intervention des ONG, des gouvernements, des institutions internationales et le développement de "réseaux d'assistance personnelle" parmi les travailleurs.

Dans un article sur les freelances et les micro-travailleurs d'Asie du Sud-Est et d'Afrique subsaharienne, Mark Graham et al. mettent en évidence quatre risques liés au travail numérique (Graham, Hjorth, & Lehdonvirta, 2017b). D'abord, les travailleurs ont un faible pouvoir de contestation à la fois de par le déséquilibre entre l'offre et la demande de travail, qui entraîne une concurrence entre les travailleurs, et de par la décentralisation des normes locales qui devraient normalement "réguler les relations de travail". Le deuxième point concerne l'exclusion économique de certains travailleurs, notamment à l'encontre des travailleurs du continent africain qui subissent des discriminations au point de les obliger à mentir sur leur localisation. Au sujet de la captation de la valeur créée par ces activités, les auteurs observent un phénomène de réintermédiation par les plateformes, leur permettant alors d'accaparer une partie de cette valeur, au détriment des travailleurs. Enfin, sur le développement des compétences, les auteurs constatent que les plateformes permettent aux travailleurs de développer des compétences et d'accéder à des activités à plus forte valeur ajoutée. Cependant, les travailleurs demeurent éloignés du cœur d'activité de l'entreprise donneuse d'ordre.

Peu de travaux permettent finalement d'appréhender les effets négatifs et positifs de cette activité sur les travailleurs et les pays hôtes. Si l'on excepte les quelques travaux focalisés sur les travailleurs (Graham et al., 2017a; Idowu & Elbanna, 2019a; Malik et al., 2017; Roomaney, Van Belle, & Tsibolane, 2018; Schriener & Oerther, 2014), il est clair que la recherche manque d'études empiriques consacrées à cette thématique, pourtant déjà traitée dans d'autres industries considérées comme "extractives".

* * *

À travers cette revue de littérature, nous avons proposé de situer l'objet de recherche "travail des données" à l'intersection de deux champs d'études : un premier qui porte sur l'étude de l'intelligence artificielle et un second qui porte sur les plateformes numériques de travail. La recherche sur le travail des données s'est en effet d'abord constituée autour du terme "micro-travail". Derrière ce terme, elle s'est principalement intéressée à l'externalisation de tâches d'annotation et de saisie de données des clients (entreprises ou chercheurs) à une foule de travailleurs indépendants, le tout coordonné par des plateformes numériques de travail. L'enjeu a alors été de caractériser ce nouveau mode de distribution du travail en insistant sur le fait qu'il correspond à un élargissement des "périphéries de l'emploi", c'est-à-dire un travail réalisé soit gratuitement soit en étant peu rémunéré et peu protégé. C'est finalement à travers des travaux comme ceux de A. A. Casilli (2019); F. A. Schmidt (2019); Shestakofsky (2017); Tubaro, Casilli, et Coville (2020a) que la recherche a fait directement le lien entre ce que font les micro-travailleurs et le développement de

SIA, rattachant dès lors cet objet à une critique plus générale de l'intelligence artificielle. Pour conclure ce chapitre de revue de littérature, nous revenons sur la pertinence du concept de plateforme, puis présentons nos perspectives de recherches à la lumière des recherches sur l'intelligence artificielle.

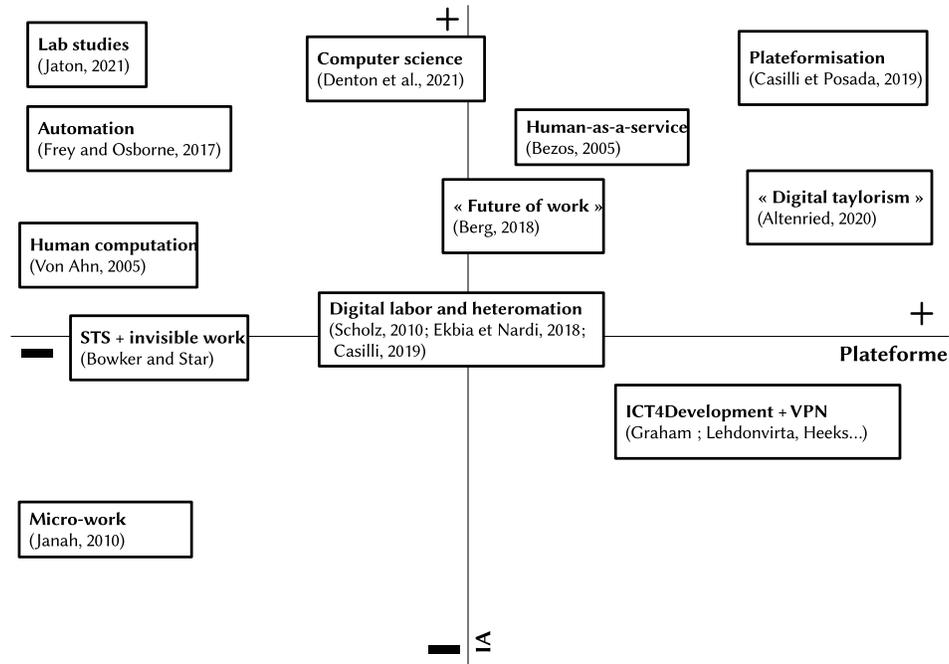


FIGURE 6 – Cartographie des concepts présents dans littérature scientifique lié au micro-travail. Lecture : les concepts sont classés selon qu'ils évoquent davantage les plateformes ou l'IA

Dans cette revue de littérature, nous avons constaté trois lacunes de la recherche sur le micro-travail. Peu de recherches sont consacrées à d'autres pays donneurs d'ordres que les É.-U. et plus généralement à d'autres aires linguistiques que l'anglais. Ensuite, la littérature s'est beaucoup intéressée aux relations entre les travailleurs et les plateformes, mais très peu aux clients de ces dernières. Enfin, beaucoup d'études portent sur le modèle de plateforme "Mturk" et sur le travail gratuit. Pourtant la littérature a commencé à analyser d'autres formes d'organisation du travail qui se situeraient entre la forme pure du marché (Mturk) et la forme "entreprise". [Lehdonvirta, Kässi, Hjorth, Barnard, et Graham \(2019\)](#) a par exemple suggéré que les PNT sont fondamentalement différentes des BPO. Les premiers permettent aux entreprises clientes de distribuer des projets divisés en tâches plus petites à un grand nombre de prestataires de travail indépendants, tandis que les seconds permettent aux clients d'externaliser des projets entiers. Les avancées récentes de la littérature montrent qu'il existe des organisations hybrides; elles se présentent comme des PNT, mais intègrent des niveaux de gestion et de supervision comme dans les BPO et les entreprises conventionnelles. Selon [Tubaro \(2021\)](#), ces organisations hybrides fournissent du "travail profond", c'est-à-dire celui de la plateforme très ouverte qui se rapproche le plus de la place de marché.

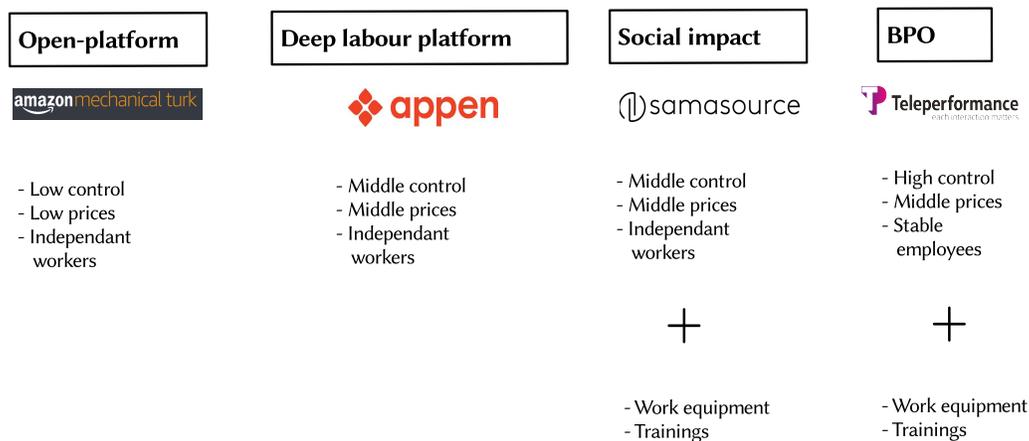


FIGURE 7 – Modes d’externalisation du travail des données

Le traitement du travail des données dans la littérature s’inscrit donc dans une remise en question plus générale du recours à des plateformes numériques de travail. Cette thématique de recherche a été l’objet de travaux antérieurs consacrés à la question du statut des travailleurs de plateformes (Régniez et al., 2020). L’enjeu était d’envisager des alternatives plus protectrices pour ces travailleurs (par exemple : le salariat). Cette question, tant juridique que sociologique, est actuellement en passe d’être tranchée par la Commission Européenne. En effet, l’objectif de la proposition de directive européenne *”relative à l’amélioration des conditions de travail dans le cadre du travail via une plateforme”* est de faciliter la requalification des travailleurs des plateformes en salariés, principalement pour les plateformes de mobilité comme Uber ou de livraison comme Deliveroo. Toutefois, la question n’est pas encore résolue pour les micro-travailleurs qui risquent d’être écartés de cette directive.

Nous avons également pu saisir la variété des plateformes et la nécessité d’analyser ce phénomène en tenant compte des différences d’organisation du travail et de la variété des secteurs. De ce point de vue, nous considérons que le concept de *”plateforme”* capte très imparfaitement un ensemble de réalités sociales liées plus globalement à l’externalisation et à la numérisation du travail et qui concourt à cette *”extension des périphéries”* de l’emploi (Berg, 2019). Aussi, il paraît utile de ne pas souligner le caractère exceptionnel du mode d’organisation platformisé, mais plutôt d’en montrer la variété. Une manière d’éclairer cette question revient finalement à se demander : est-il possible de catégoriser sous une même appellation des travailleurs ayant des rythmes différents, des activités différentes et opérant dans des secteurs variés et sur des plateformes qui ne se ressemblent jamais les unes aux autres (Cornet, Le Ludec, Wahal, & Joulin, 2022). Cela revient alors à s’interroger non pas sur une éventuelle *”platformisation”*, mais plutôt d’analyser ces différentes entreprises sous l’angle du degré d’intégration des travailleurs des données dans le processus de production de donneur d’ordres.

C'est aussi dans le cadre de recherches antérieures réalisées sur les micro-travailleurs qu'émerge l'hypothèse de chaînes de sous-traitance internationale en matière de travail de données. Les travaux précédemment cités montrent que le travail des plateformes "non localisées" est mondialisé et suit un axe Nord / Sud, avec d'un côté les donneurs d'ordres et de l'autre les travailleurs. Ces éléments de la littérature rendent donc plus que probable l'idée que les entreprises françaises s'appuient sur une main-d'oeuvre étrangère, présente notamment dans les pays d'Afrique francophones. Une partie de la littérature a analysé ce phénomène à travers le concept de "Virtual production network" (VPN) (Wood, Graham, Lehdonvirta, & Hjorth, 2019b). À travers cette idée, cette recherche présente les travailleurs des plateformes comme *désencastrés* des normes locales de travail de leur pays d'habitation. Toutefois, ce concept pose plusieurs questions. Est-ce que cela s'applique indifféremment à tous les travailleurs des données ou du numérique? Est-ce que ce "désencastrement" est uniquement le fait des plateformes numériques de travail? Est-ce que cela a même du sens de parler de désencastrement, alors que dans certains à bas revenus, le travail informel est tellement développé que les statuts d'emplois facilitent justement le travail de plateforme sans protection sociale? Enfin, si l'on postule que les chaînes de valeur de l'IA correspondent à des aires linguistiques, pourquoi insister sur le caractère virtuel de chaînes de production alors qu'elles se superposeraient en fait à l'histoire coloniale des anciens pays colonisés?

De son côté le champ d'études sur les algorithmes et l'intelligence artificielle a finalement très peu tenu compte du travail des données. Certes, de récentes recherches en sciences informatiques portent sur l'archéologie des grands ensembles de données. Ces travaux montrent de nombreuses erreurs d'annotation dans les ensembles de données "canoniques" de l'industrie de l'IA et les impacts potentiels sur les biais algorithmiques (Denton et al., 2020a; Paullada et al., 2020), mais reste encore embryonnaire. La grande majorité de la littérature en sociologie de l'IA, s'est au contraire plutôt focalisée sur les figures du concepteur, de la machine et de l'utilisateur. Au départ, certaines recherches ont questionné la notion même d'intelligence de ces machines en soulignant la différence entre humains et machines dans la capacité à faire "automatiquement" des actions (H. Collins, 1998). Plus récemment, c'est l'opacité de ces systèmes qui a été traitée, sous un angle technique et plus marginalement organisationnel. L'intérêt de ces recherches est d'avoir proposé des cadres d'analyse et de régulation pour rompre avec cette dissimulation des SIA. Toutefois, ces recherches disent finalement peu de choses sur la fabrication des modèles d'IA.

D'autres travaux placent la focale sur les utilisateurs. Dans le domaine du travail, c'est le "management algorithmique" qui a fait l'objet de nombreuses analyses. Ces dernières montrent à la fois les conséquences sur les travailleurs en matière d'asymétrie d'information, mais aussi certaines pratiques de "braconnage" qui visent à comprendre et à détourner ces algorithmes pour en tirer profit. Dans tous les cas, ces recherches et notamment l'approche juridique montrent à quel point il s'agit en fait de décisions managériales humaines. Toujours focalisés sur les utilisateurs, nous avons évoqué d'autres travaux sur la *non-neutralité* des

SIA. L'intérêt principal est de montrer le lien entre les choix de conception et les effets sur les utilisateurs : exclusion de l'usage du système, mésusage, biais algorithmiques. Enfin, nous avons évoqué la littérature qui porte sur la contribution des micro-travailleurs à l'intelligence artificielle. Certains de ces travaux montrent bien le travail essentiel qui est réalisé pour rendre possible l'automatisation. C'est tout l'argument de Crawford (2021) qui résume : *"l'IA n'est ni artificielle ni intelligente. Au contraire, l'intelligence artificielle est à la fois incarnée et matérielle, faite de ressources naturelles, de carburant, de travail humain, d'infrastructures, de logistique, d'histoires et de classifications."* Crawford (2021).

Finalement, le problème n'est pas dans la définition de l'intelligence, ou même dans les "biais" de l'IA des acteurs qui participent à la production des modèles, mais plutôt dans la division du travail de l'IA. Nous proposons alors d'analyser la succession des étapes permettant de définir et de "résoudre" une question posée par l'automatisation.

Notre thèse s'intéresse quant à elle aux SIA déployés de manière opérationnelle et non aux modèles "scientifiques", comme on l'a souvent retrouvé dans la littérature. Contrairement au monde académique, nous pensons que les SIA commerciaux ont un mode de fabrication et de fonctionnement propre, qu'il est nécessaire d'étudier. Par ailleurs, nous considérons que l'analyse des évolutions du travail liées à l'IA ne pourra se faire que dans l'étude minutieuse de ces systèmes et de leur déploiement. L'objectif de cette thèse est de combler ces lacunes de la littérature en contribuant à une meilleure connaissance des impacts de la technologie sur le travail. Ainsi, comment les services commerciaux basés sur l'IA sont-ils fabriqués? Quelles sont les conséquences du développement de ces systèmes sur le travail?

Pour répondre à cette problématique, nous proposons d'articuler trois niveaux d'analyse. Tout d'abord, l'étude du *travail de l'IA* en tant que pratique et en tant que mode d'organisation du travail. Comme pratique professionnelle d'abord, afin d'analyser les étapes qui permettent de produire des modèles d'IA : qu'est-ce que le travail du clic? Qu'est-ce que l'annotation implique en termes de discussion? Comment les annotateurs interagissent-ils avec d'autres travailleurs de la chaîne? Comme organisation du travail ensuite, car le travail de l'IA s'est structuré selon une division internationale du travail, qui passe par l'externalisation de certaines activités : comment et pourquoi ce travail est-il externalisé? Comment les entreprises gèrent-elles l'externalisation de certaines parties de leurs activités? *In fine*, qui profite de cette sous-traitance?

Un deuxième niveau d'analyse porte sur l'étude des *conséquences de l'IA sur le travail*. Nous avons déjà souligné les lacunes de la littérature concernant cette dimension. On y retrouve principalement des débats sur la fin du travail, sur les reconfigurations organisationnelles sous-jacentes à la pénétration de l'IA dans l'économie et sur l'évolution des métiers. En termes de questionnement, cela provoque des débats tels que : les avocats vont-ils être remplacés par des IA? Comment l'administration publique se transforme sous l'effet de l'intelligence artificielle? Comment les agents publics vont devoir s'adapter à l'IA? Si nous ne

critiquons par l'orientation générale de ces questions, nous considérons qu'elles basent trop souvent sur des craintes plutôt que sur des réalités³⁸. Nous proposons de contribuer à ces débats à travers deux approches. La première se propose de vérifier empiriquement deux hypothèses qui lient travail et automatisation : la fin du travail et la déqualification. La seconde aborde plus particulièrement la question des conditions de travail et des impacts de l'externalisation de certaines activités dans des pays en voie de développement, en l'occurrence Madagascar.

Un dernier niveau d'analyse porte sur *ce que le travail de l'IA explique du fonctionnement de l'intelligence artificielle*. Dans cette dernière thématique, nous proposons d'apporter notre pierre à la littérature qui cherche à définir ce qu'est l'IA, en l'abordant sous l'angle du travail. De H. M. Collins (1992) à Crawford (2021), la littérature en sciences sociales s'est positionnée frontalement contre les discours techno-utopistes sur l'automatisation, que ce soit en critiquant les capacités des modèles d'IA ou en soulignant le travail humain qui est nécessaire pour automatiser. Ces critiques se sont aussi traduites par des réflexions sur la *non-neutralité* des SIA et par des propositions de régulation dont l'objectif est d'encadrer le fonctionnement de ces systèmes et leur biais. Dès lors, on se demandera : qu'est-ce qu'*automatiser* veut dire ? Qu'est-ce que les machines peuvent ou ne peuvent pas faire ? Comment réguler l'IA au regard du travail humain nécessaire à son fonctionnement ?

Mis à part la typologie de Tubaro, Casilli, et Coville (2020a), d'autres recherches ont tenté de faire le lien entre l'organisation du travail d'annotation et les résultats des modèles (Miceli, Schuessler, & Yang, 2020a). En creux, cela souligne bien critique du travail des données. La démonstration de l'importance de ces travailleurs fait également l'objet de plusieurs analyses en sciences informatiques. Ainsi, Emily Denton et ses collègues (2020b) proposent quant à elles un programme de recherche pour ramener "les gens" dans les discussions sur l'intelligence artificielle en documentant tout le travail nécessaire pour créer des jeux de données pour l'apprentissage automatique. Amandalynne Paullada et al. (2020) vont plus loin en soulevant des inquiétudes sur la façon dont les scientifiques ou les entreprises collectent les données, en particulier sur le fait que le jugement des annotateurs de données peut entraîner un biais intégré dans un grand système socio-technique. Notre thèse vise précisément à contribuer à documenter empiriquement les pratiques quotidiennes des entreprises d'annotation de données.

En matière de sociologie de la classification, le travail de Bowker et Star (1999) constitue alors un cadre essentiel pour comprendre les classifications se créent. Ils y définissent la classification comme *"une segmentation spatiale, temporelle ou spatio-temporelle du monde"*. Les classifications, nous expliquent les auteurs, doivent répondre à certains critères pour fonctionner. Tout d'abord, il ne doit pas exister d'ambiguïté dans la classification : *"les catégories sont mutuellement exclusives"*. En tant qu'annotateur d'images d'animaux,

38. Comme en témoigne d'ailleurs la présentation que nous avons fait de l'étude de Stanford sur la pénétration de l'IA dans l'économie et qui montre que l'IA se déploie dans finalement peu de secteurs.

je dois pouvoir ranger les images d'animaux dans des catégories qui ne se recoupent. Mais, ce qui paraît évident pour les animaux l'est peut-être déjà moins pour des images de vêtements. Comment distinguer indubitablement certains types de cols, ou certaines coupes? Avant même l'annotation, comment même créer ces catégories. Par ailleurs, il faut également que le système soit complet, ce qui signifie que la classification fournit une *"couverture totale du monde qu'il décrit"*. En tant qu'annotateur, que faire avec ce qui est *"laissé dans l'ombre"* par les classifications de ceux qui produisent les modèles. Que faire avec cette les données qui rentrent dans la fameuse catégorie "autre"? Enfin, les classifications fonctionnent souvent avec des normes qu'elles contribuent également à établir. Ces normes sont des ensembles de règles qui relient des communautés hétérogènes et sont souvent contrôlées, jusqu'à même parfois être inscrits dans la loi. La question qui se pose alors consiste à se demander qui produit ces normes dans le cadre des SIA. Les développeurs? Les clients des entreprises d'IA? Ou bien ce sont peut-être aussi les travailleurs des données qui participent à ce travail de classification?

Chapitre 3 - Les coulisses de l'accès à un terrain *multi-situé*

Au tout début de cette thèse, peu d'études analysent alors l'utilisation de services de travail de la donnée par les entreprises. Un rapport de Corporaal et Lehdonvirta (2017) étudie l'utilisation des plateformes de "freelancing" par les grandes entreprises américaines. Le travail de Van Belle, Bonne, Cocquyt, et Garbutt (2019) porte quant à lui spécifiquement sur la manière dont les entreprises belges considèrent le recours aux services de micro-travail et identifie de fortes contraintes en matière de sécurité des données et de qualité du travail. Ils constatent surtout que seulement 4,5% de leur échantillon a déjà externalisé des activités liés au micro-travail. Dans le cadre de l'enquête DIPLab, nous avons participé à un entretien avec le dirigeant d'une entreprise d'IA française proposant des assistants vocaux (A. Casilli et al., 2019a). Lors de l'entretien, ce dernier nous explique les différentes stratégies d'externalisation mises en place, afin de trouver des locuteurs parlant différentes langues (français, anglais et même japonais) en ayant recours à différents services parfois situé à l'étranger. L'idée de l'existence d'une chaîne de valeur délocalisée servant au travail de la donnée commençait alors à émerger, tout en étant inexploré par la littérature.

Notre thèse s'appuie sur cette hypothèse, ainsi que sur l'intérêt d'étudier les clients des entreprises qui vendent des services liés au travail des données. La présence de ces chaînes de valeur nous semblait alors offrir une possibilité encore peu exploitée jusqu'alors d'analyser les deux "bouts" de cette chaîne : les clients finaux (entreprises d'IA) et les travailleurs de la donnée. Sans vraiment de certitude sur notre capacité à identifier et à analyser ces chaînes, nous avons décidé de procéder en plusieurs phases, opérant des allers et retours entre les différents acteurs de cette chaîne.

La conséquence de cette approche par les chaînes de production est la réalisation d'un terrain "multi-situé". Il concerne des organisations très différentes situées dans des pays différents, mais pourtant connectées entre elles par la réalisation d'un produit commun basé sur l'IA. Dans ce chapitre, nous verrons que cette approche par les deux "bouts" la chaîne a été fructueuse. Dans certains cas, c'est à travers des start-ups d'IA que nous sommes parvenu à accéder aux travailleurs des données. Dans d'autres cas, c'est à travers les entretiens avec les travailleurs que nous avons pu accéder à certains types de donneur d'ordres (par exemple : les grandes entreprises françaises).

Dans ce troisième chapitre, nous présentons les orientations méthodologiques de notre thèse et notamment sa dimension "multi-située". La première section revient sur notre travail exploratoire, ainsi que sur l'enquête réalisée auprès d'employés d'entreprises d'IA françaises. Dans une deuxième section, nous présentons le terrain réalisé à Madagascar. L'objectif de cette section est de raconter les coulisses de notre enquête en expliquant les négociations pour l'accès à notre terrain, notamment avec les entreprises d'annotation. Enfin, dans une dernière partie, nous détaillons les différentes techniques d'enquêtes mobilisées.

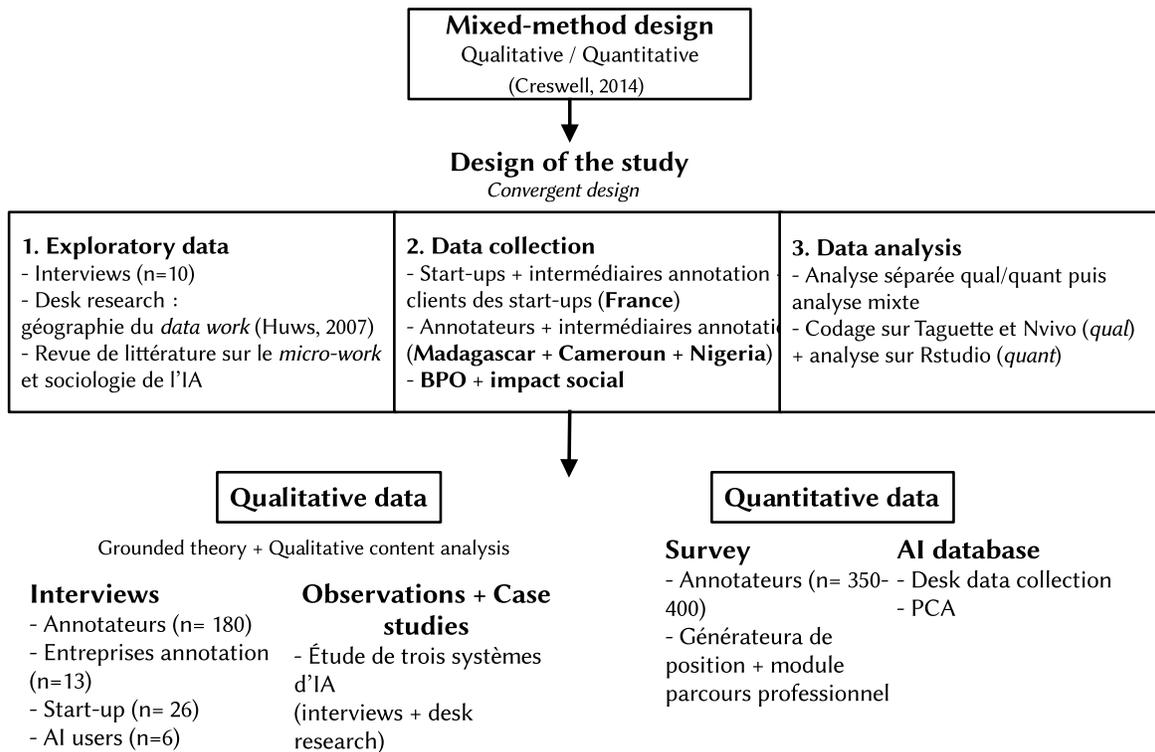


FIGURE 8 – Schéma général de notre méthodologie

3.1 Un premier terrain "français" pour comprendre les pratiques d'externalisation du travail de l'IA

Dès le départ de cette étude, une première difficulté émerge quant au cadrage de notre objet ainsi qu'à l'accès à notre terrain. Face au manque d'éléments empiriques à propos de l'externalisation du travail des données et plus généralement des services informatiques dans l'espace francophone en comparaison avec l'espace anglophone, comment identifier les terrains potentiels ? Comment accéder à des travailleurs reconnus dans la compe étant largement invisible (Irani, 2013) ? Comment avoir accès aux entreprises d'IA en dépit de la culture du secret qui entoure le travail des données ?

Un premier travail a consisté à identifier des pays et des services susceptibles de fournir des services liés à l'intelligence artificielle en menant un travail exploratoire. Par la suite, nous avons testé plusieurs approches pour collecter des données sur l'ensemble de la chaîne de valeur, la première étant "par le haut" et est passée par la collecte de données auprès d'entreprises françaises. Nous verrons que cette approche s'est avérée très fructueuse, puisqu'en plus de mieux comprendre les ressorts de l'externalisation, cela nous a également

permis d'accéder à des entreprises d'annotation.

De manière très concrète, ce terrain s'est déroulé en plusieurs phases. Premièrement, un travail exploratoire, a permis de tester plusieurs pistes et de rencontrer quatre informateurs privilégiés venant de Madagascar (Miary, Éric et Andry) et de l'univers des start-ups d'IA (Julie) qui s'avèreront décisifs pour l'enquête pour : l'aide au recrutement de travailleurs à Madagascar, la relecture et la traduction partielle du questionnaire en malgache. Une série d'entretiens avec des start-up françaises ont ensuite permis de tester l'hypothèse de Madagascar et de trouver des contacts d'entreprises d'annotation.

3.1.1 Un travail exploratoire qui a permis de cadrer notre objet en écartant certaines pistes présentes dans la littérature

La littérature sur l'externalisation de services numériques offre de précieuses orientations pour identifier des terrains d'études potentiels. Plusieurs travaux analysent le développement des tâches liées aux données et aux services informatiques en soulignant la dynamique Nord/Sud de ces échanges économiques (N. Coe, 1997; Congress of the U. S., 1985; Huws, 2007; HUWS, 2014; Mitter & Pearson, 1992), par exemple, entre les États-Unis et la Jamaïque, les Philippines ou l'Inde. Cette tendance est confirmée par la littérature sur les plateformes de tâches en ligne (Graham et al., 2017b).

La littérature souligne également que la connaissance de langues étrangères comme l'anglais ou le français était un facteur déterminant pour l'attraction d'activités de services numériques. Penny Gurstein (2002) observe ainsi deux tendances dans l'externalisation de travail de données : une externalisation domestique avec l'atypisation du travail (contrats courts, missions réalisées hors du salariat...) et une externalisation dans des pays proches culturellement ou géographiquement (frontière commune ou langue commune). Dans une enquête sur les transactions réalisées sur un site de freelancing, Gefen et Carmel (2008) montrent que dans l'ensemble, les clients sont réticents à sous-traiter à l'étranger alors que les clients des É.-U. sont plus disposés à le faire, mais uniquement dans des pays proches géographiquement et culturellement, comme le Canada. En dehors des États-Unis, les pays de langue anglaise de l'échantillon (Irlande, Royaume-Uni, Australie, Inde...) sont plus prompts à externaliser lorsque celle-ci a lieu au sein du même groupe de pays. Enfin, le prix n'est globalement pas un facteur déterminant. C'est plutôt l'existence de précédentes relations entre clients et entreprises sous-traitantes qui déterminent le choix des clients. Cet article supporte donc l'idée que les coûts de transaction sont plus importants lorsqu'un contrat concerne des entreprises de deux pays différents. Il montre également que plus la distance culturelle est grande, moins le contrat a de chances d'être conclu.

La littérature consacrée au développement de l'externalisation de services informatiques suggère encore cette dimension en identifiant différentes formes d'externalisation (Gonzalez, Gasco, & Llopis, 2006) : on-

shore c'est-à-dire à l'intérieur des frontières du pays d'origine de la firme, *nearshore* c'est-à-dire dans un pays proche géographiquement ou limitrophe, *offshore* c'est-à-dire dans un pays éloigné géographiquement, mais proche culturellement, *farshore* c'est-à-dire dans un pays éloigné géographiquement et culturellement. Ils indiquent encore que l'Inde est devenu une destination d'importance pour le *offshoring*, du fait de sa proximité culturelle avec les É.-U. et le Royaume-Uni qui se traduit par la maîtrise de la langue anglaise. Les pays de l'Europe de l'Est comme la République Tchèque, la Pologne, la Hongrie et l'Ukraine sont également devenues des pays permettant le *nearshoring* ou le *offshoring*, notamment pour l'Allemagne, et ce, du fait de leur proximité culturelle avec les pays européens de l'Ouest, d'une connaissance linguistique, de bonnes compétences techniques. Pour les États-Unis, nous avons déjà évoqué les deux destinations *nearshore* que sont le Canada et le Mexique. Le cas de l'Irlande est également souligné en tant que destination *nearshore* pour les États-Unis et le Royaume-Uni du fait de sa proximité culturelle. Leur analyse illustre une régionalisation des échanges de services informatiques facilitée par l'existence de liens culturels voire de relations post-coloniales.

Notre projet de recherche étant centré sur l'externalisation du travail de données opéré par les entreprises francophones, il s'agissait donc de donner du sens à cette littérature en construisant notre terrain de recherche. Notre intuition initiale était que le travail des données fait partie d'un secteur plus large de l'externalisation de services. Cela nous a conduits à mener un premier travail exploratoire portant sur :

- l'analyse des exportations de services numériques des pays d'Afrique francophones, utilisées comme proxy de l'implantation d'entreprises d'annotation de données
- dix entretiens à la fois avec des employés de start-ups françaises, avec des travailleurs du numérique de Madagascar

Concernant les exportations de services numériques, nous avons comparé seize pays d'Afrique francophones et de l'océan Indien, qui comptent entre 6 et 73% de locuteurs français³⁹. Ce travail, basé sur les données de l'Organisation Mondiale du Commerce (2017), fait ressortir Madagascar comme étant le deuxième pays en termes d'exportations de services informatiques (73 millions \$), derrière le Maroc (639 millions \$) et devant la Tunisie (69 millions \$) et l'Île Maurice (52 millions \$) et loin devant des pays comme l'Algérie (39 millions \$) ou le Sénégal (20 millions \$)⁴⁰. Ce travail montre que Madagascar possède une industrie de l'externalisation de services numériques très importante. Entre 2019 et 2020, nous réalisons deux entretiens avec Éric. À cette époque, il est manager dans une entreprise d'externalisation de services d'assurance. En parallèle, il fonde son entreprise de sous-traitance. Composée de 40 salariés, cette entreprise travaille exclusivement pour la start-up *Automatik* pour laquelle elle labellise des images. L'analyse de ces entretiens confirme également deux enjeux particulièrement importants de l'externalisation : la maîtrise de la langue et la connaissance du

39. Algérie, Bénin, Burkina Faso, Burundi, Cameroun, Congo, Côte d'Ivoire, Madagascar, Mali, Maroc, Maurice, Mauritanie, Niger, Rwanda, Sénégal

40. L'utilisation de ce proxy, peut être critiquable notamment si l'on se place dans la perspective d'un travail platformisé qui est alors parfois très informel et donc mal déclaré dans ces chiffres.

pays d'origine de l'entreprise donneuse d'ordre. Ainsi, en ce qui concerne Madagascar, Éric, nous indique que c'est grâce à leur accent peu repérable que les travailleurs malgaches attirent des investisseurs, notamment dans l'industrie des centres d'appels⁴¹. Au-delà de l'accent, Éric nous indique que ce que les investisseurs cherchent, c'est une main-d'oeuvre nombreuse et suffisamment qualifiée en français. Cette demande pousse même certains malgaches à passer un diplôme de français⁴².

C'est aussi parfois la connaissance du pays d'origine du donneur d'ordres qui est demandée lors du recrutement, connaissance facilitée dans ce cas par les relations historiques et diplomatiques entre la France et Madagascar. L'analyse des données qualitatives, que ce soit les entretiens avec des start-ups ou avec des dirigeants d'entreprises d'externalisation présentes à Madagascar confirmera le fait que la Grande Île est un pays majeur de l'externalisation de services numériques et en particulier pour le travail des données.

En plus de ces deux entretiens avec Éric, la littérature sur l'externalisation de services nous a donné des indications sur les caractéristiques des pays concernés par l'attraction de ces industries. Ainsi, la proximité culturelle et la langue sont des facteurs qui expliquent l'externalisation entre les deux pays. Par conséquent, la liste de pays potentiellement concernés dans le cas de la France est forcément limité par le nombre de locuteurs français. Ces éléments se verront être confirmés par nos premiers entretiens exploratoires réalisés avec des travailleurs de Madagascar. Sur la base de ces entretiens, nous avons ensuite fait l'hypothèse que le secteur de l'externalisation du travail des données s'inscrivait dans des dynamiques commerciales pré-existantes en matière de délocalisation d'activités de services informatiques. Ces données de l'OMC nous ont alors permis de confirmer Madagascar comme un candidat sérieux pour l'attraction de l'industrie des données.

3.1.2 Une collecte de données sur la chaîne de valeur de l'IA pour mieux comprendre les acteurs et leurs rôles

Pour comprendre la place du micro-travail dans la chaîne de valeur de l'IA et le processus de production de l'IA, nous avons d'abord mis en place deux stratégies de recherche complémentaires. Tout d'abord, nous avons mené des *desk analysis* sur 22 entreprises d'IA. Ces entreprises ont été sélectionnées sur la base de deux études de marché dédiées au marché de l'annotation de données. L'objectif de cette première phase était d'identifier les activités nécessaires à l'intelligence artificielle en analysant la proposition commerciale de ces entreprises et leur utilisation d'annotateurs de données (micro-travailleurs) pour l'IA.

41. "Le principal concurrent de Mada, c'était le Maghreb. Mais, les clients français n'étaient pas satisfaits de l'accent des conseillers dans le Maghreb alors qu'à Mada si quelqu'un parle très bien français on ne fait pas de différence."

42. "certains se sentent obligés de passer le diplôme de l'Alliance française. Beaucoup de jeunes étudient dans établissements catholiques ou dans des établissements où la maîtrise n'est pas top. Si tu ne passes pas par un lycée français tu vas avoir du mal à parler la langue française."

Sur la base de cette première phase exploratoire, nous avons constitué une base de données de 127 entreprises du secteur de l'IA en utilisant des registres d'entreprises tels que Crunchbase.com ou Owler. Cela nous a permis de rechercher les concurrents des entreprises de notre analyse initiale par itération. Entre juin et juillet 2020, nous avons systématiquement inclus des concurrents et des concurrents de concurrents dans notre base de données jusqu'à ce que nous trouvions les mêmes entreprises. Pour atténuer les biais éventuels, cette base de données a été complétée par une base de données d'entreprises d'IA plutôt axées sur l'annotation à partir de données collectées en mai 2020 selon la méthode de la "boule de neige"⁴³.

Sur la base de nos études de cas, nous avons ensuite codé de manière binaire notre base de données en fonction des caractéristiques des entreprises. Un premier ensemble de variables décrit le type d'activités proposées par l'entreprise : "collecte de données", "outils d'annotation", "outils de développement", "développement de modèles", "puissance de calcul", "optimisation de modèles", "hébergement de données". Le second ensemble de variables décrit le recours à une main d'œuvre d'annotation : "workforce", "impact sourcing", "crowd-sourced", "outsourced", "in-house". Cette base de données nous a permis d'utiliser des méthodes de classification, notamment l'analyse de correspondance multiples.

L'objectif de cet exercice était de pallier un déficit d'informations provenant initialement d'acteurs commerciaux, dont la spécialité est de produire des rapports d'intelligence économique sur des secteurs économiques particuliers. Ces informations nous paraissaient pertinentes, puisqu'elles permettaient d'obtenir une première liste d'entreprises de production des données et une première segmentation des acteurs de la chaîne. Néanmoins, du fait de leur origine axée sur des objectifs non scientifiques, il nous semblait pertinent de reproduire ce type de travail d'une manière plus systématique. Les résultats de ce travail sont présentés dans le chapitre cinq.

3.1.3 Un premier terrain pour comprendre l'organisation et les pratiques d'annotation des entreprises d'intelligence artificielle françaises

Entre février 2020 et novembre 2021, nous avons ensuite conduit 25 entretiens avec des employés appartenant à des start-ups françaises (Table 37). Le premier objectif de ces entretiens était de collecter des données sur la manière dont les entreprises externalisent ou non le travail des données et la manière dont ils interagissent avec leurs travailleurs délocalisés ou non. Par ailleurs, nous souhaitions mieux comprendre le positionnement de ces entreprises. Quels sont leurs modèles d'affaires? Quels sont les produits qu'ils proposent? Quels sont leurs clients? Font-ils appel à des partenaires ou à d'autres services, notamment pour la partie hébergement de données? Comment développent-ils leurs modèles?

La grille d'entretien a été construite en fonction de ces thématiques. Il en résulte entre trois et six blocs de

43. Cette phase de collecte de données a été réalisée par Mandie Joulin, sous notre direction

questions qui ont évolué en fonction de la position de la personne enquêtée dans l'entreprise et qui portent sur :

1. la trajectoire professionnelle de la personne
2. l'histoire de l'entreprise
3. l'acquisition des clients et la nature des projets
4. le processus de production du modèle d'IA
5. l'externalisation du travail des données
6. les partenariats avec d'autres entreprises (par exemple : pour l'hébergement de données)

Ces entrées de terrain ont été obtenues par différents moyens. Nous avons tout d'abord mobilisé notre réseau personnel ce qui nous a permis par effet boule de neige d'accéder à d'autres entreprises d'IA et à des entreprises d'annotation. Certains enquêtés ont également été directement ajoutés et contactés par le biais de LinkedIn. Nous avons également pris contact avec l'entreprise Automatik, directement depuis leur site web. Ces derniers ont largement contribué au travail d'enquête en nous permettant d'interroger plusieurs employés et en nous donnant accès à leur partenaire malgache pour l'annotation. Enfin, dans le cadre d'un cours donné dans une école d'ingénieur, nous avons été mis en contact avec leur incubateur de start-up ce qui nous a permis d'obtenir des contacts supplémentaires.

Ce premier terrain sur les start-ups d'IA nous a également permis de trouver des sous-traitants en matière de travail de la donnée. Notre objectif n'était pas de réaliser étude systématique sur les entreprises d'IA françaises, ni sur les grandes entreprises qui développent des projets d'IA. Ces dernières sont d'ailleurs difficiles à étudier dans leur ensemble. En effet, une entreprise peut avoir une petite équipe qui travaille sur un sujet lié à l'IA, sans que l'ensemble de l'entreprise soit au courant de ce qu'il s'y passe (Burrell, 2016). Une piste de recherche future pourrait être de réaliser un questionnaire auprès des start-up françaises d'IA, par exemple en diffusant un questionnaire par le biais d'une association professionnelle comme France Digitale qui publie tous les ans une cartographie des start-ups françaises d'IA.

Néanmoins, cette méthode ne permettrait pas de capter les entreprises qui n'appartiennent pas au secteur technologique, mais qui mettent en place des projets d'IA (les grandes entreprises), ce que notre terrain a permis de capter à travers les entretiens réalisés avec les travailleurs de Madagascar. L'apport de notre approche est donc d'avoir su traiter de la chaîne de valeur sur plusieurs "sites" d'enquête, nous permettant de reconstituer des chaînes, parfois en passant par les start-ups d'IA, parfois en passant par les dirigeants des entreprises d'annotations et dans d'autres cas encore en passant par les travailleurs des données. Enfin, comme on peut le voir dans le tableau ci-dessous, il faut souligner que notre échantillon reste diversifié en termes de type de données utilisées et de taille. En revanche, nous avons rencontré beaucoup plus d'entreprises spécialisées dans un seul produit, plutôt que des "agences d'IA" proposant des services sur mesure selon les cas rencontrés chez leurs clients.

Nom	Type de données	Nombre d'employés	Type de produit
ModeAI	Photos	-10	Spécialisée
MLDupli	Photos / Textes	10	Spécialisée
MLArchitecture ²	Hébergement données	4000	Infrastructure
CityVeillance	Photos / Vidéos	50	Agence d'IA
Automatik	Photos	60	Agence d'IA
Clik	Photos	10	Agence d'IA
AIVoice	Son	80	Spécialisée
CallAI	Son	-10	Spécialisée
Papers	Textes	-10	Spécialisée
LegalAI ¹	Textes	20	Spécialisée
CCenter ²	Son	6000	Spécialisée
AIMedic ¹	Images / Vidéos	65	Spécialisée
BankAI	Textes	-10	Spécialisée
GitAI	Tout	155	Infrastructure logiciel

¹Cette entreprise a depuis été rachetée par une multinationale

²Cette entreprise n'est pas française au contraire de toutes les autres entreprises de notre corpus

TABLE 1 – Description des entreprises d'IA étudiées

La première partie de notre enquête a eu une influence déterminante sur l'accès aux travailleurs des données, puisque nous avons constaté que les entreprises d'IA étudiées ne passaient pas par des plateformes, mais plutôt par des entreprises formelles situées à Madagascar. En effet, nous nous attendions à des formes d'externalisation plus proches de ce qui a été étudié dans la littérature, notamment à travers des plateformes comme Foule Factory ou Amazon Mechanical Turk. Ce modèle, "*platform-as-market*" permet aux chercheurs d'avoir accès à des travailleurs sans nécessairement avoir besoin d'en négocier l'accès, puisque ces entreprises permettent de réaliser des questionnaires. Toutefois, il ne s'agissait pas d'un modèle très utilisé par nos enquêtés. Par conséquent, il a été nécessaire de "négocier" un accès au terrain, en passant par les dirigeants d'entreprises d'annotation et par des clients de ces services.

3.2 Le terrain en Afrique francophone : l'accès aux entreprises d'annotation et à leurs travailleurs

Dans cette deuxième section, nous racontons les coulisses de notre enquête en insistant sur nos difficultés lors de l'accès aux travailleurs des données. En effet, une deuxième approche de notre objet a consisté à plu-

tôt passer directement par des entreprises d'annotation et par des travailleurs à Madagascar. Nous verrons que, malgré la "culture du secret" caractéristique d'une industrie qui invisibilise ses travailleurs, nos différentes entrées se sont avérées fructueuses. Cette section porte également sur le montage du projet, depuis la négociation de l'accès au terrain, jusqu'à l'accès aux travailleurs. Cette phase de l'enquête a d'ailleurs plus compliqué à mettre en place du fait de la crise de Covid qui nous a empêchés de nous rendre sur place durant plus de deux ans.

3.2.1 A l'autre bout de la chaîne de valeur : négocier un accès aux intermédiaires d'annotation et à leurs travailleurs

La mise en place de l'enquête avec les travailleurs a fait l'objet d'une négociation avec plusieurs acteurs et d'un montage administratif permettant le bon déroulement de notre travail. De manière générale, la "négociation" avec des acteurs du monde de l'entreprise est consubstantiel des enquêtes sur le travail. Ainsi, dès la fin de la Seconde Guerre Mondiale l'ouverture de champs spécialisés en *sociologie du travail et de l'entreprise* rend nécessaire ces accès⁴⁴. Dans les *digital studies*, Alexandre Mallard (2014) décrit les différentes phases de l'étude des technologies de l'Information et de la Communication (TIC) appliquée aux "univers productifs". Il souligne notamment le développement d'enquêtes sur les relations entre les entreprises et les clients, par exemple à travers la figure du centre d'appel avec des enquêtes sur les conditions de travail en France (Buscatto, 2002b; Jaeger, 2002) et dans des pays hébergeant des activités de centre d'appels pour la France (Tine, 2018). La littérature française et internationale consacrée à la figure des modérateurs peut également être mobilisée pour comprendre les enjeux de négociation de terrain (Beunet, 2013; ROBERTS, 2019). Ces différentes recherches s'appuient principalement sur des données qualitatives, et pour certaines sur des approches ethnographiques "incognito", par exemple en se faisant recruter par ces entreprises (Beunet, 2013; Buscatto, 2002a; Tine, 2018). Cette démarche d'enquête est malheureusement impossible pour nous, puisqu'il aurait été très compliqué de se faire embaucher comme annotateur à Madagascar, bien qu'elle aurait certainement eu des vertus en matière de dévoilement des conditions de travail. Par conséquent, il était nécessaire de trouver d'autres approches.

Par ailleurs, le fait de "négocier" avec des entreprises n'est pas anodin puisqu'elle contribue à conditionner le déroulement de l'enquête. Récemment, le premier numéro de la revue *Cambouis* s'est spécifiquement intéressé à ce sujet dans le contexte des enquêtes ethnographiques. Le texte d'introduction du numéro soulignant que la "négociation d'entrée est une étape charnière [...], car elle conditionne la possibilité même de l'enquête" Aubry, Kuehni, et Scalabrini (2021). Il semble cependant particulièrement important de s'appuyer sur cette littérature et sur pour éclairer cette phase de l'enquête, qui constitue en tant que telle des

44. Lucie Tanguy (2011) souligne l'inscription de la sociologie du travail dans des objectifs politiques et économiques de "productivité" qui incite le ministère du Travail d'alors à créer un Institut des sciences sociales du travail (ISST) ce qui rend possible aux sociologues l'accès aux entreprises et aux financements.

données mobilisables dans la compréhension des mécanismes d'externalisation du travail des données. Bi-zeul (1998). C'est tout le sens d'un article de Muriel (Darmon, 2005) qui considère que : *"le temps de l'enquête ne se réduit pas aux moments évidents et explicites de recueil des matériaux, mais, dans un mouvement inverse, redéfinit comme matériau potentiel tout événement qui prend place lors de l'enquête. Ce type de posture conduit à accorder une place à l'analyse de ce qui pourrait sembler n'être qu'une condition de possibilité de l'enquête et du début des « choses sérieuses » : les négociations de terrain et leur résultante, acceptation ou refus."*

Dans leur introduction, Marion (Aubry et al., 2021) disent encore l'épaisseur de ce que le chercheur peut mettre derrière le terme de négociation. Il s'agit d'échanges entre différents acteurs qui permettent d'obtenir un accord sur la présence du chercheur sur le terrain⁴⁵. La "présence" des chercheurs se traduit également par le fait de rendre publics ses résultats, qui peut aussi faire l'objet de négociations, et ce, surtout si le terrain est considéré comme "difficile" ou secret. Sur ce point, la littérature sur l'univers de la modération, de l'externalisation de services informatiques et du "micro-travail" offre des prises intéressantes pour penser cette phase de l'enquête. Ainsi, nous considérons s'agit d'un terrain "difficile" au regard de différentes caractéristiques. L'"invisibilité" des travailleurs au sein de la chaîne de production a parfois rendu difficile le travail de négociation avec des start-ups ou des entreprises d'annotation. Ces dernières ne souhaitant pas nécessairement dévoiler le mode de fonctionnement de leur entreprise. En annexe, nous ajoutons par exemple la réponse à un mail reçu en vue de la diffusion du questionnaire qui illustre la méfiance des entrepreneurs de la sous-traitance⁴⁶ (Voir : l'annexe B.1). Enfin, l'organisation du travail qui repose sur des chaînes "d'externalisation en cascade" Appay (1998) créé possiblement des conditions de travail précaires dont les travailleurs n'ont pas nécessairement envie de rendre compte.

En suivant la littérature sur le micro-travail, il a été décidé de payer les répondants du questionnaire et des entretiens à un taux légèrement supérieur au SMIC français. Le fait de compenser le temps des répondants est assez classique dans la littérature sur le micro-travail et fait même l'objet d'importantes discussions à propos des paiements optimaux, c'est-à-dire qui limitent les biais tout en étant le plus éthiques possible Williamson (2016). Cette littérature discute notamment les impacts sur la qualité et le nombre de réponses, ainsi que sa pertinence pour des questions d'éthique.

Partant du constat que le secteur de l'intelligence artificielle s'appuie sur des populations précaires peu ou pas rémunérés, il nous semblait évident de compenser les travailleurs pour leur participation à l'enquête. Plus globalement, la question du paiement des personnes enquêtées est aussi discutée en sciences sociales

45. "Le terme de « négociation » [désigne] les processus de communications, de transactions et d'échanges entre [le] chercheur-euse et son terrain dont l'objet concerne une entente ou un accord relatif à l'établissement et aux modalités d'une relation d'enquête [et] pour rendre leur présence possible. Ils et elles peuvent solliciter une autorisation formelle ou un accord informel, exposer leurs questions de recherche et leurs méthodes d'enquête ou rester discrets et discrètes, estimant que toute formalisation des raisons de leur présence nuirait à l'enquête" (Aubry et al., 2021)

46. Il a tout de même été possible de réaliser deux entretiens, un avec le CEO et un autre avec un chef de projet à Madagascar

que ce soit pour des entretiens ou des questionnaires (Hammett & Sporton, 2012; Head, 2009; Heberlein & Baumgartner, 1978). Face aux principes de participation volontaire à la science et au risque d'influence sur les résultats, il convient de préciser que le fait de compenser les personnes enquêtées est consensuel chez les chercheurs qui étudient le travail des données et des plateformes (Molina, Tubaro, Casilli, & Santos-Ortega, 2023). En effet, dans ce contexte, les chercheurs demandent du temps à des travailleurs qui serait utilisé par ces derniers pour gagner de l'argent. Par conséquent, il paraît logique de ne pas faire peser une partie du coût de la recherche sur ces derniers. Cette décision a eu déjà des impacts sur le déroulement de l'enquête, puisqu'il a fallu trouver un moyen administratif de payer des travailleurs dans des pays du continent africain.

Une des solutions proposées par l'administration de l'école était de donner des tickets cadeaux, pour des enseignes comme la FNAC, solution vite écartée tant elle paraissait inappropriée (voire indécente), au regard de la situation des pays ciblés. Une autre solution parfois employée, a été d'utiliser notre salaire pour payer certains travailleurs. Toutefois, cette solution n'était pas viable du fait du volume de répondants, que ce soit au questionnaire et aux entretiens. Une dernière solution a été de passer par des entreprises d'annotation ou des start-ups, ainsi que par un prestataire présent à Madagascar qui a permis le recrutement et le paiement des travailleurs à Madagascar. Aussi a-t-il fallu démarcher des entreprises et donc négocier nos entrées dans ces entreprises d'annotation.

Une première convention de recherche a été passée entre notre école, une start-up d'IA française (Automatik) et son prestataire d'annotation à Madagascar (Entreprise de sous-traitance A). Automatik est une des premières entreprises pour laquelle nous avons la certitude qu'elle externalise du travail des données à Madagascar, aussi nous avons pris contact avec eux, via le formulaire de contact de leur site Internet afin de leur proposer de collaborer à une enquête sur leurs collaborateurs. Par ce biais, nous avons été mis en contact avec Julie, chargée de la RSE qui s'occupe de l'impact environnemental de l'entreprise ainsi que des externalités sociales négatives et qui se pose alors des questions concernant les conditions de travail des travailleurs de leur sous-traitant. En échange d'un financement servant à payer les annotateurs, nous nous sommes engagés à rendre un rapport à usage interne sur les conditions de travail chez leur sous-traitant. Le même procédé a été utilisé pour l'entreprise de sous-traitance B. Concernant l'enquête auprès de l'entreprise G et celle sur Facebook, nous avons passé un contrat prévoyant le paiement des annotateurs participant à l'enquête. Dans la convention, plusieurs éléments encadrent le déroulement de l'enquête et l'utilisation des données, en termes d'anonymisation, de propriété et d'utilisation des données :

- *"L'école s'engage à ne pas divulguer les données brutes récoltées dans le cadre la recherche à un tiers ou aux autres parties de la convention."*
- *"La start-up s'engage à ne pas réclamer la propriété des données brutes récoltées dans le cadre de ce partenariat."*

- *”L’école s’engage à remettre à la start-up un rapport de recherche à usage interne”*
- *”La start-up s’engage à respecter l’indépendance et la neutralité des chercheurs de l’école, et ne bénéficie pas d’un droit de regard ou de validation des productions écrites dans le cadre des travaux scientifiques réalisés par les chercheurs de l’école”*

Concernant l’entreprise G, nous avons ajouté et contacté son fondateur via LinkedIn. Ce dernier a rapidement accepté un entretien, moment qui a également permis de discuter d’une éventuelle passation du questionnaire auprès de ses salariés, qu’il a fini par accepter en échange d’un paiement du temps passé par les annotateurs. Pour ce qui est de l’entreprise B, c’est le CEO d’une start-up, qui lors d’un entretien, a évoqué cette entreprise et nous a donné le contact du responsable de la partie ”data” de l’entreprise. Un premier échange avec ce dernier s’est révélé infructueux, puisqu’il n’a pas donné suite à notre proposition d’enquête. C’est lors d’un autre entretien, cette fois avec le responsable ”data” d’une start-up d’IA médical, que nous obtiendrons une autre mise en contact, qui cette fois s’avèrera plus fructueuse. Ainsi, il accepte le principe et propose une rencontre à Paris avec lui et son CEO afin de discuter des modalités de l’enquête. Dans les deux cas, le caractère scientifique de l’enquête et notre rattachement académique dans une grande école d’ingénieur ont pu faciliter les discussions.

La passation du questionnaire via des groupes Facebook s’est déroulée différemment. Contrairement aux autres ”terrains”, l’objectif était d’accéder à des travailleurs des données par d’autres biais, sans l’accord intermédiaire de leurs employeurs. Ainsi, nous avons rémunéré une entreprise pour recruter et payer des travailleurs de Madagascar, en diffusant des annonces sur Facebook ([voir : annexe B.4](#)). Par la suite, nous avons trié les CV en incluant les personnes ayant eu des expériences dans l’annotation ou la saisie de données ou occupaient ces emplois au moment de la diffusion de l’annonce. Nous avons choisi d’étendre à la saisie de données, ainsi qu’à la modération de contenus étant donné la porosité des activités que ce soit dans les entreprises de BPO ou sur les plateformes. Pour les entretiens, nous avons également inclus d’autres activités comme les *call centers* de manière à avoir des points de comparaison.

Certains entretiens ont aussi été négociés de manière opportuniste avec des travailleurs de Madagascar, du Cameroun et du Nigeria. Pour les premiers, l’accès au terrain a été conclu sur LinkedIn, en prenant contact avec des travailleuses exerçant sur une plateforme d’annotation de données françaises (N=17). Par ailleurs, certains contacts travailleurs de Madagascar ont été obtenus par le biais d’autres travailleurs (N=4), mais également par le biais de dirigeants d’entreprises rencontrés à Madagascar (N=3). Concernant le Nigeria, il s’agit d’entretiens réalisés par le biais d’une étudiante nigériane, présente dans un de nos cours au sein d’une école de commerce spécialisée dans l’IA et qui avait déjà fait de l’annotation de données.

	A	G	B	Fb
Type contrat ¹	Convention	Contrat	Convention	Contrat
Passation	09/21-04/22	01/21-02/22	01/22-05/22	05/22-10/22
Modalités	Mixte	Mixte	Mixte	Mixte
Réponses totales ²	84	47	122	39
Réponses partielles	8	4	129	1
Réponses complètes	76	47	80	38
Nombre de salariés	120	60	300	X

¹ La convention prévoyait la remise d'un rapport à l'entreprise

² Réponses partielles et complètes sans les doublons

TABLE 2 – Calendrier et modalités de la collecte de données principale

Si notre manière de procéder pour l'accès au terrain se justifie par les impératifs de l'enquête, elle n'est pas sans conséquence potentielle et sans questionnements quant à notre propre positionnement vis-à-vis des entreprises étudiées. De notre point de vue, il est très étonnant que des entreprises aient autant ouvert leur porte lors de l'enquête, malgré l'existence d'une convention de recherche. Avant de revenir sur les risques de ces accords, demandons-nous comment expliquer et donner du sens à l'accord de ces entreprises ?

Pour nous, un premier élément réside dans notre position au moment de l'enquête : doctorant à l'Institut Polytechnique de Paris et précédemment rapporteur au Conseil national du numérique, institution qui dépend du ministère de l'Économie, des Finances et de la Souveraineté Industrielle et Numérique. Ces éléments biographiques sont susceptibles d'avoir rassuré les entreprises contactées, voire même de les avoir encouragés à nous répondre du fait du prestige de Polytechnique en France. A cet élément général, il faut ajouter des éléments propres à chacune des entreprises. Concernant *Automatik*, la négociation s'est déroulée avec la responsable RSE de l'entreprise avec comme objectif affiché d'en savoir plus sur les conditions de travail de leur prestataire à Madagascar. Chez eux, comme chez l'entreprise B, cet objectif était prépondérant et faisait d'ailleurs partie de la négociation, puisque nous devons fournir un rapport "interne" à ces deux entreprises. Dans d'autres entreprises, nous avons pu constater l'intérêt qu'avaient les dirigeants à rencontrer des chercheurs ayant une connaissance sur d'autres entreprises du marché et sur des clients potentiels.

Au-delà de ces justifications, il convient également de revenir sur les risques éventuels et les avantages de ce mode de passation. Une première limite importante de ce procédé d'enquête est liée à la nécessité d'être passé par les dirigeants de ces entreprises pour négocier l'accès au terrain. Si nos résultats quantitatifs et qualitatifs ne semblent pas avoir été trop affectés par ce facteur du fait de la multiplication des entre-

tiens et des points d'accès aux travailleurs, nous pouvons faire l'hypothèse d'une sous-déclaration par les travailleurs de leurs problèmes. La présence de notre traducteur, Andry, a d'ailleurs été particulièrement importante pour mieux comprendre le contexte des travailleurs au regard du secteur. D'autres éléments sont cette fois liés aux rapports avec les dirigeants des entreprises de sous-traitance. Le premier problème concerne l'appropriation du travail que nous leur avons rendu, ce que nous avons anticipé en précisant dès le départ que ce rapport ne devait être utilisé qu'en interne. Il y avait également un risque important d'être empêché de dévoiler certains éléments liés aux clients et aux travailleurs, malgré l'anonymisation des données.

Le contrat passé avec *Automatik* illustre bien les risques liés au dévoilement d'informations. Dans ce cas, nous avons été confrontés à un dilemme éthique. Nous avons en effet constaté d'importantes irrégularités dans le fonctionnement de leur sous-traitant à Madagascar : le contrat de travail des travailleurs était en fait un contrat d'indépendant déguisé en CDI. Face à cela, nous avons préféré indiquer ce problème au donneur d'ordres même si nous prenions le risque que l'entreprise se sépare de son sous-traitant. Pour minimiser ce risque, nous avons conseillé l'entreprise de manière informelle afin de nous assurer que les travailleurs ne gardent leur travail et surtout que le sous-traitant leur propose un contrat formel. Nous avons également vivement conseillé à l'entreprise française de négocier une augmentation de salaire. Après plusieurs mois de négociation, entre Julie et la dirigeante de l'entreprise à Madagascar, l'entreprise française a pris en charge cette augmentation de salaire et le passage à des contrats CDI.

Ce dernier cas montre également l'avantage de ce mode de passation. Ainsi les temps d'échanges servant à coordonner la passation de l'enquête et à rendre nos résultats ont aussi permis de mieux comprendre l'organisation du travail entre les entreprises donneuses d'ordres et les entreprises sous-traitantes. Notre "présence" s'est prolongée jusqu'à pouvoir suivre une négociation entre *Automatik* et son sous-traitant pour l'amélioration des conditions de travail. Enfin, ce temps de négociation et notre "présence" auprès de l'entreprise d'IA a permis d'accéder à des employés d'entreprises d'externalisation et de réaliser des entretiens avec ces derniers. Dans certains cas, cela a même facilité l'accès aux clients finaux des entreprises d'IA. Nous avons par exemple interrogé un gestionnaire de supermarché qui utilise une solution de vidéo-surveillance automatique, ainsi qu'un fonctionnaire en charge de la mise en place d'une solution similaire dans une ville de la banlieue parisienne. En multipliant les points d'accès aux différents niveaux de la chaîne, nous avons donc reconstitué de nombreuses chaînes de valeur.

3.2.2 Sur le terrain. Quelles différences entre le distanciel et le présentiel ?

Le caractère international de l'enquête réalisé avec des annotateurs de Madagascar, du Cameroun et du Nigéria rend difficile une "co-présence" physique facilitant le recrutement de travailleurs, et ce, surtout dans les conditions d'enquête permises par le Covid-19. L'ensemble de l'enquête s'est donc déroulé de manière

mixte. Ainsi, les deux tiers des entretiens réalisés avec travailleurs de Madagascar se sont déroulés sur place, la plupart du temps dans les locaux de leur entreprise. Concernant les travailleurs d'autres pays étudiés, ces entretiens se sont uniquement déroulés à distance. Le questionnaire a quant à lui été très majoritairement passé à distance, sauf quelques dizaines de réponses réalisées depuis les locaux d'une entreprise à Madagascar.

Pour nous, les deux modes d'enquêtes présentent des avantages et des inconvénients. Ainsi, les entretiens réalisés en à distance ont été réalisés sans caméra la plupart du temps, du fait de difficultés de connexion dans les pays concernés. Il a donc parfois été plus difficile d'avoir une proximité avec la personne enquêtée susceptible de faciliter la discussion. En revanche, cela a tout de même permis de se rendre compte de difficultés propres à Madagascar (par exemple : connexion internet ou matériel informatique pas nécessairement récent). Nous pensons enfin que cette distance "physique" a tout de même pu encourager les travailleurs à répondre avec plus de détachement vis-à-vis de notre présence de chercheur que lors des entretiens en présentiel. En effet, il a été assez fréquent que nous soyons assimilés à des clients ou à de futurs entrepreneurs de la sous-traitance, du fait de notre position sur place (des hommes blancs).

L'enquête à distance s'est déroulée sans traducteur entre le français et le malgache. Par conséquent, les entreprises ont sélectionné les travailleurs les plus doués en langue française. Ce biais s'est résorbé par les entretiens en présentiel, puisque nous avons alors fait appel à un traducteur, ce qui a évidemment d'autres conséquences sur la restitution des paroles de nos enquêtés. Sur ce point, précisons également que la relation avec notre traducteur, par ailleurs juriste d'entreprises, nous a permis de beaucoup mieux comprendre le contexte de Madagascar et le secteur de l'externalisation de services.

La présence physique au sein des locaux a également permis d'observer le processus de production de l'annotation en train de se faire et de mieux comprendre le déroulement d'une journée. Il a également été utile de se rendre dans les bureaux, afin de constater les éléments liés aux infrastructures locales (difficulté de circuler, absence de transports en commun) ou bien de caractéristiques géographiques (la nuit tombe tôt) qui déterminent les conditions de travail. Enfin, nous avons pu constater les caractéristiques des locaux et les différences entre les entreprises sur ce point.

Au cours de cette enquête, nous avons multiplié les accès au terrain en maximisant l'effet boule de neige lié au déroulement de l'étude. Ainsi, le fait de travailler sur les start-ups d'IA a permis d'obtenir des contacts et de faciliter notre accès à des entreprises de sous-traitance. Inversement, certaines entreprises d'IA ont accepté d'être mises en contact avec nous, par le biais des entreprises de sous-traitance. S'il est clair que le fait de passer par les dirigeants d'entreprises de sous-traitance, voire les clients a pu occasionner des biais, le fait de multiplier les accès et les entretiens a été un moyen de mitiger au maximum ce risque. Nous sommes également revenus sur les différences entre le distanciel et le présentiel, soulignant les apports de chacun

des modes de passation. Dans le contexte d'une enquête portant sur une population urbaine travaillant dans le domaine du numérique, nous considérons que cette nécessité a plutôt permis d'offrir un point de vue complémentaire par rapport au terrain en présentiel. Néanmoins, l'enquête portant sur un pays très éloigné de la France en termes de niveau de vie et de normes de travail, l'enquête en présentiel a permis d'apporter énormément d'éléments de contexte propre à améliorer la compréhension des résultats obtenus.

3.3 Exposé des techniques d'enquête et d'analyse

Pour terminer ce chapitre, nous détaillons les différentes techniques d'enquête et d'analyse utilisées dans notre thèse. Nous présentons cette fois l'enquête par méthode mixte réalisée sur les travailleurs des données à l'aide d'un questionnaire et d'une grille d'entretien. Nous détaillons d'abord la construction de grille d'entretien utilisée pour étudier les travailleurs des données et ses évolutions. Nous présentons également notre méthode de codage en vue de l'analyse des entretiens. Enfin, dans une seconde section, nous abordons le processus de construction de notre questionnaire, ses évolutions et les modalités de passation ⁴⁷.

3.3.1 Méthodologies d'analyses qualitatives

La construction de notre grille d'entretien s'est d'abord inspirée de celle utilisée lors de l'enquête Digital Platform Labor portant sur les micro-travailleurs français (A. Casilli et al., 2019a). L'objet de cette étude étant des travailleurs réalisant des tâches courtes sur une plateforme de travail, nous avons rapidement été obligés de transformer substantiellement cette dernière, eu égard aux éléments expliqués précédemment : présence d'entreprises formelles de sous-traitance et travailleurs situés à Madagascar. Ainsi, nous avons retiré tous les éléments propres au travail des plateformes, par exemple, la participation à des forums d'entraide dans la réalisation des tâches ou bien les éléments liés à l'accès à la plateforme (modalités d'inscriptions...) La grille a ensuite continué à évoluer à mesure que nous nous sommes rendu compte des différences entre les deux terrains. À titre d'exemple, il nous a fallu retirer tous les blocs portant sur l'organisation du travail des micro-travailleurs français, qui pour la plupart avait un travail principal, au contraire des travailleurs étudiés à Madagascar dont le travail des données est l'activité principale.

À l'inverse, nous avons approfondi certains éléments qui ne ressortaient pas de l'enquête DIPLab. Ayant un accès à des travailleurs employés sur plusieurs années et uniquement sur des projets d'IA, il a été possible de poser davantage de questions sur les projets d'IA, la manière dont ils étaient pilotés par les entreprises d'IA et de sous-traitance et leurs évolutions dans le temps. Par ailleurs, nous avons également certains éléments propres au contexte de Madagascar et notamment le rapport à l'apprentissage des langues, constatant

47. Nous avons collaboré activement avec Maxime Cornet, doctorant à Télécom Paris pour la réalisation de cette enquête, comme en témoigne les publications co-écrites ensemble. Par ailleurs, Arnaud Leliève a réalisé le nettoyage des données quantitatives

l'importance de cette variable dans la mise en place de la chaîne de sous-traitance. Enfin, la grille a été adaptée à chaque afin de tenir compte des réponses au questionnaire et de s'en servir pour appuyer certains éléments, notamment liés aux conditions de travail. Par exemple, lorsque certains travailleurs ont indiqué se sentir opprimés ou mal payés au travail.

Les entretiens ont tous fait l'objet d'une prise de note très exhaustive lors de leur passation, ce qui a été facilité par la réalisation d'une partie d'entre eux à distance. Nous avons ensuite fait une sélection des meilleurs entretiens en vue d'une retranscription réalisée par une prestataire. Nous avons ensuite eu recours au logiciel NVivo pour réaliser le codage des entretiens. Pour ce faire nous nous sommes basés sur le livre de Flick (2014) sur l'analyse de données qualitatives, en ayant recours à des approches : la "grounded theory" et la "qualitative content analysis". La "grounded theory" est portée par ce que contiennent les données. L'objectif étant de développer des théories en faisant des allers-retours entre le terrain et la théorie. Flick (2014) identifie plusieurs phases :

1. le codage initial consiste à décrire les données en déterminant les catégories qui s'appliquent ;
2. la deuxième étape est le "focused coding" et sert à identifier à partir de la précédente phase les grandes catégories conceptuelles.
3. La dernière phase est la phase de théorisation ("theoretical coding") qui consiste à importer des catégories conceptuelles d'ailleurs.

La "qualitative content analysis" correspond à une description systématique des données à travers le codage. Les étapes sont plus nombreuses et passent surtout par un travail plus important de définition des questions de recherche. Sur cette base, il faut ensuite construire un guide de codage basé sur des catégories soient "data-driven" ou "theory-driven". Puis, il convient de réaliser un premier codage pour tester et évaluer le guide de codage avant de stabiliser les catégories et d'analyser le résultat. Notre processus de codage a mixé les deux approches en empruntant certaines phases à l'une ou à l'autre. Ainsi, nous avons commencé par une relecture de certains entretiens (représentatif des différents types d'acteurs). Puis, nous avons réalisé un premier guide de codage, basé sur : plusieurs questions de recherche, de grandes catégories d'analyses et une description des catégories. Nous avons ensuite testé ce guide en essayant de décrire les données, sans nécessairement le faire à l'aide de catégories conceptuelles. Cette phase a permis de stabiliser nos catégories d'analyses (Niveau 1 du codage) et de trier les catégories moins importantes (niveau 2 et 3). Puis, nous avons retravaillé le codage pour intégrer ces dernières à de grandes catégories conceptuelles. Le résultat final est présenté Figure 51. La difficulté principale de ce travail a été de traiter le nombre important d'entretiens, ce qui nous a poussé à parfois privilégier certains entretiens au détriment d'autres. Si l'ensemble des entretiens a fait l'objet d'un codage, ce n'est pas nécessairement le cas pour l'ensemble des catégories d'analyses. Par ailleurs nous avons à la fois traités des entretiens intégralement retranscrits et d'autres partiellement retranscrits.

3.3.2 Méthodologies d'analyses quantitatives

Dans cette deuxième partie méthodologique, nous revenons sur les différentes étapes de conception du questionnaire. Ce dernier a été construit sur la base de trois questionnaires concernant les micro-travailleurs français (A. Casilli et al., 2019a), les micro-travailleurs indiens (Gray & Suri, 2019) et du monde entier (Berg et al., 2018). Par ailleurs, il a fortement évolué en tenant compte de deux entretiens exploratoires réalisés avec des annotateurs de Madagascar et deux tests du questionnaire. Nous détaillons également les conditions de réalisation de ces tests et le déroulé de sa passation.

Ce questionnaire a été initialement construit sur la base de trois questionnaires utilisés dans le cadre trois enquêtes utilisant des méthodes mixtes, le premier concerne les micro-travailleurs français (A. Casilli et al., 2019a), le second les micro-travailleurs indiens (Gray & Suri, 2019) et le troisième vise les micro-travailleurs du monde entier de plusieurs plateformes (Berg et al., 2018). Ces enquêtes étant centrées sur le micro-travail, elles étaient conçues pour étudier des travailleurs indépendants opérant sur des plateformes numériques de travail (PNT).

Aussi, un premier travail a consisté à supprimer certaines questions propres aux PNT, ou à les adapter si besoin. Par exemple, il était inutile de demander précisément les jours ou les heures travaillés par les travailleurs malgaches, alors que sur cette plateforme, cette donnée permet de comprendre la temporalité de la réalisation des tâches. En revanche, il était intéressant de poser la question du travail de nuit et du week-end pour mieux comprendre l'organisation du travail dans les entreprises d'annotation. Un deuxième travail a consisté à adapter le questionnaire au contexte malgache, que ce soit pour les modalités de réponse ou bien pour les questions. Pour ce faire, nous avons demandé à l'Institut National de la Statistique de Madagascar (INSTAT), l'accès au questionnaire de la dernière enquête emploi. Nous avons reçu une version datant de 2008 et entièrement en malagasy. Aussi, il a été décidé de le traduire automatiquement en français à l'aide de l'API de traduction de Google. Ce faisant nous avons pu reprendre les modalités propres à Madagascar (par exemple : pour les secteurs d'activités ou le niveau d'emploi). Nous avons ensuite ajouté des modules spécifiques : sur la maîtrise de la langue française, car cela correspondait à des hypothèses sur les raisons de l'installation des activités d'annotation à Madagascar et sur le travail d'annotation réalisé via des plateformes numériques de travail. Enfin, un module "bonheur au travail" repris de l'enquête menée en France par Christian (Baudelot & Gollac, 2002) a été ajouté, afin de collecter des données sur les conditions de travail des annotateurs. Nous avons également repris le générateur de position déjà prévu dans le questionnaire de Antonio (A. Casilli et al., 2019a) qui a été adapté au contexte de Madagascar (Table 41).

La première version du questionnaire a été codée sur LimeSurvey, un logiciel de passation de questionnaire qui a été choisi pour trois raisons : nous disposons déjà d'une licence dans notre équipe de recherche, il est *open source*, dispose de serveurs dans l'Union européenne et avait déjà été utilisé lors d'une précédente

enquête. Une fois ceci fait, nous avons organisé un premier test avec un travailleur de l'entreprise A.

Pour ce test, notre contact au sein de la première entreprise étudiée a choisi un travailleur ayant un diplôme en sciences sociales afin que nous puissions avoir le meilleur retour possible sur notre questionnaire. Ce test a été réalisé à distance à deux et enregistré. Une personne posant les questions et une autre se focalisant sur les réactions de la personne interrogée. Ce premier test a mis en lumière plusieurs points : le manque de littératie administrative et informatique, un niveau de français qui pouvait être moyen chez certaines personnes interrogées, et globalement la longueur et la difficulté du questionnaire. Cependant, la suite de l'enquête montrera que ce premier test a été réalisé avec un travailleur peu prompt à répondre à notre questionnaire. Pour faciliter la passation du questionnaire, il a donc été décidé de le simplifier et de le traduire en partie. Sur le premier point, nous avons retravaillé la formulation des questions (par exemple : "à votre domicile" devient "chez vous"). Ensuite, nous avons supprimé certaines questions jugées inutiles, ou qui se répétaient avec la grille d'entretien. Enfin, nous avons demandé une traduction des modules les plus compliqués du questionnaire. Le deuxième test du questionnaire a duré 37 minutes et a été réalisé avec une autre annotatrice. Mise à part la question sur la profession des parents, le questionnaire a beaucoup mieux fonctionné. Il a donc fallu procéder à peu de modifications pour aboutir à une version finale. Nous avons par exemple supprimé la question ouverte : "décrivez votre travail". Le questionnaire sera finalement entièrement traduit par un prestataire et adapté aux différents modes de passation de l'enquête. Nous reproduisons ce questionnaire final dans son intégralité (Voir : l'annexe B.5)

Notre questionnaire comporte finalement 96 questions réparties en 16 blocs de questions. Un premier bloc concerne des informations socio-démographiques basiques et l'origine sociale des répondants (Q1 à Q20). Un deuxième bloc sert à collecter des informations sur la composition du foyer Q21 à Q23. Un troisième bloc est centré sur les compétences des annotateurs : éducation initiale, formation professionnelle et niveau en langue française (Q24 à Q30). Le quatrième bloc sert mieux comprendre l'accès des annotateurs à l'informatique et à Internet (Q31 à Q34). Puis, un cinquième bloc est focalisé sur le travail réalisé par le biais d'une plateforme numérique de travail (Q36 à Q41). La question 42 est un module servant à retracer le parcours professionnel des annotateurs de données. De la question 45 à 68 sont des questions servant à mieux connaître les conditions d'emploi et de travail avec notamment des questions s'inspirant du module bonheur au travail. Enfin, les deux derniers blocs sont le générateur sur lequel nous revenons plus en détail dans la partie suivante et le générateur de noms.

En ce qui concerne le capital social, notre questionnaire comprend un "générateur de positions", c'est-à-dire une série de questions visant à établir l'accès relationnel des répondants à une série de professions telles qu'avocat, ingénieur, enseignant ou chauffeur (Lin & Erickson, 2008). Le capital social du répondant est l'agrégation de tous les contacts avec ces professionnels, en tant qu'indicateur de l'accès potentiel aux connaissances, informations et autres ressources liées à leurs professions (Li & Verhaeghe, 2015; Van

Der Gaag & Webber, 2008). Parce que les professions sont inégalement prestigieuses dans la stratification sociale, cet instrument a également été utilisé pour déduire la structure de classe de la société (Savage et al., 2013). Nous posons une question qui est une adaptation du générateur de position de Van der Gaag, Snijders 2004 et qui vise à mesurer le capital social des individus : *« Connaissez-vous quelqu'un dans chacune de ces professions ? Il s'agit de personnes dont vous connaissez le nom et prénom, et avec qui vous pourriez parler ou échanger (en disant un peu plus que bonjour ou bonsoir!) si vous les rencontriez par hasard dans la rue. »*

Comme l'indiquent Lin et Erickson 2008 : *« la considération la plus importante est le choix des professions. »*. Dans la littérature, trois éléments sont censés garantir la robustesse de la mesure : la simplicité du générateur de position, la représentativité de toute l'échelle sociale dans le générateur de position et représentative du contexte de passation de l'enquête. Initialement, nous nous basés sur la version de l'enquête DIPLab décrite dans l'article de Paola Tubaro et al. (2022)⁴⁸. Concernant le nombre de profession, notre test a montré qu'il fallait qu'il soit moins important afin qu'il soit plus facile à passer. Par ailleurs, nous avons choisi d'ajouter des professions propres au secteur informatique que nous étudions dans cette thèse. Dès lors, comment garder un nombre relativement restreint de professions, tout en conservant une représentativité des positions sociales représentés ? Comme l'indiquent Verhaeghe, Van de Putte et Roose 2013 : *« Plus il y a de professions, plus les ressources du réseau social sont évaluées. »*. En outre, il est nécessaire d'avoir des intitulés de professions simples et clairs pour que les répondants puissent les comprendre. Notre générateur de position doit être le plus proche possible et le plus de la structure des emplois à Madagascar. Molina et al. 2020 ont par exemple ajouté quatre professions basé sur le contexte local dans leur générateur de position pour l'Espagne et au Mexique. Enfin, il est nécessaire que cette liste soit « représentative » des niveaux d'emploi et de classes sociales : *« Les professions doivent être choisies parmi l'ensemble des niveaux et des classes professionnelles et doivent provenir de différents secteurs économiques afin de refléter l'ensemble de la distribution des ressources dans une société. »*. Par ailleurs, Hällsten, Edling et Rygdren 2015 montrent que limiter le nombre d'occupations peut conduire à avoir une mesure plus qualitative⁴⁹. En conséquence, nous avons réduit leur nombre (48 à 22) tout en gardant quatre professions informatiques, des professions pivots (10) et des professions "context-based" basées sur la structure de l'emploi à Madagascar, par exemple opérateur de données ou "manager dans la pêche" (Table 41).

Il était également important de valider la modalité de passation de ce module : entièrement en ligne. van der

48. Il s'agissait d'une "liste de 48 professions, basée sur une classification professionnelle standard (Professions et Catégories Socio-professionnelles, PCS, à 2 chiffres, 2003), certains éléments étant désagrégés en descriptions plus concrètes (...) Huit professions liées à l'informatique et aux technologies numériques [sont incluses] : développeur de logiciels, data scientist, community manager/influencer, web designer, digital business analyst, digital consultant, ingénieur informatique et technicien informatique."

49. « the composition of occupations in this list and/or in the measure of social capital is crucial to the effectiveness of the social capital measure. We show that some are pivotal. Omitting them would severely limit the statistical power of the social capital measure, however, there are also some occupations that only contribute measurement error. Removing such occupations would positively impact the effectiveness of the social capital measure in predicting outcome variables. To some extent the logic is captured in the phrase less is more—by removing a few occupations, we can get a better measure. »

gaag, Appelhof, et Webber (2012) indiquent que le taux de réponses est meilleur en face à face que par voie postale ou par Internet. Toutefois, Marlow (2005) offre un exemple d'un générateur de position qui apparaît après un long questionnaire qui a été passé en ligne et pour lequel il y a eu un faible taux de valeurs manquantes. Par ailleurs, pour des raisons de facilité de codage du questionnaire, nous avons opté pour la "dichotomisation" des réponses ("Oui/non"). Graaf, Kalmijn, Kraaykamp, et Monden (2010) ont également utilisé une méthode de dichotomisation (Oui/Non) pour les questions et montrent que cette façon de faire fonctionne parfaitement.

Le générateur de position s'appuie sur l'attribution de scores à chaque profession de façon à mesurer la stratification sociale d'un échantillon. Le *Social Interaction and Stratification Scales* (CAMSIS) propose des scores et des variations de ces scores en fonction des pays (Meraviglia, 2016). Malheureusement, étant donné qu'il n'existe pas de version nationale pour Madagascar contrairement à la France ou même à d'autres pays comme les Philippines, nous avons décidé d'utiliser la version internationale dont les professions sont basées sur l'International Standard Classification of Occupations (ISCO) (voir : annexe B.4), et ce, en essayant d'adapter les professions.

Pour conclure cette (longue) partie méthodologique, il est important d'aborder la manière dont a été passé le questionnaire. Notons d'emblée que, mises à part pour la passation via Facebook, nous avons systématiquement fait passer le questionnaire avant de mener nos entretiens afin de disposer des données quantitatives au moment des entretiens et d'adapter certaines questions de la grille". Par conséquent, nous avons sélectionné les individus susceptibles de faire l'entretien, sur la base des réponses aux questionnaires (voir : annexe B.3). Ensuite, la majorité des questionnaires ont été passés en ligne, sur les ordinateurs de bureau de travailleurs et donc sur leur temps de travail. Seule une trentaine de réponses ont été réalisées en présence d'une aide, dans des locaux dont nous disposons pour l'occasion à Madagascar. Pour faciliter la compréhension de l'enquête, nous avons rédigé un brief qui a été envoyé aux annotateurs et à leurs managers de manière à ce qu'il soit mis au courant des modalités de l'enquête, et surtout de leur possibilité de refuser la participation à cette dernière. Nous reproduisons ce brief type en annexe dont l'objectif était de les informer des enjeux, du déroulement de l'enquête et de leurs droits au regard du RGPD (voir : annexe B.2). En moyenne, les annotateurs ont passé entre 30 et 40 minutes à répondre à ce questionnaire. Par ailleurs, la Table 2 donne un certain nombre d'informations concernant le taux de réponse qui est de : 70% chez l'entreprise A, 40% au sein du sous-traitant B et quasiment 100% pour le sous-traitant G. Pour Facebook, cette mesure n'a pas vraiment de sens puisque nous avons procédé différemment des trois entreprises avec lesquelles nous avons des contrats. Dans ces entreprises, nous avons été mis en relation avec un chef de projet qui nous a aidés à diffuser le questionnaire, soit par mail soit, dans un cas, sur un outil d'information interne (similaire à Slack). Dans le cas de Facebook, une annonce a été diffusée sur des groupes Facebook qui servent pour le recrutement de travailleurs des données (voir : annexe B.4)

Partie 2 - Comment s'organise le processus de production de l'IA ? Un agencement marchand qui repose sur la qualité des données, une organisation du travail qui s'appuie sur des *hiérarchies électroniques*

L'analyse de la littérature sur le travail des données montre que cette dernière s'est largement focalisée sur le travail intermédié par des plateformes numériques de travail (PNT). On a pu constater les mécanismes d'invisibilisation et de minoration du caractère essentiel de la contribution des travailleurs des données. [Irani \(2015a\)](#) analyse l'utilité du micro-travail pour des entrepreneurs du numérique qui peuvent ainsi accéder facilement à des travailleurs humains interchangeables, ce qui leur permet, *in fine*, de se recentrer sur leurs activités à "haute valeur ajoutée" et de valoriser leurs activités d'innovateurs auprès des fonds d'investissement ⁵⁰.

La minimisation du travail humain est également caractéristique de la recherche en intelligence artificielle. Cette vision de l'organisation du travail de la donnée se révèle ainsi dans des publications en sciences informatiques qui se proposent d'optimiser le fonctionnement de la foule, comparant les micro-travailleurs à des micro-processeurs. Ainsi, Chen Cao et ses collègues (2016) considèrent ainsi que : *"À mesure que l'utilisation du crowdsourcing augmente, il est important de penser à l'optimisation des performances. À cette fin, il est possible de considérer chaque travailleur comme une HPU (Human Processing Unit) et de s'inspirer de l'optimisation des performances des ordinateurs traditionnels ou des nœuds en nuage dotés d'unités centrales."* Les travailleurs sont alors assimilés à des processeurs humains dont l'utilisation doit être maximisée, comme des ordinateurs, comme une puissance de calcul "humaine" (*Human Processing Unit*). Cette perception du travail des données renvoie à l'idée qu'il s'agit d'un travail que *"tout le monde peut faire"*, pour reprendre les mots d'un employé de startup interrogé. Par conséquent, il s'agirait d'un travail qui peut être organisé de la manière la plus flexible possible, hors des cadres d'emplois traditionnels et notamment à travers des plateformes numériques de travail comme Amazon Mechanical Turk.

50. *"En externalisant les tâches fastidieuses, (...) les entrepreneurs concentrent leur propre travail sur ce qui a une plus grande valeur (...). AMT devient alors une infrastructure non seulement pour le traitement des données, mais aussi pour produire la différence entre les "innovateurs" et les travailleurs "subalternes" (...) AMT n'est donc pas seulement un moyen de distribuer le travail cognitif ou immatériel, bien qu'elle le soit aussi. Mais c'est aussi un système qui organise les travailleurs pour le plaisir des programmeurs, en adaptant les travailleurs aux formes de production expérimentale et d'innovation de l'industrie tardive."*

Une récente controverse sur l'utilisation de la plateforme Amazon Mechanical Turk en psychologie invite pourtant à quelques commentaires sur ce mode de collecte des données pour l'IA. En effet, des estimations indiquent que la qualité des données serait en baisse, impactant la qualité des études (Chmielewski & Kucker, 2020; Webb & Tangney, 2022), et ce alors que pas moins de "45% des articles publiés dans les meilleures revues de psychologie comportent au moins une étude réalisée sur MTurk". En parallèle, une analyse de données bibliométriques d'articles scientifiques utilisant Mturk collectées durant notre thèse montre deux éléments. D'abord, son utilisation en sciences informatiques (*computer science*) est en baisse, suggérant d'autres modes de collecte de données dans la recherche en intelligence artificielle. Par ailleurs, la plateforme américaine est surtout utilisée par des chercheurs des USA, ce qui suggère qu'il existe d'autres modes de production des données dans d'autres pays.

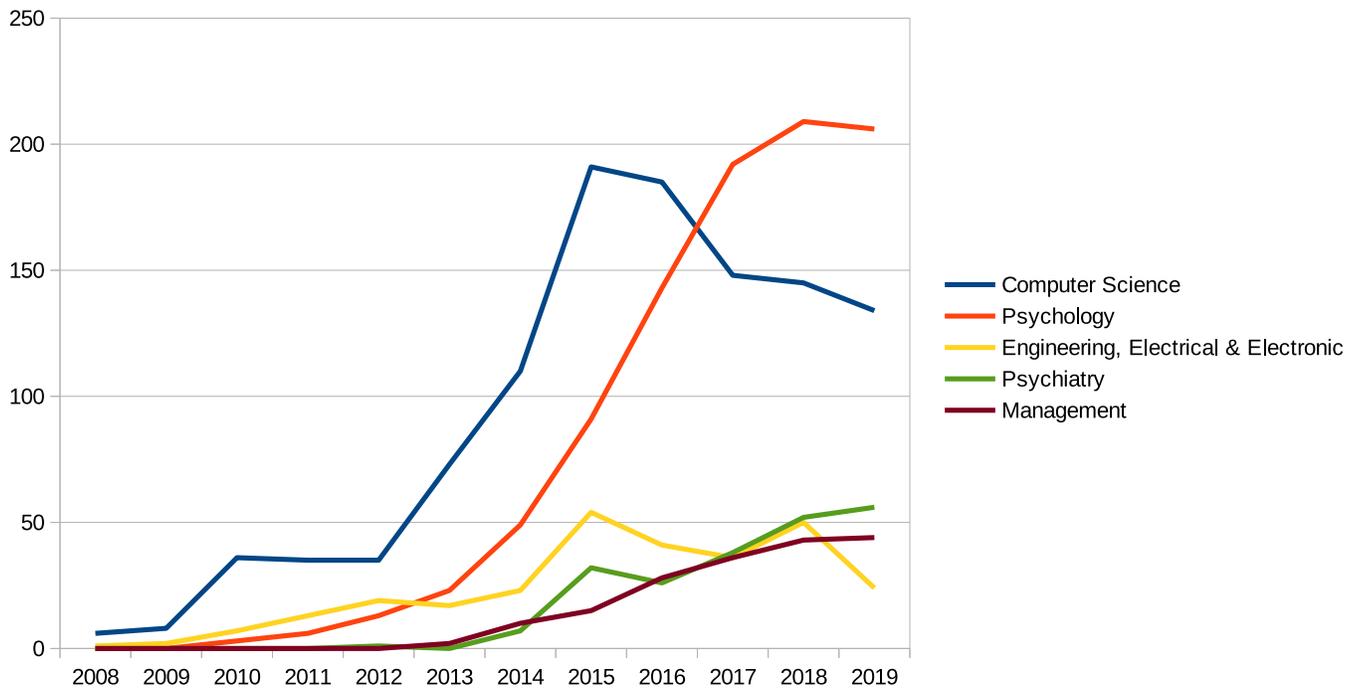


FIGURE 9 – Nombre de documents par année pour le corpus d'articles scientifiques "Mturk". Lecture : ce corpus correspond à une collecte d'articles utilisant la plateforme Amazon Mechanical Turk quelque soit la discipline scientifique

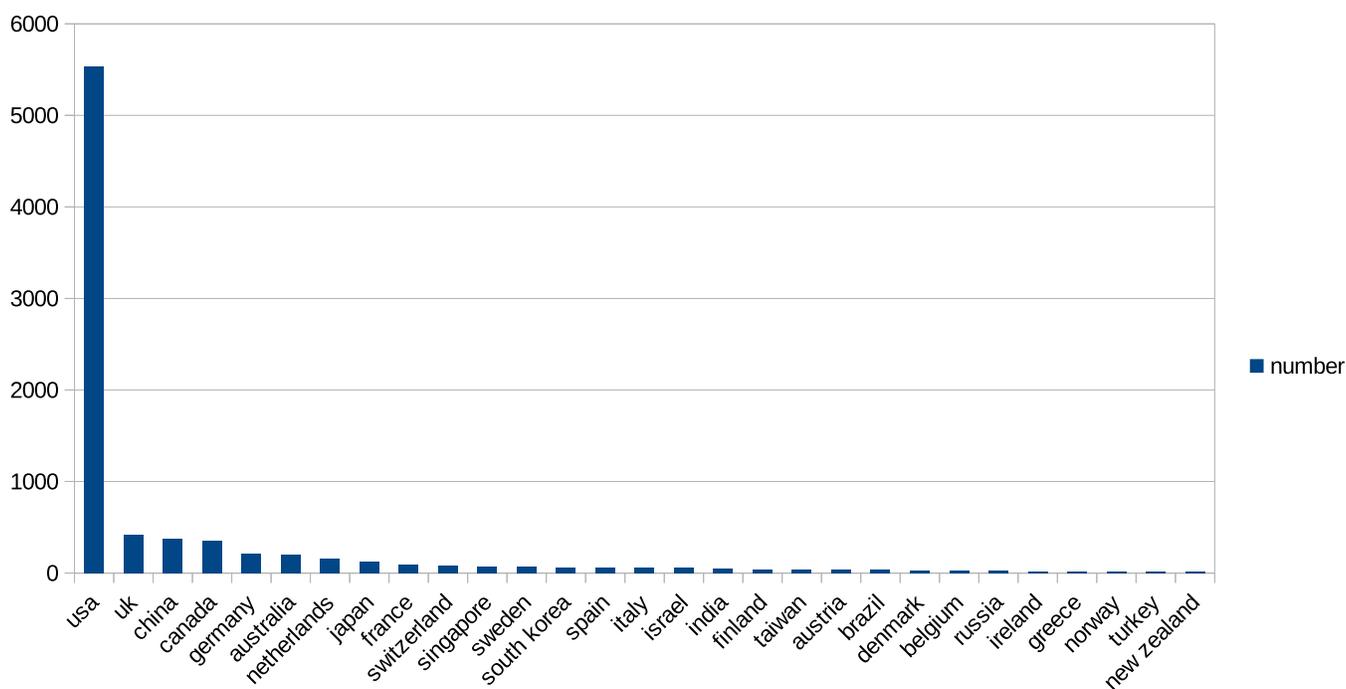


FIGURE 10 – Nombre de documents par discipline scientifique

Le lien entre les biais de l'IA et la qualité des données créées par les travailleurs des données a tout de même fait l'objet de plusieurs analyses critiques en sciences informatiques. Ainsi, Emily Denton et ses collègues (2020b) proposent un programme de recherche pour ramener le travail humain dans les discussions sur l'intelligence artificielle en documentant tout le travail nécessaire pour créer des jeux de données pour l'apprentissage automatique. Amandalynne Paullada et al. (2020) vont plus loin, en soulevant des inquiétudes sur la façon dont les scientifiques ou les entreprises collectent les données. En particulier, elles insistent sur le fait que le jugement des annotateurs de données peut entraîner des biais, qui sont ensuite intégrés au sein de vastes systèmes socio-techniques, posant alors des problèmes pour les utilisateurs de ces derniers. L'enjeu qui sous-tend la question des biais de l'IA est en effet lié à celui de la qualité des données collectées et du degré de maîtrise du processus de production de la part des entreprises qui fabriquent l'IA (Miceli, Schuessler, & Yang, 2020b).

D'autres articles en sciences sociales, suggèrent d'ailleurs l'existence d'autres formes d'organisation du travail des données impliquant une externalisation plus proche de la sous-traitance classique à travers le recours à des entreprises de *business process outsourcing* (BPO) (Miceli et al., 2020b; F. A. Schmidt, 2019). Ainsi, Lehdonvirta et al. (2019) soulignent que les PNT sont fondamentalement différentes des BPO. Les premières permettent aux entreprises clientes de distribuer des projets divisés en tâches plus petites à un grand nombre de prestataires indépendants, tandis que les seconds permettent aux clients d'externaliser

des projets entiers. Entre ces deux extrêmes, existeraient également des organisations hybrides : elles se présentent comme des PNT, mais intègrent des niveaux de gestion et de supervision similaires à ceux des BPO et les entreprises conventionnelles. Selon Tubaro (2021), ces organisations hybrides fournissent du "travail profond" (*deep labour*) associant finalement la fragmentation de la gestion de la force de travail et la stabilité contractuelle. Poursuivant cette perspective de recherche, nous pensons qu'il existe un continuum allant des PNT "pures" aux BPO, mais qu'au sein de ce continuum les organisations du travail se rapprochent d'une forme hiérarchique et donc de formes plus traditionnelles de salariat.

L'ensemble de ces éléments nous amène à plusieurs questions. Tout d'abord, outre les plateformes à la Amazon Mechanical Turk, quels sont les modes d'organisation du travail des données? Est-ce que les entreprises se comportent de la même manière que les chercheurs en matière de recours à des plateformes? Observe-t-on les mêmes tendances dans l'espace francophone que celles constatées dans la sphère anglophone? Finalement, on peut se demander dans quelles mesures ce processus de production peut être organisé "n'importe comment"? Autrement dit, la forme que prend l'organisation du travail des données est-elle déterminée par des facteurs autres que par le coût du travail? Et, si oui, quels sont-ils? Dans ce chapitre, *notre hypothèse principale est que le recours aux plateformes est limité par les enjeux liés à la qualité des données, plus déterminants dans le cas des entreprises. En corollaire de cette hypothèse : les entreprises d'intelligence artificielle passeraient par des travailleurs plus intégrés à leur processus de production ce qui leur permet de mieux contrôler la qualité du processus et les compétences des travailleurs.*

Chapitre 4 - L'intégration logistique des travailleurs des données : la qualité de l'annotation comme limite du recours au "marché pur"

Dans le cas de l'annotation de données réalisée pour les entreprises francophones, la question du modèle organisationnelle reste ouverte tant le point de vue des clients des plateformes d'annotation a été ignoré par la littérature à l'exception de quelques travaux (Nguyen, 2014; G. B. Schmidt & Jettinghoff, 2016; Van Belle et al., 2019), principalement focalisée sur le recours à des plateformes numériques de travail. Dans ce chapitre, nous retraçons le processus de production de l'IA en mobilisant deux approches théoriques complémentaires.

En s'interrogeant sur le mode d'organisation du travail des données, notre travail renvoie directement à deux questions classiques en économie : pourquoi la firme existe-t-elle alors que le marché est supposé offrir une allocation optimale des ressources? Comment ce marché s'agence-t-il en termes de valorisation du travail? Cette question prend un tout autre sens lorsque l'on s'intéresse au recours à des plateformes numériques de travail qui se caractérisent par la mise en relation de différentes "faces" d'un marché. Dès lors, pourquoi certaines entreprises mettent-elles en place des systèmes hiérarchiques plutôt qu'un recours au marché "pur".

Une première dimension théorique de ces deux questions nous invite à prendre en compte le facteur explicatif des coûts de fonctionnement du marché. Dans certains cas, ce dernier est moins efficient qu'une organisation fondée sur l'autorité de l'entrepreneur et sur la relation de subordination qui unit ce dernier et ses employés. Plusieurs théories complémentaires coexistent pour définir la nature de l'entreprise et en étudier les mécanismes de coordination, que ce soit à l'intérieur ou à l'extérieur du périmètre des firmes (Baudry, 2003) (p. 4). Une première catégorie de théories sont dites contractualistes, car elles postulent que le marché est l'organisation la plus efficace. Ce sont les défaillances de ce dernier qui expliquent l'apparition des firmes. La notion de "contrat", au sens juridique, est au centre de l'analyse. La firme y est conçue comme une "structure de gouvernance", façonnée par la rationalité limitée et l'opportunisme des agents, ainsi que par la spécificité des actifs nécessaires à la fabrication du produit final de l'entreprise. Il s'agit par exemple d'un sous-traitant ou d'un salarié qui ne réalise pas la prestation demandée par la firme en y mettant suffisamment d'effort. Sur le second point, Williamson distingue cinq types d'actifs spécifiques : physiques (des machines spécifiques pour fabriquer certaines pièces), localisés (la proximité géographique), humains (l'accumulation de compétences d'un salarié), dédiés (un système d'information de contrôle entre une firme et son sous-traitant) et incorporels (la reconnaissance de la marque de la firme focale). Dès lors, il propose trois types de structures de gouvernance : le marché, la hiérarchie (la firme) et la forme hybride qui existe notamment dans le cas de relations de long terme entre une firme et son fournisseur.

La théorie des incitations s'est quant à elle intéressée à l'organisation qui assure la maximisation de l'effort des individus en postulant toujours que le marché est systématiquement le système le plus efficient. Toutefois, Alchian et Demsetz (1972) observent le développement de formes de travail en équipe, au sein desquelles ce qui est produit collectivement est plus important que la somme de ce qui est produit individuellement, justifiant à nouveau l'existence des firmes. Pour maximiser l'effort individuel, il sera alors nécessaire de mettre en place un système de surveillance et de mesure des contributions individuelles.

Jensen et Meckling étendent l'approche contractuelle à l'ensemble des relations contractuelles en incluant donc tous les acteurs qui interagissent avec la firme (salariés, clients et fournisseurs), les uns et les autres étant reliés par des contrats. Dans cette approche, la firme est vue comme une construction artificielle abritant de nombreux contrats de différentes natures, et qui, en tant que personne morale, est responsable devant les tribunaux des problèmes liés à ces contrats. La firme permet dans ce cadre une centralisation et une simplification des relations contractuelles, en évitant la multiplication des contrats entre les individus, et en contraignant les parties prenantes à l'exécution de ces contrats.

Un dernier versant de la théorie des incitations est proposé par Bengt Holmström, Paul Milgrom et John Roberts. Ils opposent deux systèmes d'incitation : par le travail salarié (subordination, méthode de production définie par un autre agent et salaire fixe) et par le travail indépendant (choix de la méthode de production, propriété des outils de production et rémunération à la production) ⁵¹. Le système incitatif favorisera donc le salariat dans les cas où la mesure de la performance est coûteuse et qu'il est difficile d'évaluer les activités des salariés. En revanche, le travail indépendant sera plus efficace s'il est aisé de mesurer la performance.

La théorie des contrats incomplets va enfin intégrer les bénéfices et les coûts de l'intégration des sous-traitants, en insistant sur les droits de propriété. La source de l'incomplétude des contrats réside dans la difficulté de vérifier *ex post* certaines variables déterminantes de la relation, comme le niveau d'investissement d'un co-contractant dans de nouvelles machines plus performantes, ou même l'investissement de ses salariés. Au-delà de la difficulté de vérifier empiriquement ces modèles, deux critiques limitent la portée de ces théories contractualistes. Il nous semble réducteur de penser que le rôle de la firme est uniquement disciplinaire. Ensuite, ce schéma explicatif conduit à *"à attribuer la source de la performance d'une firme non pas à la façon dont on crée de la valeur, c'est-à-dire à la façon dont une firme parvient à être plus performante par rapport à une autre dans sa fonction productive, mais à la manière dont la firme met en place les mécanismes d'incitation appropriés"* (Baudry (2003) p.27)

51. *"le système incitatif combine trois dispositifs, l'autorité, la propriété et le système de rémunération. Il s'agit alors de rechercher la meilleure combinaison possible de ces trois dispositifs, considérés comme endogènes, pour que le travailleur alloue ses efforts de manière optimale, sachant qu'il existe trois facteurs exogènes susceptibles d'influencer cette combinaison : l'incertitude sur le futur, le degré de spécificité des actifs sur lesquels portent les échanges et les coûts de mesure de la performance (output) des travailleurs.* (Baudry (2003) p.21)

Une dernière manière d'expliquer l'existence des firmes repose sur l'analyse des mécanismes de "hiérarchie" et de "coopération" qui s'y exercent. Ici, Baudry (2003) propose de distinguer la hiérarchie et l'autorité, le premier désignant par exemple le fait qu'un sous-traitant se réfère aux instructions de la firme donneuse d'ordre, notamment en cas de conflits sur la réalisation de la tâche. L'autorité se définit quant à elle par la délégation du pouvoir de décision d'un agent à un autre⁵².

La hiérarchie implique donc une subordination, tandis que l'autorité s'établit entre deux agents par "pur consentement" dans le cadre de la relation contractuelle. Le fondement de la hiérarchie réside dans la propriété des moyens de production : "Autrement dit, l'employé est dépendant économiquement de son employeur". Le droit du travail traduit cette inégalité. L'employeur dispose d'un pouvoir de direction économique qui lui permet de supprimer ou de transformer son activité. Il dispose également d'un pouvoir de direction des personnes qui lui permet de décider de recruter un salarié et de déterminer ses tâches. Enfin, il bénéficie d'un pouvoir disciplinaire lui permettant de prononcer des sanctions à l'encontre des salariés. Toutefois, cette relation de subordination est limitée par les conditions réelles d'exercice de la tâche. Ainsi, il n'est pas possible pour le dirigeant d'entreprise de donner des ordres à chaque instant et qui tiennent compte des éventuelles évolutions de cette tâche. Il s'agit donc plutôt d'un ensemble de règles qui encadrent l'action du salarié.

En réponse à ces limites, des mécanismes de contrôle sont mis en place, par exemple, à travers des indicateurs chiffrés sur l'activité des salariés (nombre de tâches réalisées). Toutefois, ce contrôle est nécessairement imparfait : il n'est pas toujours faisable pour certaines tâches, est coûteux et ne supprime pas totalement l'initiative individuelle. Il peut donc être complété par des mécanismes d'incitations monétaires ou non monétaires : salaire fixe à l'*input* (comportement du salarié) soit un salaire à l'*output* (en fonction de la performance). Dans le contexte d'une division internationale du travail, l'analyse des relations inter-firmes a également fait l'objet de développement en théorie de la firme autour de l'émergence des "firme-réseaux". Ce concept acte l'existence d'entreprises intermédiaires entre la forme pure du "marché" et celle de "l'entreprise" qui s'appuient sur de nouvelles structures de gouvernance afin de coordonner des réseaux de sous-traitants, notamment dans des secteurs comme l'automobile.

La caractéristique principale de la firme-réseau est de coordonner d'autres agents économiques "juridiquement indépendants" (par exemple : un réseau de sous-traitant). Ces acteurs concourent à mettre en commun différentes compétences pour la réalisation d'un produit ou d'un service. En théorie, la firme-réseau fait donc appel au marché à travers le recours à d'autres entreprises sous-traitantes. Toutefois, la complexité des produits finis et la nécessité de coordonner différentes sphères de compétences impliquent de mettre en place des mécanismes d'incitation qui diffèrent à la fois de ceux du marché (*prix*) et de ceux de l'entre-

52. C'est par exemple le cas, lorsqu'un restaurant consent à déléguer le pouvoir de décision concernant les livreurs à la plateforme Deliveroo.

prise (le *contrôle* et la *hiérarchie* qui s'appliquent aux salariés). Cette intégration logistique passe par des mécanismes de communication entre la firme focale et les sous-traitants, mais aussi par des dispositifs de remontée de données sur le travail réalisé par le sous-traitant d'un côté et sur les besoins de la firme focale de l'autre⁵³.

Globalement, la littérature s'accorde sur l'existence de caractéristiques spécifiques aux firme-réseaux, à travers les nouveaux dispositifs d'incitation qu'elles développent pour contrôler leurs sous-traitants : l'intégration *organisationnelle* avec la collaboration plus poussée des salariés d'un sous-traitant au sein de la firme focale, des dispositifs d'*intégration logistique* qui se traduisent par des systèmes de remontée de l'information (par exemple : sur l'état des stocks...) et par la coordination du *contrôle de la qualité* à travers les normes ISO et les cahiers des charges à remplir de la part des fournisseurs. Cela pose à nouveau la question des frontières de la firme et de la pertinence de la distinction entre le marché et la firme.

Sur ce point, la théorie des firmes propose plusieurs hypothèses concernant les firme-réseaux que nous reprenons ici, la première thèse étant celle du continuum. Ici, la firme-réseau est "hybride" empruntant à la fois au marché et à la hiérarchie : "(...) *l'économie n'apparaît plus ainsi comme une articulation de marchés et de hiérarchies, mais plutôt comme un vaste réseau de contrats (...) la firme-réseau se situe en définitive le long du continuum, la distinction avec la firme au sens strict étant une affaire de degré et non de nature.*" (Baudry, 2003) (p. 107). Par conséquent, la firme peut maintenir les incitations du marché, tout en évitant les "*distorsions bureaucratiques*" liées au modèle hiérarchique. Cette hybridité est toutefois limitée par la spécificité des actifs qui fera pencher les firmes vers un modèle purement hiérarchique.

Baudry (2003) souligne plusieurs limites à cette vision de la firme-réseau (p.106). Ainsi, selon lui, sans tomber dans le modèle "hiérarchique", les firmes arrivent à faire diminuer les coûts de coordination à travers des dispositifs d'incitations spécifiques comme l'intégration logistique. Par ailleurs, il considère que cette forme de gouvernance demeure floue, tant il est impossible de réellement distinguer la forme marché de la forme hybride : les attributs sont les mêmes, seul leur degré diffère. En effet, dans les deux cas, la relation d'emploi est remplacée par une relation commerciale de manière similaire à la forme du marché.

La théorie des "*chaînes de valeur mondiales*" (CVM) s'inscrit quant à elle dans une tradition d'investigation

53. "Nous pensons que la firme-réseau organise, entre des firmes juridiquement indépendantes (...) Les firmes en position de fournisseurs ne sont en fait qu'une articulation dans un ensemble économique plus large. Ceci signifie que ces firmes, tout en étant juridiquement indépendantes, concourent à un même processus de fabrication (...) Nous voulons par là souligner le fait que, dans ce type de relation, seul le produit livré à la clientèle finale par la firme-pivot est un produit collectif et est donc, substantiellement, une marchandise relevant de la sphère de l'échange proprement dite (...) Comme les activités des firmes membres du réseau relèvent de la sphère de la production, l'impératif de coordination est dans ces conditions essentiel. Or, si le marché n'est pas une structure de gouvernance adéquate pour gérer la relation, les dispositifs d'incitation et de coordination ne peuvent pas être les mêmes que ceux qui prévalent à l'intérieur de la firme, puisque la relation met face à face des firmes juridiquement indépendantes. C'est pourquoi, au sein de la firme-réseau, il existe des mécanismes incitatifs et des dispositifs de coordination, comme l'intégration logistique, totalement spécifiques." (Baudry, 2003)

à l'intersection des études commerciales et de la sociologie économique. Elle vise à comprendre l'externalisation de la production de biens et de services, dans le contexte de la mondialisation, trait distinctif de ces théories. Hopkins et Wallerstein (1977) considèrent ainsi la mondialisation comme "un réseau de processus de travail et de production dont le résultat final est un produit fini". Le cadre d'analyse des chaînes globales de commodité (CGC) explore les facteurs à l'oeuvre dans la constitution de ces chaînes en décrivant les relations inter-firmes en termes de structure de gouvernance (Gereffi, Korzeniewicz, & Political Economy of the World-System Conference, 1994). Cette littérature montre comment certaines entreprises contrôlent d'autres firmes dans le cadre de relations de sous-traitance.

Certaines recherches montrent toutefois que le pilotage de certaines chaînes de valeurs est effectué par d'autres acteurs que les firmes, par exemple les traders qui négocient l'achat et la vente de la production (Gibbon, 2001). La gouvernance peut être également bicéphale, associant deux acteurs à sa tête (Fold, 2002). Un dernier apport vient du constat d'agencement particulier de chaînes de valeur : les réseaux modulaires (Sturgeon, 2002). La théorie des réseaux modulaires s'est développée en réponse aux recherches en sociologie économique et notamment autour de l'existence de formes "hybrides" entre marché et hiérarchie : des arrangements contractuels de long terme et/ou une dépendance mutuelle entre deux entreprises. Dans ce cas, c'est le réseau, c'est-à-dire les liens sociaux entre les entrepreneurs, qui permet de réguler les échanges entre les firmes. En ce sens, le réseau inter-firme peut constituer une structure de gouvernance, une forme sociale d'organisation qui se distingue donc des marchés et des firmes hiérarchiques. La théorie des réseaux modulaires tente de répondre à la question suivante : quelle structure alternative aux réseaux permettrait d'obtenir des bénéfices équivalents en termes de "transmission et l'acquisition de connaissances et compétences nouvelles" ? Sturgeon (2002), à travers les relations interfirmes dans l'électronique, montre que le développement de standards qui s'appliquent à toute l'industrie permet de faciliter l'échange d'informations sur les transactions entre les firmes. Ainsi, "les standards imitent la confiance". Dès lors, une firme donneuse d'ordre pourra connaître les caractéristiques précises de la réalisation d'un service sans forcément développer des relations de confiance avec ses fournisseurs. Par ailleurs, dans le cadre des réseaux modulaires, un produit est divisé en un ensemble de tâches qui seront réalisées indifféremment par plusieurs fournisseurs. Finalement, les réseaux modulaires se caractérisent par l'indépendance de chaque unité de production, mais aussi par une intégration logistique permise par la standardisation et l'échange d'information.

Prolongeant ces analyses, Gereffi, Humphrey, et Sturgeon (2005) identifient une typologie des relations interfirmes entre marché et hiérarchies : les gouvernances *par le marché*, *modulaire*, *relationnelle*, *captive* et *hiérarchique*. La gouvernance par le *marché* se caractérise par l'échange de produits facilement codifiables et par des échanges d'informations relativement simples entre fournisseurs et acheteurs, qui permettent de faibles coûts de coordination. Les chaînes de valeur *modulaires* sont caractérisées par la fourniture de produits plus complexes qui sont fabriqués selon les instructions des acheteurs. Dans ce cas, même si le

produit est complexe, la codification des processus de production permet de réduire le *”besoin de suivi et de contrôle direct de l’acheteur”*. Le troisième type de chaîne de valeur est *relationnel*, car les produits sont si complexes que les spécifications ne peuvent être codifiées explicitement. Ces chaînes de valeur se caractérisent par une forte dépendance entre acheteurs et fournisseurs et par l’échange de *”connaissances tacites”*, ce qui rend nécessaire des interactions fréquentes. Les *chaînes de valeur captives* font, quant à elles, référence à des produits complexes pour lesquels la codification est difficile et où les capacités des fournisseurs sont faibles. Dans ce cas, les acheteurs ont besoin de contrôler leurs fournisseurs et de les rendre dépendants. Enfin, les *chaînes de valeur hiérarchiques* désignent des produits complexes pour lesquels il est difficile de trouver des fournisseurs compétents. Par conséquent, il est nécessaire pour les acheteurs de développer ces produits en interne. Ici, la théorie des GVC semble plutôt s’inscrire dans la thèse du continuum évoqué dans la section précédente : il existerait une pluralité de structures de gouvernance entre le *”marché”* et la *”hiérarchie”*. Toutefois, le fait de souligner l’existence de certaines structures hybrides comme la forme modulaire et relationnel semble également accréditer la thèse de la spécificité des firme-réseaux.

Le travail de [N. M. Coe, Dicken, et Hess \(2008\)](#) vise ainsi à palier certains manques de la littérature existante sur les GVC. Le traitement des firmes fondé sur la dichotomie entre les firmes pilotes et les firmes sous-traitantes est peu satisfaisant, car il ne tient pas compte de la variété des rapports de pouvoirs. De plus, les chaînes de valeur sont des entités dynamiques, ce qui a pour conséquence de faire évoluer le rapport de pouvoir entre les acteurs. Les relations intra-firmes sont également très peu considérées alors que la structure interne d’une firme et les relations qui s’y jouent sont essentielles dans la structure de la production. Cette dimension renvoie à la firme comme un réseau interne imbriqué dans des réseaux plus vastes (par exemple : institutionnels) qui s’oppose à une vision légaliste de la firme dans laquelle les frontières de la firme sont clairement délimitées par le droit. Le concept de GPN part plutôt du principe que : *”les frontières entre l’internalisation et l’externalisation des fonctions sont en constantes évolutions, impliquant une reconfiguration complexe des frontières organisationnelles, y compris la prolifération de diverses formes de collaboration interentreprises.”* ([N. M. Coe, Johns, & Ward, 2007](#)).

Les startup françaises que nous avons étudiées externalisent effectivement cette activité pour certaines d’entre-elles, mais pas vers une foule d’utilisateurs anonymes sur une plateforme en ligne. C’est également le cas pour certaines grandes entreprises américaines citées dans les projets réalisés par les annotateurs. Comment alors expliquer cette tendance à contre-courant du recours au marché *”pur”*? Nous verrons que cela correspond à une nécessité de maîtrise de la qualité des données qui passe par une relation directe avec les fournisseurs permettant aux entreprises d’IA de contrôler les travailleurs presque comme s’ils étaient leurs propres employés. Ainsi, cela implique de mettre en place des processus variés de contrôle, caractéristiques de l’intégration logistique : outils numériques de communication et contrôle managérial.

4.1 Décider du mode d'externalisation du travail des données. Une force de travail stable pour garantir la qualité des données et maîtriser la volatilité des modèles d'IA

Notre thèse a déjà expliqué les différentes perspectives ouvertes par les recherches autour du *digital labor* en matière de lien entre IA et travail des données. Ainsi, nous avons souligné les différentes manières d'accéder aux données traitées dans la littérature : travail de la foule (platformisé), micro-travail (impact sourcing) et travail humain de calcul (travail gratuit). Dans cette première sous-section, nous montrons que de nombreuses startup d'IA adoptent une organisation type comparable aux firme-réseaux coordonnant le travail de sous-traitants pour la production de données. Ce faisant, nous nous inscrivons dans la lignée de Tubaro (2021) en identifiant une évolution du marché des données vers une plus grande intégration des sous-traitants. Paradoxalement, chez de nombreux enquêtés, la contribution des travailleurs des données est pourtant minimisée en même temps qu'ils en soulignent l'absolue nécessité pour atteindre les promesses de l'intelligence artificielle. Comment se décide alors l'externalisation de la phase de production de données ? Comment les entreprises trouvent-elles des prestataires pour s'occuper de cette activité ?

L'enjeu de cette sous-partie est d'interroger les facteurs qui motivent les entreprises françaises d'IA à faire le choix de la sous-traitance directe. Ce faisant, nous questionnons également le mode d'organisation du travail de l'IA et la structure de cette industrie. Nous verrons en effet qu'un autre enjeu essentiel est celui de la nécessité d'établir des standards de qualité des données annotées et d'en contrôler la conformité. La question de la qualité a ainsi été largement soulignée dans nos entretiens comme déterminante pour produire un modèle performant. Cette dimension de notre travail illustre l'existence d'une "économie de la qualité" pour reprendre ce concept issu de la sociologie économique.

Cette perspective théorique met en évidence l'importance cruciale de la qualité dans la production, la consommation et la valorisation des biens et des services. Se faisant, elle se distingue des approches traditionnelles de l'économie en ce sens qu'elle ne se limite pas à considérer la quantité produite, mais accorde une attention particulière à la valeur intrinsèque des produits. La qualité n'est pas simplement une caractéristique accessoire, mais plutôt un élément central dans la détermination de la valeur économique. Ainsi, des dimensions tangibles, telles que la durabilité, la fiabilité et la performance, mais également des aspects plus subjectifs, telles que la réputation de la marque, la satisfaction du consommateur et l'expérience d'achat constituent des facteurs déterminants dans l'agencement du marché (Kjellberg et Helgesson, 2007). L'économie de la qualité explore également les mécanismes sociaux qui sous-tendent la création et la perception de la qualité (Granovetter, 1985 ; Zelizer, 1994). Elle met en lumière le rôle des normes sociales, des réseaux de confiance et des interactions entre les acteurs économiques dans la construction de la qualité et s'intéresse aux stratégies des entreprises visant à maximiser la qualité de leurs produits ou services, tout en mainte-

nant une rentabilité économique. Cette perspective vient également s'articuler avec des questionnements qui portent sur les modèles organisationnels permettant ce niveau de qualité.

Les recherches en sciences sociales et en sciences informatiques mentionnées dans notre premièrement illustrent très largement les problèmes liés aux biais contenus dans les données utilisées pour l'entraînement des modèles d'IA et plus généralement les risques des biais de l'IA pour leurs utilisateurs. Au regard du champ théorique de l'économie de la qualité, le processus de développement de l'IA pose donc un certain nombre de questions. Quelles méthodes de collecte, de nettoyage et de validation des données sont mises en place pour garantir la qualité des données utilisées dans le développement de modèles d'IA afin d'en réduire les biais? Comment les entreprises d'IA intègrent-elles les enjeux de validation de la qualité des données et de détection des biais tout au long du cycle de vie des projets d'IA? Quelles pratiques sont développées pour garantir la traçabilité et la transparence du processus de production de l'IA? Comment les entreprises mettent en œuvre des mécanismes de responsabilité et d'audit pour surveiller et améliorer en permanence la qualité des modèles d'IA en production et atténuer les risques associés aux biais et aux erreurs? Pour répondre à ces questions, nous avançons comme *hypothèse principale dans cette partie que le facteur de la qualité des données est justement celui qui détermine une organisation du travail plus professionnelle ce qui conduit in fine à une plus grande intégration des travailleurs des données, et ce, par rapport au modèle idéal-typique du travail de plateforme.*

Dans cette partie, nous abordons d'abord le processus de décision de l'externalisation du travail de la donnée chez les entreprises d'intelligence artificielle. Nous verrons que ce choix répond d'abord à un enjeu de qualité des données. Puis, nous verrons, à travers le positionnement des entreprises d'IA vis-à-vis du monde académique que le choix de prestataire de type BPO correspond à un objectif d'optimisation des modèles et d'industrialisation des processus de production au regard de la réalité du terrain le tout en minimisant les biais potentiels. Nous verrons qu'il s'agit d'une activité fortement socialisée et organisée justifiée par la nécessité de maîtriser les procédures d'annotation de données.

4.1.1 Formaliser l'organisation du travail d'annotation dans les entreprises d'IA françaises : la qualité des données par la stabilité de la force de travail

L'objectif de cette première sous-section est de comprendre comment les entreprises d'IA mettent en place l'externalisation du travail des données, c'est à dire la manière dont ils décident ou non d'externaliser cette partie de la production de l'IA, et les raisons qui expliquent ces choix. Nous y analysons leurs pratiques concrètes d'annotation, y compris lorsque cette phase de la production est complètement internalisée. Nous verrons d'ailleurs que certaines entreprises passent d'un travail d'annotation internalisé à l'externalisation de cette activité.

Dans certains cas, nous avons constaté que l'annotation a été complètement internalisée par les entreprises d'IA. Un premier exemple concerne CallAI, une startup de moins de 10 employés qui développe des algorithmes de reconnaissance des échanges téléphoniques destinés à des centres d'appels. En analysant les conversations des clients, ce produit d'IA doit faciliter le travail des opérateurs téléphoniques en proposant automatiquement les meilleures réponses possible aux clients. Lors d'un entretien avec le CEO de la startup, il a été étonnant de constater qu'il réalise lui-même l'annotation, que ce soit pour créer des données de conversation, c'est-à-dire en enregistrant des phrases, ou pour transcrire de manière textuelle des phrases prononcées lors des échanges téléphoniques⁵⁴. Cette entreprise a rapidement constaté les limites des jeux de données existants, qui ne correspondaient pas aux cas d'usages qu'elle rencontrait avec ses clients. Elle a donc dû créer son propre outil d'annotation⁵⁵. C'est notamment l'adaptation au contexte du client final et la reconnaissance de produits spécifiques qui rend nécessaire la production de données propre à l'entreprise. Le dirigeant cite par exemple le cas des appels téléphoniques qui contiennent des noms de produits.

Encadré n°1 : Personnaliser son modèle en créant ses propres outils d'annotations

Q : Et la détection des mots-clés, vous l'adaptez par client ou ça va détecter les mots-clés pour tous les clients pour enrichir la database quoiqu'il arrive ?

R : On l'adapte par client. Donc quand le client vient nous voir, il nous dit Moi, j'aimerais bien repérer quand on dit tel nom de produit et que ça affiche tel type d'information . Donc on définit que ce produit c'est un mot-clé en lui-même dans notre outil. L'idée est vraiment de faire du personnalisé. Et en plus des annotations d'une qualité remarquable, sur des données difficiles! Aujourd'hui, si tu vas sur Google Speech-to-Text et que tu cliques sur démo, tu peux tester, tu peux dire des noms de produits spécifiques (je prends souvent l'exemple Xperia ou notre produit Live Insights), et en fait, il va galérer parce qu'on est en langue française et il ne comprend pas que je switch pour citer un nom de produit en anglais, donc il va me sortir n'importe quoi, et Xperia je pense qu'il ne le reconnaît même pas du tout. Donc c'était pour pallier ce genre de problématique où en fait, il y avait des outils qui faisaient de la transcription audio, mais qui n'étaient pas suffisants pour pouvoir être sûrs du contexte, donc là c'est vraiment du contexte spécifique où on va n'écouter rien d'autre que le mot-clé. On ne va même pas chercher à traduire le texte. On va juste dire Ah, il y a un des mots-clés qui est sorti , et on va afficher du contenu par rapport à ce mot-clé.

CEO - CallAI

Dans l'entretien, le CEO précise également qu'il aimerait externaliser une partie de ce travail, mais que

54. "Du coup, pour l'instant c'est souvent moi qui le fais et je suis bien aidé par l'équipe. À la volée, je dirais qu'on a dû faire 200 000 annotations depuis deux ans." (CEO - CallAI)

55. "on a été obligés de créer plus ou moins notre outil d'annotation/de labélisation (c'est de la donnée audio). Il y avait un certain nombre de datasets qui étaient de base disponibles, mais ça ne correspondait pas forcément à ce qu'on voulait faire." (CEO - CallAI)

cette possibilité dépendra de leur capacité à lever des fonds auprès d'investisseurs et à conquérir des parts de marché. Dans ce cas, c'est donc la taille de l'entreprise et ses moyens financiers qui limitent encore l'externalisation de l'annotation. De ce point de vue, le marché de l'annotation est structurellement lié au mode de financement de l'économie numérique.

Une autre entreprise, Clik, internalise également le travail d'annotation. Elle propose plusieurs services d'IA au cas par cas, selon les demandes des clients, avec une spécialisation dans la reconnaissance d'images. Il s'agit d'un cas typique d'*agence d'IA*, dans le sens où elle ne développe pas un produit en particulier. Cette entreprise a par exemple travaillé sur des projets d'automatisation du floutage comme sur des détections d'anomalies sur des bâtiments. Au tout début de l'existence de l'entreprise, il a été décidé d'embaucher d'abord un annotateur, avec lequel nous avons pu discuter. Pendant cette période, le travail d'annotation est partagé entre lui et les chefs des projets d'IA qui sont davantage spécialisés dans la conception des modèles⁵⁶. Avec l'augmentation des besoins en données liée à l'arrivée de nouveaux clients, il a fallu créer une équipe d'annotateurs⁵⁷. Au moment de l'entretien, Jin gère désormais une équipe d'une petite dizaine de personnes dédiées à l'annotation embauchées sur des contrats à temps partiel en télétravail. Au quotidien, il doit récupérer les besoins des chefs de projet IA en matière de données, puis réaliser des guides d'annotation pour former son équipe à un travail d'annotation toujours varié. C'est principalement à cause du traitement de données sensibles, en l'occurrence des visages, que la startup a préféré cette organisation du travail de la donnée à l'externalisation, potentiellement plus respectueuse du RGPD selon le CEO.

Dans la catégorie des entreprises qui internalisent la production de données, nous avons également rencontré le cas de AIVoice, une entreprise qui développe un programme de réponse vocale automatique destiné à équiper les centres d'appels. Cette solution a par exemple été utilisée par certains de ces centres durant la crise sanitaire, afin de gérer les pics d'appels, notamment au moment des annonces du gouvernement sur les nouvelles règles applicables. C'est lors d'un entretien avec F.W, une annotatrice que nous avons mieux compris l'organisation de l'entreprise. Elle s'articule autour d'une équipe de recherche en IA qui comporte des d'annotateurs comme F.W, qui a la particularité d'avoir une formation en traitement automatique du langage (TAL). Ces dernières, toutes des femmes, sont embauchées en CDD ou en stage pour des durées très courtes (3 à 6 mois)⁵⁸.

56. "Q : Et au tout début, quand tu rejoins Clik, il y a que toi qui fait de l'annotation ?

R : Moi et les chefs de projets. Mais moi essentiellement avec parfois aussi les chefs de projets quand il faut aller un peu plus vite et qu'on doit rendre compte de résultats auprès des clients. " (Chef de projet data - Clik)

57. Quand j'ai intégré Clik, je faisais essentiellement de l'annotation. Ils avaient besoin de labéliser pour avancer certains projets. Comme ils obtenaient de bons résultats, ils ont décidé de prolonger mon contrat. Comme on avait pas mal de projets en cours, et de plus en plus besoin d'annoter, ils ont décidé de monter une équipe de labeliseurs, dont je m'occupe (...) Avec les projets en cours, c'est clair qu'un seul labeliseur ne suffisait plus. Donc on a pris un, puis deux, puis trois, puis quatre, pour qu'on puisse avancer les projets et répondre aux besoins des clients dans les temps. (Chef de projet data - Clik)

58. "Au niveau de l'équipe, il y a une trentaine de personnes avec plusieurs profils, les commerciaux, les développeurs back-end, le support (management), il y aussi le pôle recherche et développement avec les ingénieurs en intelligence artificielle qui utilisent

Lors de l'enquête, nous avons également interrogé Eleonor qui dirige le département de recherche et développement de l'entreprise, et surtout le pôle dédié à l'annotation (pôle data) qui fait lui-même partie de ce département. Elle raconte qu'au tout début de l'entreprise c'est une consultante externe, docteure en linguistique, qui met en place le processus de production de l'annotation. C'est elle qui a déterminé la manière dont les données devaient être annotées : en interne⁵⁹. À son arrivée, Eleonor a proposé de garder ce système plus susceptible de garantir la qualité des données. Comme l'explique Patricia, responsable du produit, la difficulté pour mettre en place une sous-traitance réside dans les caractéristiques intrinsèques du produit qui impliquent une très grande maîtrise du français de la part des annotatrices. En effet, ces dernières doivent retranscrire des séquences sonores, mais aussi déterminer les émotions de ces séquences.

Encadré n°2 : La maîtrise de l'annotation comme variable importante de la qualité du modèle

Q : Ma question était de savoir si vous aviez eu vent d'autres entreprises en IA similaires à la votre et de comment eux faisaient leur travail d'annotation ?

R : Je ne sais pas. Pour nous l'annotation est une part essentielle de notre travail dans la construction et l'élaboration des projets d'IA et de callbots, donc on n'imagine pas une seconde, ni la sous-traiter, et encore moins à l'étranger. Après, ça dépend ce qu'on annote. Aujourd'hui, nous on annote de la voix, en français. Les annotatrices doivent maîtriser parfaitement la langue (langue maternelle), sans accent. Quand on annote la partie émotion, l'annotatrice doit être très fine sur la langue et l'émotion traduite par ce mot ou cette phrase. Donc on n'imagine pas une seconde la possibilité de sous-traiter, et encore moins à l'export, ce genre d'annotation. Après, il y a des annotations plus simples et avec moins d'enjeux qui relèvent plus de la masse que de la finesse. L'annotation peut être très diverse. Mais je ne maîtrise pas le marché de l'annotation en tant que marché.

Patricia - AIVoice

En matière d'internalisation du processus d'annotation, nous avons également inclus le cas de LegalAI. Il s'agit d'une entreprise qui propose un outil d'IA servant à automatiser la détection des éléments importants des documents juridiques (contrats...). Pour le développement de leur modèle, ils se basent sur les

les données et les annotateurs c'est assez fluctuant car c'est des missions courtes (CDD 3 mois et les personnes restent trois mois), normalement c'est trois annotateurs, mais là on est 5. Il y a un pic de demande et on est deux stagiaires. Je pense que depuis le confinement, les clients de l'entreprise considèrent que c'est de plus en plus important d'avoir des téléconseillers disponibles, donc c'est vrai qu'il y a une demande de ces solutions d'IA qui sont plus pratiques.”(Floriane - Annotatrice AIVoice)

59. “Au début, ce n'était pas à 100% interne. Il y avait une personne qui travaillait en freelance, qui était docteur en linguistique et qui a sa propre boîte de transcription/annotation. Elle accompagne les entreprises dans cette tâche. C'était elle la conseillère de AIVoice sur cette partie. Il y a eu tout un protocole d'annotation qui a été rédigé par elle, qui s'y connaît bien, qui a fait sa thèse là-dessus. Ensuite, il y a eu l'embauche de ces trois personnes qui ont été managées par cette personne. Une fois par semaine, il y avait une réunion où elle analysait les résultats, le protocole d'annotation, etc.” Eleonor (CTO AIVoice)

larges modèles linguistiques comme Chat-GPT qu'ils adaptent aux demandes spécifiques des clients. Le directeur de la startup cite par exemple les cas d'une banque, d'un cabinet d'avocat ou d'une entreprise de construction dont les besoins diffèrent. L'entreprise a choisi d'avoir recours à la "junior entreprise", une entreprise étudiante, d'une université anglaise. Des étudiants s'occupent de l'annotation depuis les locaux de la startup ⁶⁰(Guillaume - CTO LegalAI). Comme l'explique ce dernier, l'intérêt principal de cette manière de produire les données est de pouvoir compter sur des profils ayant des compétences juridiques. Par ailleurs, ces étudiants étant inscrits dans un master international, la startup a accès à des ressources linguistiques variées, puisque le produit final est censé être multilingue. Un dernier aspect a été largement évoqué par le CTO, celui d'avoir une proximité forte entre les équipes qui développent le modèle et celles qui produisent les données, d'où l'intérêt de faire venir les étudiants dans leurs locaux. Le fait de maîtriser le droit facilite leur compréhension des tâches d'annotation qui leur sont demandées.

Encadré n°3 : Trouver des "ressources humaines" pour produire des annotations spécialisées

Ce sont des étudiants de Queen Mary qui sont en général en Master légal. Queen Mary a une espèce de junior entreprise qui permet aux étudiants d'être rémunérés. On a souvent des étudiants au profil international qui viennent dans ce master (...). En général à la fin de l'année, ils partent dans un cabinet d'avocats fiscalistes ou autre. On essaie de prendre des profils internationaux parce que le modèle est multilingue. Par exemple, on a des requêtes en suédois, on a des gens qui nous demandent si on est fort en polonais. Donc le profil, c'est ça, étudiant en master pour devenir avocat et en général qui parle plusieurs langues.

Guillaume - CTO LegalAI

Si l'on a constaté quelques cas d'entreprises qui internalisent l'annotation, la plupart finissent néanmoins par avoir recours à l'externalisation. C'est le cas pour MLDupli, une startup qui développe un système automatique pour détecter les produits de luxe contrefaits, vendus sur les sites de e-commerce. Il s'agit d'un des premiers cas rencontré lors de notre enquête et pour lequel j'ai réalisé deux entretiens, chacun avec une *data analyst*. Le service vendu par l'entreprise consiste, sur la base d'un catalogue de produits fourni par le client final, à collecter les données des catalogues de e-commerce, puis à réaliser une détection des faux produits. Globalement, le rôle des data analystes consiste à maintenir à jour les scripts permettant d'extraire les données des sites de e-commerce et, le cas échéant, d'en créer de nouveaux, selon les besoins des clients.

60. "Ensuite on a une équipe data qui produit de la data, qui est pilotée par notre CEO et cette équipe-là, on a l'équivalent d'une ressource, même si elle change tous les jours, parce qu'on a des étudiants avocats qui viennent un jour par semaine, et on en a quatre/cinq. Et le reste des équipes, ça va être marketing, vente et customer success. Et tous ces gens sont à Londres."

Toutefois, les deux data analystes indiquent qu'en plus de réaliser ce qu'elles qualifient de tâche plus "intensive", une partie de leur travail consiste à faire du "travail manuel" (*manual work*) afin de fournir mensuellement au client les informations produites par la start-up : des fichiers contenant tous les produits contrefaits et le lien vers leurs annonces. Il s'agit de nettoyer ces tableurs, en enlevant les faux positifs détectés par l'algorithme et de s'assurer que les scripts captent bien l'ensemble des produits visés par l'entreprise cliente, comme l'explique Jeanne, une des deux data analystes : *"pour vérifier la couverture des outils de scraping. Cette tâche n'est pas automatisée, nous avons des scripts pour le faire, auparavant. Vous découvrez des produits sur Amazon, lorsque vous le faites, mais si vous le faites, vous trouverez des produits qui ne sont pas de Dior, et vous trouverez des équivalents, nous devons nettoyer la base de données et nous assurer que nous les obtenons tous."* En d'autres termes, il s'agit par exemple d'aller vérifier si le produit détecté est bien contrefait. Avant le recours à un prestataire externe ce travail était réalisé toutes les fins de mois par l'ensemble des salariés de la startup, afin de tenir les délais de livraison des données, et ce d'autant plus lorsque les clients commençaient à affluer.

C'est le constat de l'ampleur du travail manuel nécessaire à la livraison des données qui va la pousser à demander à son CEO le recours à des prestataires pour réaliser ces tâches à sa place. Ils commencent par embaucher une personne en sous-traitance. Cette expérience ne fonctionne pas très bien, la raison invoquée étant le manque de compétences en informatique de la personne initialement recrutée⁶¹. Aussi, la startup décide de faire appel à un sous-traitant situé en Arménie qui travaille déjà avec eux sur des missions de développement web⁶². Pour la partie "données", cette même entreprise, met alors en place une équipe de 5 personnes en Arménie : *"Pour notre projet, ils ont embauché des étudiants, ce qui revient moins cher. Je pense qu'ils les ont payés à l'heure, car nous avons 5 personnes qui travaillent et 2 à temps partiel."* (Carole - Data analyst - MLDupli) Pour gérer cette équipe, K. utilise la plateforme de gestion de projets Slack qui lui permet d'être en contact régulier avec son équipe et de lui assigner directement les tâches à réaliser. Par ailleurs, un outil a été développé en interne pour que l'équipe en Arménie puisse réaliser le travail : *"Nous avons une plate-forme spéciale pour l'appariement, ils (les développeurs) créent une plate-forme. Ils ont accès aux captures d'écran, au titre (du produit). Ils peuvent faire des recherches."*

L'entreprise RectiTAL propose quant à elle un service de traitement automatique du langage destiné à détecter les éléments importants des documents juridiques (par exemple : pour des banques). Au sein de cette entreprise, nous avons pu questionner Mathilde, en charge de gérer l'annotation pour la startup ainsi que le directeur de l'entreprise de sous-traitance qui s'occupe de l'annotation et les travailleurs des données

61. *"Nous avons un sous-traitant, juste un freelance qui était un ami d'un data scientist, une femme qui travaillait pour nous, à temps partiel. Cela n'a pas très bien fonctionné parce que le salaire est trop bas, qu'elle ne parle pas anglais, qu'elle ne se sent pas intégrée et qu'elle n'a pas assez de compétences en informatique. Elle a environ 40 ans."*(K. Data Analyste - MLDupli)

62. *"Pour le front-end, ils ont deux ingénieurs de WeDoApps, l'un est à plein temps et basé en Ukraine, et l'autre est au Chili."* (Data analyst - MLDupli)

malgaches. Elle a été spécifiquement recruté pour mettre en place le travail de collecte et d'annotation, après une thèse en traitement automatique du langage. Elle constate à son arrivée qu'il manque encore les données d'entraînement, bien que l'architecture et les modèles à tester soient prêts. Ces données vont être collectées directement sur le Web, puis triées par Mathilde ⁶³. Après la collecte, il fallait également produire les annotations de données. Dans ce cas-là, cela consiste par exemple à mettre un cadre autour d'un titre dans un document ou autour d'une date, ou autour d'une marque. C'est Mathilde qui va faire ce travail pendant plusieurs semaines, mais, comme dans le cas de MLDupli, elle va finir par demander à avoir de l'aide dans cette tâche. Comme dans d'autres cas, un aspect est alors déterminant dans la mise en place de l'externalisation du travail des données : le fait que la startup ait réussi à lever des fonds.

Un dernier cas d'externalisation nous semble particulièrement intéressant. Il s'agit de AIMedic, une startup qui opère dans le secteur médical et qui propose un outil de détection automatique des anomalies cardiaques sur la base de vidéos d'électro-cardiogrammes (ECG) ⁶⁴. Étant donné qu'il s'agit d'un produit destiné à la santé, l'annotation nécessite une expertise médicale de la part de l'annotateur, car ce dernier doit être capable de dire, sur la base d'une vidéo, si le cœur est en train de battre normalement ou non.

Pour produire les premières données annotées, l'entreprise s'est d'abord appuyée sur un partenariat avec un hôpital des États-Unis qui a fourni des données déjà annotées : *"on avait le diagnostic de l'algorithme embarqué dans l'appareil, parce qu'on n'est pas les premiers à faire des algorithmes. En fait, chaque fabricant d'appareils fait un algorithme (...) mais en fait les médecins ne l'utilisaient pas trop parce que ce n'était pas assez précis. Ça lançait beaucoup de fausses alertes, de peur d'éviter un problème."* (Nicolas - CEO AIMedic). Concernant l'annotation, AIMedic a d'abord eu recours à l'expertise d'un cardiologue français très réputé, notamment pour ses contenus pédagogiques sur les ECG ⁶⁵. Par la suite, leur collaboration a pris fin brutalement, mais les données qu'il avait produites ont toutefois continué à être utilisées comme base de connaissance pour les annotations réalisées ultérieurement. Cette fois, ce sont les data scientist chargés du

63. Q : *Et pour faire ça, vous utilisez quoi ? Des thèses ?* R : *On a fait du web scraping de plein de documents et on avait une énorme base de données de je ne sais plus combien de gigas, et après je les trie et je les nettoie. Enfin, super chiant ! Et personne ne veut faire ça et c'est extrêmement demandé parce que sans ça, tu ne peux rien faire. Donc on avait trois/quatre personnes pour les modèles. Mais ils ne pouvaient rien faire, parce qu'ils n'avaient pas la donnée. Du coup, ils attendaient la donnée. C'est complètement débile ! (...) C'est je suis thésard, je fais tels modèles neuronaux hyper complexes avec telle architecture, j'ai utilisé telle fonction de BERT, c'est state of the art, mais j'attends pour le faire tourner !"* (Mathilde - CTO - BankAI)

64. Dans ce cas, nous avons eu l'occasion de questionner deux employés de la startup, ainsi qu'un dirigeant de leur prestataire pour l'annotation

65. *"On avait un des co-fondateurs de l'entreprise qui était la personne qui leur avait permis de recevoir le prix de la Banque Publique d'Investissement, et donc ils avaient eu, je crois, 700 000 euros de leur première levée de fonds (à peu près la moitié, et le reste c'étaient des business angels). Ce qu'ils font c'est qu'ils missionnent un expert du domaine et ensuite Yann et Jia avait présenté des slides à cette personne, et c'était Pierre Taboulet, qui est très connu dans la cardiologie en France. Son site est vraiment celui que tous les étudiants en médecine vont consulter. Et il a kiffé le projet et il est très motivé par l'IA. Il se levait tous les jours à 5 heures pour annoter pendant deux heures avant d'aller bosser. Il a annoté une quantité de données vraiment phénoménale ! Je pense que l'on peut parler de quelque chose de l'ordre d'entre 20 000 et 40 000 ECG sur quelques années. Et en plus des annotations d'une qualité remarquable, sur des données difficiles !"* (Cédric - Lead data Team - AIMedic)

développement du modèle qui ont produit les données annotées. L'avantage était alors de pouvoir discuter quotidiennement des annotations difficiles à faire. En revanche, entre la formation et l'annotation effective ce travail prenait énormément de temps aux ingénieurs censés fabriquer les modèles d'IA⁶⁶. Pour éviter à ses ingénieurs de faire de l'annotation, AIMedic a choisi de se tourner vers un de leur premier client afin de continuer faire de la labelisation⁶⁷. Ce client, un centre de radiologie qui pratique les ECG, est composé de techniciens qui ont donc l'habitude de lire ce type d'examen médical. En plus d'utiliser le produit au quotidien, ces derniers devaient donc également produire des annotations.

Encadré n°4 : Trouver des annotateurs "experts"

Une solution assez cool qu'on avait trouvée au début, c'était qu'on avait des utilisateurs de notre produit, qui sont des techniciens ECG, donc globalement la même connaissance que des cardiologues, mais qui n'ont pas le diplôme, qui utilisaient notre produit commercial pour faire des rapports et on s'était dit avec leur chef on fait deux contrats ; un contrat pour notre produit normal, et un autre contrat où on vous achète un service à un prix pas cher, qui permet de garder vos équipes utilisées . Typiquement, ça l'arrangeait parce que il voulait faire croître son équipe (...)il pouvait utiliser notre contrat pour utiliser ses ressources et ne pas perdre d'argent avec des gens qui n'avaient rien à faire.

Cédric - Lead data Team - AIMedic

L'entreprise finira par avoir recours aux services d'une entreprise d'annotation à Madagascar, et ce, même si les annotations produites dans ce contexte n'étaient pas utilisables pour l'outil final. En effet, l'entreprise doit aussi se conformer aux normes médicales en matière d'autorisation des dispositifs médicaux, ce qui implique d'avoir de toute manière recours à des annotations réalisées par des experts occidentaux du domaine⁶⁸.

66. "Q : Tu as parlé de donner des petites formations en cardiologie. Qui en reçoit ? Juste les data-scientists ou n'importe qui qui rentre dans la boîte, y compris les annotateurs qui vont traiter les données, reçoivent une formation cardio ? Réponse : C'est tout le monde dans la boîte. Les data scientists encore plus parce qu'en fait au jour le jour ils sont amenés à annoter, donc souvent on annote ensemble et on a des groupes de consensus ou quand des gens participent on fait des revues ensemble et on discute des cas difficiles, ce qui fait progresser tout le monde. Pour ce qui est des annotateurs externes, c'est peut-être le bon moment pour en parler. Pendant très longtemps, quand je suis arrivé et pendant encore au moins deux/trois ans après que je sois arrivé, c'était fait uniquement en interne." (Cédric - Lead data Team - AIMedic)

67. "Et puis on s'est dit, finalement, les employés, ce sont souvent des data-scientists et on aimerait bien qu'ils fassent autre chose (...) on voudrait un truc qui scale. Ça a été mon sujet pendant longtemps (...) ça a commencé il y a trois ans environ." (Cédric - Lead data Team - AIMedic)

68. "Par contre, il y a aussi des cas où malheureusement on ne peut pas utiliser cette ressource-là, même si on a confiance et que l'on pense que c'est une ressource qui est vraiment fiable. Dans certains cas, par exemple réglementaire, il y a des credentials bien spécifiques qui sont demandés. Par exemple, on a travaillé avec la FDA, on a eu une autorisation déjà et on travaille sur des mises à jour assez régulièrement ; et la FDA demande que ce soient des techniciens certifiés américains et des cardiologues certifiés américains, et ce sont des gens complètement différents et beaucoup plus chers. Donc on a dû travailler avec des profils comme ça" (Cédric - AIMedic)

Selon nos interlocuteurs, l'utilisation des plateformes numériques de travail pour produire des données est fondamentalement différente du travail avec des sous-traitants à long terme. Pour les employés d'Automatik, une autre startup d'IA, des sous-traitants avec qui ils sont directement en relation fournissent des ensembles de données de meilleure qualité que les travailleurs de la plateforme⁶⁹. Dans les projets d'Automatik, les annotateurs doivent ainsi développer une expertise dans des tâches spécifiques. Les managers et les employés de la startup soutiennent et forment les travailleurs à la fois lorsqu'ils effectuent des fonctions essentielles, comme la détection des vols ou la distinction des différents plats sur un plateau et lorsqu'ils manipulent les outils d'annotation. Ceci, à son tour, affecte le flux d'informations et la forme des chaînes de production dans les organisations étudiées. La qualité et la spécificité des annotations de données sont donc les principaux critères pour écarter les plateformes de micro-travail. Dans toutes les entreprises étudiées, ce critère est rapidement ressorti des entretiens et vient contrebalancer l'argument du prix ce qui explique que le modèle BPO soit considéré comme le meilleur moyen d'aboutir à ce niveau de qualité en intégrant davantage les travailleurs dans le processus de production.

Nos résultats montrent que le développement de l'"expertise" des travailleurs et la nature dynamique des guides et des étiquettes d'annotation sont essentiels pour les projets d'IA à long terme. Dans le cas d'une caisse de cantine développée par Automatik, le processus de développement des compétences est formalisé par la création d'"annotateurs experts", spécialisés dans l'identification d'un certain type de plat (entrées, plats principaux, etc.). Par ailleurs, l'activité est organisée de manière à ce que les annotateurs développent une expertise sur les données : *"Q : Pourquoi pas des plateformes ? R : Franchement, c'est la qualité et le prix. En moyenne, on paye un annotateur 500 euros par mois. Mais si je devais comparer à ça à faire mes datasets moi-même... Et la qualité, c'est si je comparais ça à MTurk, etc. Il y a quand même un savoir-faire et une expertise qui monte au fur et à mesure et ils font des missions qui correspondent à notre typologie de clients et à nos produits. On peut avoir des vrais retours produits."* (Théophile - CEO - Automatik)

On retrouve cette même idée de compétence à propos de l'entreprise AIMedic qui a besoin d'annotateurs qualifiés en médecine. C'est finalement la variable *optimisation des ressources humaines*, soit le temps passé par leurs propres employés qui va les inciter à externaliser cette activité, avec la difficulté de trouver des personnes qualifiées dans l'analyse d'ECG. Après avoir eu recours à des cardiologues et des techniciens en radiologie français⁷⁰, l'entreprise a trouvé une solution d'externalisation à Madagascar en travaillant avec

69. *"Honnêtement, nous avons regardé Amazon et son système, mais nous l'avons rapidement abandonné. Du point de vue de l'annotation, la qualité est pourrie. Nous voulions que les gens développent une certaine expertise. Si des personnes choisies au hasard font cela pour une petite somme d'argent, la qualité [du travail] sera très probablement affectée. Si vous foirez 100% de vos annotations, cela ne sert à rien."* (Automatik CTO)

70. *"on a eu pendant longtemps des gens en interne, soit des cardiologues en interne (après Pierre, on a toujours eu des cardiologues. On a un cardio de Massy qui est là une journée par semaine et d'autres que l'on connaît qui travaillent de manière ponctuelle), soit des employés."* (Cédric - Lead data Team - AIMedic)

des médecins malgaches pour l'annotation des ECG ⁷¹.

Derrière l'enjeu de la montée en compétences des annotateurs identifié par le CEO d'Automatik, on retrouve, en creux, le fait que c'est aussi grâce à une proximité forte entre les équipes de développement et les équipes d'annotations que le modèle va être plus performant. Il s'agit alors d'avoir des moyens d'intégrer les travailleurs de manière à ce qu'il puisse y avoir des interactions entre les équipes "data" et les développeurs du modèle. Nous retrouvons cet aspect dans le discours d'Eleonor (AIVoice) qui justifie le recours à des annotatrices en interne par sa trajectoire professionnelle qui l'a amené à utiliser durant sa thèse la plateforme Figure Eight, depuis rachetée par Appen. Elle explique notamment qu'il est très difficile de contrôler le travail d'annotation sur les plateformes de crowdworking : *"Oui, je les ai utilisés pendant ma thèse. J'ai utilisé Figure Eight. Je me dis que si AIVoice a déjà entamé cette activité et que c'est pérenne, c'est déjà bien mis en place et ancré dans l'ADN de l'entreprise, je trouve ça beaucoup mieux qu'avoir recours à ce genre d'outils, tout simplement parce que c'est la qualité. J'ai eu ce problème avec Figure Eight, même si je choisissais la meilleure qualité d'annotation que je pouvais, c'est-à-dire finalement en payant plus les gens qui annotent. Le problème est que je ne peux pas vraiment vérifier la qualité. Le contrôle de la qualité est un vrai problème sur ces plateformes."* (Eleonor - CTO AIVoice) Lorsqu'elle arrive chez AIVoice, elle y découvre un modèle différent de ce qu'elle a pu utiliser auparavant. Au sein de sa startup, les annotateurs ont des compétences en langue et font partie intégrante du département de recherche et développement dont elle a la charge. Cette manière de procéder rejoint une tendance déjà évoquée dans d'autres entreprises et qui consiste à intégrer le plus possible les annotateurs au processus de développement du modèle afin que ces derniers soient plus performants. De ce point de vue, le travail des données fait partie intégrante du travail de construction des modèles.

Encadré n°5 : La proximité avec les annotateurs pour la gestion de la qualité des modèles d'IA

"Par contre, avec des gens avec qui on travaille tous les jours, avec qui on fait des réunions, que les consignes sont claires et que l'on parle tous la même langue, ça se passe beaucoup mieux. Et les résultats le montrent aussi. Nos algorithmes sont très performants et c'est donc que nos données sont bien annotées et bien nettoyées (...) C'est arrivé quelquefois sur du texte où on se rend compte qu'il y a une phrase qui comporte une erreur (...), mais c'est très rare. Quand ça vient d'une équipe en interne qui est experte en langage, c'est beaucoup moins fréquent qu'avec des plateformes comme Figure Eight ou autre."

71. "On a commencé avec une fille, une médecin généraliste de Madagascar qui venait de finir sa formation. Le [sous-traitant B] m'a expliqué un peu comment ça marchait. Apparemment, après la formation en médecine, il y a un peu un temps mort à Madagascar, pendant lequel les gens peuvent faire un stage de formation et leur business marche pour ce truc-là. Ils peuvent utiliser ce temps pour travailler dans ce domaine."(Cédric - Lead data Team - AIMedic)

Dans cette section, nous avons analysé comment se formalisent les pratiques de production de données au sein des entreprises françaises d'IA étudiées. On constate que cette phase est soit internalisée ou bien externalisée à des sous-traitants directs à Madagascar. Les raisons qui justifient ces différents modes de production nous permettent d'identifier plusieurs catégories d'analyse qui expliquent les limites du recours au marché "pur". S'il ne faut pas minimiser l'impact des coûts de production, il est impératif de souligner qu'il ne s'agit pas d'un facteur décisif du choix de ces deux modes de production précis, en comparaison avec le recours à des plateformes comme Mturk. En effet, comme on le verra dans notre dernière partie, nos estimations de la rémunération des travailleurs de Madagascar (de 0,41\$ à 1,7\$ par heure) sont nettement inférieures aux revenus estimés d'Amazon Mechanical Turk à 3,13\$ par heure en moyenne (Hara et al., 2018), tout en sachant que ce coût horaire ne représente pas le prix total de la prestation. Ainsi, nos entretiens permettent de constater que le coût pour les entreprises de sous-traitance va de 500 euros par mois et par agent d'annotation à 1500 euros. Même avec un mode d'organisation susceptible d'améliorer la qualité des données et des modèles, les travailleurs de Madagascar sont moins payés que des travailleurs des plateformes.

Lorsque les entreprises internalisent encore ce processus de production, elles soulignent des enjeux de moyens économiques. Ainsi, elles sont dépendantes de leur capacité à lever des fonds et donc de la structure du financement des entreprises technologiques en France. D'autres éléments viennent limiter le recours à des prestataires externes et notamment la régulation du secteur numérique (par exemple : le RGPD). Le choix de l'externalisation par des prestataires en contact direct est lié à deux facteurs. Tout d'abord, il s'agit d'un choix d'optimisation des ressources humaines internes à ces startups et notamment pour les profils techniques de haut niveau comme les *data scientist*. Ensuite, lorsqu'elles choisissent d'externaliser, la question de la qualité des données est revenue à des nombreuses reprises, comme la raison principale de préférer des travailleurs des données stables, à temps plein sur le projet et avec lesquels il est possible d'échanger plutôt que des travailleurs de plateformes. Cet élément est également un facteur qui peut justifier le recours à des travailleurs directement employés par l'entreprise d'IA.

Nom	Type de données	Mode de production des données	Justification
ModeAI	Photos	BPO Madagascar	Qualité
MLDupli	Photos / Textes	BPO Arménie	RH
CityVeillance	Photos / Vidéos	BPO Madagascar / Impact sourcing	RH
Automatik	Photos	BPO Madagascar	Qualité
Clik	Photos	Interne	Régulation
AIVoice	Son	Interne	Qualité
CallAI	Son	Interne	Moyens
LegalAI	Textes	Interne	Qualité
AIMedic	Images	BPO Madagascar / Interne	Régulation / RH
BankAI	Textes	BPO Madagascar	Qualité

TABLE 3 – Mode de production des données des entreprises d’IA étudiées

Ces éléments n’expliquent pas entièrement ce qui justifie ce recours à des services permettant d’atteindre une qualité suffisante. Derrière cette notion, se joue en fait d’autres éléments plus profonds. Finalement, pourquoi ce qui est en apparence, une activité que tout le monde peut faire depuis n’importe où est externalisée de cette manière-là? Qu’est-ce qu’implique cette relation directe avec les sous-traitants en matière d’organisation de la chaîne de production en comparaison avec le marché ”pur”?

4.1.2 Un impératif d’applicabilité du modèle qui diffère du monde académique : la confrontation avec le réel des modèles déployés en production

En matière d’organisation du travail d’annotation, les personnes interrogées ont aussi mis en avant une distinction entre le monde de l’entreprise et le monde scientifique qui vise plutôt à développer des modèles de grande ampleur comme ChatGPT. Cet aspect rejoint l’hypothèse que nous avons formulée dans l’introduction de ce chapitre autour du fait que les entreprises d’IA fonctionnent différemment des chercheurs en matière de production des données.

Dans l’extrait d’entretien qui suit, l’ingénieur en chef d’Automatik explique que ce qu’il observe du monde académique, c’est une compétition pour la performance du modèle en lui-même, sans forcément prendre en compte la source des données. Dans les entreprises, au contraire, l’enjeu est de proposer le service le plus efficace, basé sur le modèle le plus performant, y compris si cela passe par un travail plus important autour de la source des données et donc autour de l’annotation.

Encadré n°6 : Une forte différence entre le monde académique et le monde de l'entreprise

Il y a un peu de littérature, mais la pure computer science, on va être dans le cliché des chercheurs, mais ils sont en train de se battre pour gagner des pourcentages en améliorant leurs réseaux de neurones, mais pas en améliorant la source de données, parce qu'ils vont gagner des décimales sur un dataset qui est un peu pourri en disant Mon réseau de neurones est meilleur . Mais nous, d'un point de vue entreprise, ingénierie, etc., on n'a rien à prouver à personne et s'il faut réannoter un dataset pour gagner en performance, on va plutôt faire ça.

Corentin - Ingénieur - Automatik

Cet aspect sera confirmé par un autre entretien, cette fois-ci avec le CEO de LegalAI, qui met en avant la différence entre le monde académique et le monde de l'entreprise, car les chercheurs ont besoin de faire des *benchmark*. Il s'agit de jeux de données qui permettent à la communauté scientifique de l'IA de mesurer la performance des modèles sur des tâches spécifiques (par exemple : reconnaître des objets sur une image), et ce, de manière standard. Dans le monde réel, l'important est au contraire de maximiser les exemples et les catégories d'objets reconnus par le modèle. Cette différence rejoint également la critique des benchmarks formulée par [Raji, Bender, Paullada, Denton, et Hanna \(2021\)](#) : *"les benchmarks présentés comme des mesures de progrès vers une capacité générale [à réaliser des] tâches vagues telles que la "compréhension visuelle" ou la "compréhension linguistique" sont aussi inefficaces qu'un musée fini pour représenter "tout ce qui existe dans le monde entier", et ce pour des raisons similaires - étant intrinsèquement spécifiques, finis et contextuels."*

De ce point de vue, l'enjeu des entreprises réside davantage dans un arbitrage autour d'où investir de l'argent : faut-il recruter plutôt des data scientists en France ou plutôt des annotateurs à Madagascar ? Cette différence suggère finalement que la qualité, non pas seulement des données, mais de l'ensemble du processus d'annotation est au moins aussi importante que le modèle. Guillaume considère d'ailleurs que la valeur des annotateurs ne réside pas seulement dans le fait d'annoter, mais aussi dans leur capacité à qualifier et à faire remonter ce qu'il y a dans les données : c'est ce qu'il appelle une approche *"qualitative"* des données. Il explique que les corrections de son modèle proviennent soit des utilisateurs de l'outil (donc des clients) soit des annotateurs du fait de leur connaissance experte de ce que recouvre précisément telle ou telle donnée (*une expertise du terrain*). Finalement, c'est la donnée qui a le plus de valeur. Loin d'être une activité périphérique, la phase d'annotation est au contraire au cœur de la production de l'IA, ce qui nécessite une proximité forte avec les annotateurs.

Encadré n°7 : Investir dans le modèle ou dans les données ?

"(...)le monde académique est un peu différent de l'industrie. Le monde académique a besoin d'un

benchmark et l'important c'est qu'il n'y ait pas de contagion entre le modèle et le data set. Ça, c'est vraiment le plus important. Mais dans le monde réel, la data change et tu as beaucoup plus la main sur ta data. Et où tu mets des ressources? Tu les mets sur améliorer ta data ou améliorer ton modèle? Pour moi, la data a souvent plus de valeur(...) Mais ce n'est pas juste la production. C'est vraiment une vision un peu qualitative de la data et c'est ça qui est hyper dur à avoir, parce que ce n'est pas juste un annotateur, c'est vraiment quelqu'un qui regarde les performances du modèle par rapport à la data et qui commence à se dire mon modèle, je peux identifier que le facteur qui fait que l'on n'est pas bon ici, c'est parce que sur ce type de contrat il n'y a qu'une partie et on en prenait tout le temps deux. Si tu arrives à faire ça, tu as tout gagné, parce que tu vas pouvoir produire plus de data ici. Mais insérer ce truc-là, c'est hyper dur. Dire mon modèle n'est pas bon à cause de ça et à cause de cette data qui manque, c'est hyper dur. Et ça peut venir que d'une communication entre les équipes qui peut produire des métriques sur les résultats du modèle. Les deux corrections majeures de notre modèle, elles sont venues (...)de gens qui l'utilisaient(...)

Guillaume - CEO - LegalAI

Il rejoint Théophile, CEO d'Automatik qui considère *"que les modèles, on s'en fiche, parce que de toutes les manières ils évoluent très vite."* En effet, l'enjeu est de coller à *"la vision (...) terrain"*. Son entreprise développe par exemple un outil de détection des fibres optiques mal posées par les techniciens, et dans ce cas précis : *"ce n'est pas parce qu'aujourd'hui tu sais dire que la fibre est bien posée, que demain ce modèle saura dire que la fibre est bien posée. Les normes bougent, le matériel aussi donc du coup tes algorithmes sont un peu moins forts à reconnaître une armoire de rue des années 2000 que 2020, etc."*

Au fond, ce qui rend la donnée si précieuse réside dans les limites actuelles des modèles : il est possible de détecter des anomalies sur l'image d'une armoire fibre à un instant T, mais il est impossible de prévoir comment ces dernières vont évoluer. Dès lors, le processus de modélisation est nécessairement cyclique et c'est cette raison qui rend la donnée et les annotateurs si importants pour les entreprises : *"Il y a un paramètre qui fait que les algorithmes doivent évoluer en permanence. Donc ils n'ont pas beaucoup de valeur. Par contre, ce qui fait évoluer ton algorithme, c'est une grosse base de données et celle-là, plus il y en a, mieux c'est."* L'enjeu réside donc dans l'adaptation d'architectures d'IA ou de modèle déjà existants aux enjeux des clients comme le souligne un chargé de communication de l'entreprise CityVeillance qui développe un outil de détection d'éléments sur des images de vidéosurveillance destinées aux communes : *"R : Oui. Si tu veux, on n'a pas inventé la roue, nos ingénieurs ne sont pas partis de zéro sur les algos qu'on utilise, ils réutilisent des éléments et des packages, et nous derrière, on vient juste les adapter par rapport au besoin client, et clairement ce n'est pas de la magie. Ce sont des choses assez simples. C'est juste que l'articulation de tout ça, sous une même plateforme, c'est ça qui a été plus compliqué."* (Arthur - Chargé de communication CityVeillance). Selon lui, le travail de son entreprise est même précisément d'agir comme un intégrateur de solutions d'IA qui sont amenées à

évoluer selon les besoins du client. Toutefois, cette dimension du travail des entreprises d'IA nécessite de mettre en place un travail d'annotation supplémentaire lorsque les besoins des clients n'existent pas déjà dans le produit d'intelligence artificielle.

Encadré n°8 : Trouver des données d'entraînement

Q : Comment vous obtenez vos données ?

R : [les clients ne] nous fournissent absolument rien. On a récolté des données avant et fait des apprentissages pour des projets différents, publics ou privés, et maintenant nos algorithmes ont eu le temps d'itérer et de s'entraîner correctement pour détecter des voitures, des humains. Les cas d'applications ne sont pas très compliqués. Ce qui est compliqué c'est la compréhension des besoins clients pour venir moduler tes algos et faire en sorte qu'ils répondent à ça (...).

Arthur - Chargé de communication CityVeillance

Cet enjeu est également évoqué par le dirigeant de LegalAI qui souligne les limites des larges modèles comme ChatGPT qui, bien qu'entraînés sur de larges corpus de données, n'ont pas été suffisamment "nourris" avec des données spécifiques à un domaine d'application précis. Ainsi, selon lui, *"répondre à des questions légales à partir de Wikipédia (quand ils sont entraînés sur Wikipédia), ça va marcher sur des trucs de salaires parce que ce sont des questions simples, mais dès qu'on tombe dans des domaines très légaux, ces modèles ne marchent pas bien. Et donc je crois qu'il y a un vrai sujet d'essayer d'être tout le temps un peu state of the art dans son domaine, pour nous les avocats, et d'essayer d'offrir l'IA la plus poussée possible."* (Guillaume - CEO - LegalAI). On retrouve une autre justification du recours à des annotateurs parfois experts ou qui, en travaillant sur des projets sur de la longue durée, vont développer une expertise suffisante pour produire des annotations permettant cette adaptation de l'IA au domaine d'application. Il considère en effet que *"l'un de sujets sur le deep learning qui pour moi est le cœur de notre business c'est qu'on appelle le domain adaptation (...) c'est le fait que ces modèles sont très bons, mais ils perdent beaucoup en performance quand ils arrivent sur des domaines hyper spécifiques, typiquement le biomédical ou les avocats."* (Guillaume - CEO - LegalAI).

Au moment de l'entretien, la startup CallAI internalise le travail des données. Ces derniers envisagent toutefois d'avoir plutôt recours à une PNT comme Amazon Mechanical Turk, mais ils se heurtent à plusieurs freins pour le moment. D'abord, la capacité à mettre en place un système de vérification des annotations efficaces nécessiterait une personne de leur équipe à temps plein. Un deuxième frein tient au fait de réussir à interfacer l'outil d'annotation qu'ils ont développé eux-mêmes, pour leurs propres besoins avec la plateforme. Par ailleurs, étant donné qu'il s'agit d'une jeune startup, ils n'ont pas encore réussi à lever les fonds nécessaires à consacrer à l'annotation⁷². Un dernier frein réside dans la difficulté du travail d'annotation,

72. "En plus, on a un outil qui aujourd'hui est utilisable en interne, mais si on voulait enrôler quelqu'un d'autre sur l'outil... On

pour lequel, le CEO souligne la nécessité de produire des données de qualité : *”Ce n’est pas si simple d’annoter. On pense que c’est le truc le plus banal, mais en réalité, il faut être sûr de ne pas faire d’erreurs pour que la base de données ne soit pas dégommée, surtout quand tu es en phase de test et que tu veux prouver que ton concept marche, il faut bien nettoyer ta base de données (...) [Il faut] de la patience. De la concentration, parce que quand tu en as un certain nombre à faire. En fait, pas grand-chose peut te niquer la base de données. Il faut être bien concentré pour ne pas faire d’erreur. Après, ce n’est pas une tâche qui est hyper compliquée. Il suffit d’appuyer sur 3/4 boutons. Au fur et à mesure, puisque c’est moi qui ai fait une bonne partie de la base de données, je sais comment je l’ai faite”* (CEO - CallAI). Comme dans d’autres projets d’IA, deux types de bases de données sont nécessaires. D’abord, une base d’entraînement qui sert à créer le modèle. Dans ce cas, les données sont, selon lui, plus faciles à créer, car il s’agit simplement de prononcer des phrases pour les enregistrer ⁷³.

Il faut ensuite produire des données de vérification. Cette phase d’annotation sert à réentraîner l’algorithme en corrigeant ses détections. Dans le cas de CallAI, cela arrive pour de nouveaux clients lorsque certains extraits sonores comportent des noms de marques ou des noms de produits, en définitive tout ce n’est pas prévu au départ dans le modèle initial. Pour pallier à cela, l’entreprise a mis en place la *”live detection”* qui consiste à retranscrire à la main des extraits audios qui contiennent justement les mots clés spécifiques au client qui ne sont pas pris en compte par le modèle. Dès lors, il est nécessaire de comprendre les enjeux du client final, de connaître les termes qui seront éventuellement prononcés dans le cadre des échanges vocaux entre le consommateur et le conseiller client qui bénéficie de l’outil de CallAI. On y retrouve une tension entre une vision de l’annotation comme quelque chose que tout le monde peut faire et une vision de cette activité comme étant plus compliquée qu’en apparence, notamment parce que les erreurs de compréhension ont un impact très important sur les modèles.

En matière de qualité des données, nous avons également constaté l’importance de l’enjeu des biais de l’IA

a pensé à Mechanical Turk pendant un certain temps. Il faut avoir la sécurité sur cette plateforme. On aimerait bien mettre en place de la double vérification pour que quand ces personnes vont faire de l’annotation, qu’il y ait du monde qui repasse derrière. Il y a un certain nombre de choses à mettre en place pour être sûr qu’il n’y a pas n’importe quoi qui soit fait. D’autant que c’est un système d’annotation un peu particulier, du coup, il faudra enrôler des gens sur notre outil, ce qui probablement rajoutera un peu de complexité (...) on n’a vraiment pas du tout commencé à tester l’annotation externe. On est en cours de levée de fonds. Si on réussit, on envisage même plutôt soit d’avoir quelqu’un dans l’équipe, soit de demander à un peu tout le monde de le faire tout le temps (ce qu’on fait déjà un peu aujourd’hui), que tout le monde annote un peu. Sur ce qu’on appelle les live detections, c’est là où c’est compliqué. Mais sur les annotations liées au training tout le monde peut le faire.”(CEO - CallAI)

73. *”Il y a l’annotation qui vient de la phase d’entraînement, donc les annotations qui sont dites par les opérateurs pour entraîner l’outil. Et il y a les annotations qui viennent de l’écoute en live, une fois qu’ils ont installé l’outil et qu’ils l’utilisent. Et là, ça nous arrive d’avoir des fausses détections, c’est-à-dire des extraits de plusieurs secondes où on nous demande de recadrer l’annotation, mais en réalité il n’y a pas de mot-clé dedans. C’est juste une fausse détection. On l’annote comme tel. Et ça vient alimenter l’outil d’une manière un peu différente. Du coup, il faut avoir une bonne connaissance de l’ensemble des mots-clés qui sont dans la base de données, parce que si on arrive devant un mot-clé et qu’on lui dit que c’est un négatif alors que c’est quelque chose de la base de données, au fur et à mesure, ça va dégrader le modèle. Donc, il faut quand même avoir une certaine connaissance de ce qui a été fait pour bien continuer.”*(CEO - CallAI)

qui peuvent survenir lors de la phase d'annotation. Cet élément est mis en avant lors d'un entretien avec un ingénieur d'Automatik, qui raconte un projet d'IA centré sur la détection des braqueurs de caisse des parking automobile pour une grande entreprise de gestion de ces espaces⁷⁴. Sur ce projet d'IA, lui-même "métis" comme il le précise, raconte qu'il a dû faire lui-même la collecte de données en se rendant dans un parking, habillé en survêtement pour produire des données. Dans ce cas, il s'agit d'extraits vidéos captés par une caméra de vidéosurveillance. Le système de détection marchait "trop" bien, surdétectant systématiquement les personnes racisées. Corentin précise que c'est sur cet aspect précis que l'annotation peut avoir un fort impact sur le fonctionnement de l'IA, y compris en introduisant des biais "culturels". Face à ce constat d'une IA devenue "raciste" du fait de la méthode d'acquisition et d'annotation de données, Automatik a mis en place un autre protocole pour mitiger ces biais. Ils ont demandé à des étudiants d'une école d'ingénieurs française, tous blancs, d'aller devant les caméras du parking en survêtement afin d'apprendre au modèle, non pas à détecter une couleur de peau, mais des vêtements (en l'occurrence les joggings)⁷⁵.

Encadré n°9 : Tenter de maîtriser les biais algorithmiques

Les gens étaient super contents, le réseau de neurones était ultra-bon. Moi, je ne peux plus rentrer dans un parking Indigo ! En fait, l'IA fait ce qu'on lui demande de faire sur des données biaisées et donc on a une intelligence [biaisée]. Là, l'IA est vraiment raciste. Autre sujet de sociologie très intéressant ! D'une certaine manière, il peut y avoir des biais culturels qui peuvent être induits (...) En prenant des gens des Ponts, ça équilibrait le truc. On voulait que le réseau de neurones apprenne juste les vêtements et pas la couleur de peau. Mais il fait ce qu'on lui demande de faire. Ça touche à l'annotation aussi. Entre un bon comportement et un mauvais comportement, il y aura forcément un biais culturel dans la décision.

Corentin - Ingénieur - Automatik

Corentin souligne bien les biais qui résident dans le travail d'annotation, ce qui rejoint l'enjeu de s'assurer d'une proximité entre la production des modèles et des données. Eleonor (AIVoice), parle quant à elle de sensibilité des modèles par rapport aux données, lorsqu'elles ne sont pas de bonne qualité. Elle explique que si ces liens là n'existent pas, les modifications faites sur le modèle mettraient plus de temps à se répercuter sur le travail d'annotation. De la même manière, dans le cas où des erreurs se glissent dans le travail d'anno-

74. "On a eu un sujet intéressant à un moment c'est à quel point les biais dans l'annotation modifient les réseaux de neurones et impactent la société ou le produit. Un exemple : à un moment donné, on travaillait avec Indigo pour repérer les actions qui se passaient au niveau des caisses. Le but principal était de détecter les gens qui allaient casser les caisses pour récupérer la monnaie. Du coup, il fallait faire de l'acquisition de données. Moi je suis métis et j'y suis allé habillé en survêt Puma, chaussettes sur le pantalon, enfin le cliché..."(Corentin - Ingénieur - Automatik)

75. Notons ici, que, si le nouveau protocole tient compte des principes de *data justice* d'équilibre des jeux de données en termes de représentation des catégories de population, le principe initial qui est à l'oeuvre derrière le modèle reste critiquable : l'entreprise cible les personnes habillées en survêtement avec tout ce que cela comporte également comme biais.

tation, il sera plus simple de les comprendre et de les corriger lorsqu'il est possible d'organiser des réunions directement avec les annotateurs. En effet, les *"algorithmes sont très sensibles aux données qu'on injecte et donc si on injecte des données qui ne sont pas de très bonne qualité, c'est normal que les résultats ne suivent pas. C'est logique. Ça nous fait perdre beaucoup de temps de réaliser qu'en fait cet algorithme ne marche pas très bien parce qu'il y a quelques erreurs dans l'annotation et ça c'est très chronophage, surtout quand on parle de volume de données très élevé. On ne peut pas revenir en arrière sur 2 000 heures."* (Eleonor - CTO AIVoice).

Le lien entre la production de données de qualité et l'intégration logistique des travailleurs est également souligné par le dirigeant d'une entreprise d'annotation de Madagascar qui met en place une offre basée sur les interactions régulières entre les travailleurs, leurs managers, les dirigeants du sous-traitant B et les entreprises d'IA. Selon lui, l'intérêt de faire appel à un sous-traitant direct pour les *data scientists* est que ce dernier va mieux comprendre le processus de production de l'annotation, et donc mieux en tenir compte dans l'appréciation des performances du modèle.

Encadré n°10 : Sécuriser la qualité des données par le processus de production

Dans la mesure où le fait de vraiment sécuriser la manière dont l'annotation est produite ça permet aux data scientists qui entraînent son modèle de comprendre vraiment pourquoi son modèle se comporte ainsi. (...). Si on connaît les données, si on connaît les conditions dans lesquelles ces données ont été préparées, si on sait qui l'a fait, selon quel process, avec quel niveau de formation, etc., on est mieux capable de juger si Il y a tel biais dans mon modèle, mais en fait je comprends pourquoi... parce que dans le process on avait dit de faire comme ça, etc. . En tout cas, c'est notre positionnement qui est très orienté qualité et projet complexe.

Victor - CTO - Sous-traitant B

En effet, la performance du modèle est largement liée à la qualité du processus d'annotation : *"on va souvent parler des faux négatifs et des faux positifs qui sont des erreurs, et souvent on en parle pour les modèles, mais ça peut être répliquable à la tâche d'annotation manuelle. Quelqu'un qui voit quelque chose et qui l'annote alors que ce n'était pas la vérité va générer un faux positif et inversement, s'il voit quelque chose et qu'il se dit Il ne faut pas que je l'annote , ce sera un faux négatif."* (Victor - CTO - Sous-traitant B) L'enjeu de la qualité de l'annotation est également ressorti d'un entretien réalisé avec Dinaso, la dirigeante du sous-traitant E, qui produit des annotations pour des entreprises américaines et qui fonctionne de la même manière que le sous-traitant B. Dans le contexte du développement de la voiture autonome, le client a proposé un guide d'annotation avec des concept clés à annoter comme un *"échangeur d'autoroute"* ou une *"aire suburbaine"* qui ont été mal compris par les travailleurs malgaches. La dirigeante a donc dû énormément échanger avec le client pour préciser les définitions de ces concepts afin de faciliter la formation des annotateurs et donc

la qualité finale des données annotées.

Encadré n°11 : Les biais potentiels de l'annotation de la voiture autonome

La question des biais s'est posée très vite (...) Au début, c'était moi qui faisais les formations et on voyait le gap entre ce qu'on recevait de formaté et ce qui était au final compris. On a dû faire énormément d'aller-retour pour pouvoir leur dire Ça, ça ne marche pas ; vous nous parlez d'une autoroute à Chicago pour les véhicules autonomes ; c'est bien, vous voulez faire 200 000 annotations sur ce que c'est qu'une autoroute, mais vous ne nous définissez pas ce qu'est une autoroute. Vous écrivez juste interchange ou suburban area in Chicago , mais comment on définit ça ? En fait, c'est dans ces lancements-là qu'ils ont vu la différence.

Dinasoa - CEO - Sous-traitant E

Guillaume évoque le même type de problème au sujet des annotations de contrats juridiques. Ainsi, en demandant à des annotateurs présents sur Amazon Mechanical Turk de sélectionner une date, il indique que le problème réside dans la variabilité des annotations produites : *"Principalement, ce que je reproche à MTurk c'est que, on mettait trois personnes en parallèle sur les mêmes data, et on avait trois réponses différentes. C'est un énorme problème. La deuxième chose, c'est que c'est difficile de forcer, par exemple si je demande la date de début, certains vont sélectionner juste la date, d'autres la phrase, ou avec les parenthèses."* Il souligne en effet la difficulté de faire comprendre des consignes aux annotateurs surtout en tenant compte des ambiguïtés potentielles de l'annotation : *"Donc il faut faire des tas de guidelines, des documents énormes sur le sujet. On était en train d'embaucher quelqu'un qui est un ancien avocat/entrepreneur qui est devenu notre CEO et rapidement il a commencé à faire de la data et dire ce qu'il fallait faire."* (Guillaume - CEO LegalAI). C'est pour cette raison qu'il a préféré recruter des personnes présentes physiquement dans ses locaux ⁷⁶

Pour Jean, qui s'occupe de développer un outil de détection automatique des styles vestimentaires à partir d'images pour un site de vente de vêtements de seconde main, l'enjeu de l'annotation réside dans la capacité à tenir compte des "erreurs" de classification. En effet, lorsqu'il évoque la relation qu'il entretient avec ses annotatrices à Madagascar, prédomine l'idée que leur travail consiste à annoter et à identifier les styles vestimentaires qui posent problème, soit, parce qu'ils n'ont pas été prévus au départ, soit parce qu'il y aurait des styles trop proches les uns des autres. Dès lors, la gestion de l'annotation doit accorder de l'importance à ce qu'il qualifie d'"erreur" : *"le fait qu'il y ait une erreur c'est un sujet à caler, ce n'est pas un problème vous voyez ce que je veux dire ? parce que en fait, ce business là c'est de travailler sur des erreurs, c'est d'essayer de*

76. *"Et on s'est dit que ça serait plus simple d'avoir des gens en physique (c'était avant le Covid), et de voir les contrats avec eux et leur dire comment sélectionner (...) ça permet aux gens de tester et de voir comment l'annotation fonctionne. Donc c'est comme ça que ça s'est construit et ce qu'on appelle la data team, c'est l'équipe qui aujourd'hui utilise notre outil pour produire de la data comme le feraient nos clients, mais sur un certain type de contrats ou dans un domaine particulier."*(Guillaume - CEO LegalAI)

faire moins erreur, mais ce n'est pas vraiment un problème de faire des erreurs" (Jean - CTO - ModeAI). C'est le niveau de qualité attendue qui justifie le recours à des travailleurs avec lesquels Jean peut échanger au quotidien : *"Du coup on peut discuter un peu librement des erreurs, des problèmes et cetera ... d'une certaine façon tu as un engagement beaucoup plus important, les gens comprennent ce que l'on veut, et à terme nous on aura une meilleure qualité"* (Jean - CTO - ModeAI).

Encadré n°12 : La maîtrise de la qualité par la gestion des erreurs

J : Du coup on ne dramatise pas la notion d'erreur, on essaye de voir si cette erreur-là et de plus en plus fréquente alors quel est le véritable problème pour la résoudre (...), elles, elles sauront qu'elles ont travaillé non pas à labelliser mais à labelliser pour Miroa, une boîte qui fait de l'IA sur de la fringue de 2nde main en France - c'est aussi très engageant pour eux quel est l'impact de leur travail, et pas juste là pénibilité de la récurrence de la labellisation. C'est un cercle vertueux. [Sur Mturk] (...) c'est pas forcément un gage de maîtrise de la qualité et cetera ... c'est la vieille solution de labellisation etc. ... Là, notre sujet c'était plus de traiter la subjectivité avec le plus haut niveau de qualité potentiel, donc on n'allait pas passer sur Amazon mechanical turk

Jean - CTO - ModeAI

Souvent, les entreprises tâtonnent dans le choix d'un processus d'annotation, comme l'illustre le cas de Mathilde chez BankAI. Elle évoque également la nécessité de produire des données de qualité, grâce au recours à des BPO qui facilitent le contrôle des travailleurs. Comme chez d'autres entreprises d'IA étudiées, c'est aussi la comparaison avec les plateformes numériques de travail qui va guider le choix de l'entreprise : à la fois d'un point de vue éthique, du point de vue de la qualité et de l'accompagnement à l'annotation proposé par l'entreprise finalement choisie. Dès le départ du projet, l'utilisation d'une PNT n'est pas envisagée pour des raisons de qualité des données, car il n'y a pas forcément de contrôle sur ces dernières. Comme dans le cas de AIVoice, la capacité d'avoir une proximité avec le sous-traitant permet de réajuster la force de travail de manière qualitative (changement dans l'annotation) et quantitative (nombre d'annotateurs), en fonction des besoins de la startup. Cela permet également d'exercer un plus grand contrôle sur les données produites⁷⁷. Mathilde analyse également ce choix au regard du type de données dont l'entreprise a besoin : des données textuelles. De fait, malgré l'écriture d'un guide d'annotation, cette dernière considère que les éléments textuels sont soumis à de fortes ambiguïtés⁷⁸. Dans ces cas-là, les annotateurs ne savent pas for-

77. "R : Donc on a trouvé un compromis. À un moment, il y avait des Indiens aussi. Je ne me souviens plus de leur nom (...) c'était une boîte. On ne voulait pas utiliser les plateformes. Le crowdsourcing (...) on ne voulait pas faire ça parce qu'on avait besoin d'un truc de qualité. (...) Il n'y a pas de contrôle. [Q : Dans quel sens ?] (...) Pas de contrôle du travail qui est fait. On ne peut pas réajuster en fonction de nos besoins. On n'a personne avec qui communiquer"(Mathilde - CTO - BankAI)

78. "On débutait donc on n'avait pas des tâches hyper précises. C'est du NLP donc quand même vachement plus compliqué. Il y a beaucoup d'interprétations dans nos guidelines."(Mathilde - CTO - BankAI)

cément ce qu'est la bonne réponse en termes d'annotation, ce qui nécessite d'être en mesure d'intervenir rapidement en cas de mauvaise compréhension du guide. Un enjeu plus mineur est la dimension éthique du recours à des plateformes numériques de travail, perçues comme offrant de moins bonnes conditions de travail. Mais l'argument déterminant est le fait que BankAI n'a pas beaucoup d'expérience en matière d'annotation. Le recours à une entreprise de sous-traitance dont l'annotation est la spécialité l'a aidé dans la mise en place du processus de production de l'annotation que ce soit dans la mise en place d'un outil d'annotation ou bien pour formaliser un guide d'explication des tâches à effectuer par les annotateurs⁷⁹.

Ces différents exemples éclairent les difficultés des entreprises d'IA à mettre en place le processus d'annotation le plus pertinent possible. Ces dernières sont parfaitement conscientes des enjeux qui résident dans la qualité des données du fait de son impact sur la qualité des modèles. Par conséquent, les employés d'entreprises d'IA interrogés valorisent une organisation du travail qui respecte certains principes d'intégration des travailleurs de l'annotation à leur processus de production, ce qui limite donc le recours aux plateformes numériques de travail et donc au *marché pur*. A rebours de la littérature sur le *micro-travail*, on constate un recours accru à de la sous-traitance directe. Comme l'indique Victor, dirigeant d'une entreprise de sous-traitance, le marché de l'IA évoluant vers une meilleure compréhension des enjeux liés aux biais de l'annotation et vers des projets plus complexes, certaines entreprises d'IA tendent à valoriser le fait d'intégrer plus les annotateurs à leur processus de production : *"Je pense qu'il y a pas mal de déconvenues avec le crowdsourcing. L'IA est un domaine nouveau, l'annotation de données est un domaine encore plus nouveau. Ça a commencé de manière très basique J'ai des données, il faut les annoter, qui veut les annoter ? . Au fil du temps, le marché se structure, les besoins se précisent et on comprend mieux les biais qu'il peut il y avoir dans l'annotation de données. Le côté j'envoie mes données et annotez-les , ça commence un peu à se perdre et on arrive plus dans les vrais sujets d'annotations de données qui sont un peu plus complexes et c'est peut-être plus difficile à faire en crowdsourcing parce qu'on ne connaît pas les gens, etc."* (Victor - CTO - Sous-traitant B)

Comme on le verra par la suite, cela passe par le recours à des entreprises de BPO et par l'utilisation d'outils numériques pour faciliter le contrôle avec les travailleurs, ainsi que les échanges concernant les données. Dans certains cas, cela passe également par un accompagnement technique par les entreprises de sous-traitance. Dans cet entretien, le dirigeant d'une des entreprises les plus importantes d'annotation résume parfaitement cette organisation intégrée du travail en évoquant un important client, une entreprise américaine de e-commerce, très active dans le monde de l'IA. Les équipes de développement de ce dernier sont bien conscientes de l'impact du travail d'annotation sur les modèles d'IA, au point de considérer la compré-

79. "Q : Et il y avait d'autres raisons pour lesquelles ne pas choisir les plateformes ? R : Après, je n'avais pas envie personnellement. Parce que le micro task, j'en avais entendu parler(...) pour des questions éthiques. Mais ce n'est pas l'argument que j'ai donné. J'ai donné le contrôle qualité (...) surtout vu le peu d'expérience qu'on avait là-dedans et vu à quel point ils nous ont aidés, la boîte d'annotation, à bien mettre en place notre projet. Heureusement qu'on avait ça, sinon on aurait fait que de la merde." (Mathilde - CTO - BankAI)

hension des annotateurs comme un axe majeur du développement des algorithmes.

Encadré n°13 : Assurer un lien entre l'annotation et les développeurs d'IA

On est tombé dedans et donc on s'est retrouvé avec des équipes [de développeurs] qui mettent une valeur importante sur l'humain, parce qu'au final pour eux, l'annotateur est au centre de tout, et jusqu'à aujourd'hui. S'ils pouvaient lui parler directement, ils lui parleraient, parce que c'est lui qui est la clé de leur modèle. C'est sa compréhension qui est la clé de leur modèle. Et c'est ça qui nous a plu. On s'est rendu compte qu'il y avait un lien très fort entre ces équipes de recherche et ces personnes qui désignaient ces modèles [les data-leads/chercheurs] avec nos annotateurs. C'est là où on s'est dit On ne va pas juste être un prestataire, on ne va pas juste faire un sweatshop, mais on a un coup à jouer pour changer la manière dont on appréhende l'industrie.

Patrick - CEO - Sous-traitant E

Les entreprises d'IA commerciales se basent largement sur la recherche scientifique tout en se concentrant principalement sur des applications concrètes. Cet élément les pousse à se distinguer du monde académique par la mise en place de stratégies qui visent à perfectionner les modèles par la qualité de la source de données et non par l'amélioration des modèles en eux-mêmes. Nos résultats montrent alors que le traitement des données est d'autant plus essentiel à la production et à la maintenance de l'intelligence artificielle, que c'est précisément ce qui contribue à produire des services commerciaux fonctionnels par l'adaptation de modèles pré entraînés à des contextes et à des données spécifiques. Les entretiens mobilisés dans cette sous-section suggèrent également que cette adaptation passe parfois par le fait de compter sur des travailleurs des données disposant d'une expertise suffisante pour aborder le phénomène que les entreprises cherchent à automatiser. La stabilité de la force de travail est enfin justifiée par les évolutions constantes des modèles, ainsi que des phénomènes automatisés.

Si la littérature a commencé à documenter les apports des micro-travailleurs dans la construction des bases de données, peu de chercheurs sont capables de retracer l'ensemble du processus de production des systèmes d'IA dans un contexte d'entreprises. Ainsi, les études sur des projets comme ImageNet se focalisent uniquement sur le monde scientifique (Paullada et al., 2020). Dans d'autres enquêtes qui portent sur les laboratoires d'IA, cette phase préparatoire de l'IA ne reçoit pas beaucoup d'attention (Jaton, 2017). Nos résultats suggèrent alors plusieurs implications quant à l'analyse du processus de production de l'IA, que ce soit dans un contexte académique ou bien de politiques publiques. Ainsi, il paraît essentiel que les méthodes d'audit et d'explicabilité intègrent cette phase de préparation de données. En effet, comme on le verra dans la sous-section suivante, cette phase n'est pas exempte de risque de biais.

Les salariés des entreprises d'IA comme les dirigeants des entreprises de sous-traitance sont parfaitement

conscients des risques potentiels de biais liés à la phase de production des données. Ces risques justifient d'après eux le recours à des entreprises formelles, car ils considèrent que ça va permettre de limiter les erreurs des annotateurs en facilitant le travail répercussion des choix du modèle sur les annotateurs. Néanmoins, comme le cas raconté par Corentin le suggère, il nous semble peu pertinent de parler biais algorithmiques, mais plutôt des décisions humaines qui orientent le comportement du modèle vers certaines interprétations du problème à automatiser. D'une certaine manière, il n'est donc pas uniquement question de l'annotation en elle-même, mais aussi et surtout des décisions prises lors du processus de production du modèle. De ce point de vue, nos résultats contribuent à la fois à la littérature en sociologie de l'IA et en informatique et invitent à mieux prendre en compte trois éléments rarement identifiés comme déterminants par la littérature scientifique (Miceli, Posada, & Yang, 2022) autour de la "socialisation du modèle". D'abord, l'organisation du travail des données et ses impacts sur la qualité des données⁸⁰. Ensuite, l'évolution des décisions qui contribuent à façonner le modèle et les raisons de ces évolutions. Enfin, le processus de documentation qui permet de construire des "standards" d'annotation et les inscrire dans des documents qui cadrent le travail des annotateurs. Ces éléments feront l'objet d'un développement au cours de nos prochaines sections.

4.1.3 Maîtriser le processus d'annotation pour tenir compte de la volatilité des modèles d'IA

Dans le contexte de l'intelligence artificielle francophone, les entreprises se heurtent à plusieurs problèmes liés à la disponibilité des données pour l'entraînement des modèles. En comparaison avec la recherche publique et les grandes entreprises du numérique, la quantité de données accessibles pour l'entraînement de modèles qui s'appliquent à des cas d'usages spécifiques est plus réduite. Ainsi, dans le contexte de la recherche, les jeux de données sont souvent rendus publics. Ils sont d'ailleurs aussi utilisés par les entreprises d'IA qui soulignent leurs limites pour leurs propres usages. Les grandes entreprises d'IA comme Facebook ou Google ont des approches différentes, qui se basent sur des dispositifs d'acquisition de données intégrés à leurs services⁸¹. La taille de ces firmes et la variété des données collectées constituent des actifs déterminants auxquels n'ont pas accès les entreprises étudiées dans cette thèse.

Un autre enjeu associé au dispositif de collecte des données concerne le type d'IA déployé commercialement. L'exemple de la caisse automatique de cantine illustre à quel point le cela influence la collecte des données. Dans ce cas, l'utilisation de la caisse de cantine par les consommateurs peut nécessiter de nouvelles phases d'annotations liées à l'émergence de nouveaux éléments sur les plateaux-repas. On comprend également que la nécessité de perfectionner constamment les modèles impacte fortement le travail d'annotation. Comment les données sont acquises, travaillées, transformées et incorporées dans les modèles par

80. Et bien évidemment sur les travailleurs

81. C'est précisément l'objet de toute la littérature qui concerne les mécanismes d'extraction de valeurs basées sur le travail des internautes

les entreprises d'IA? Comment circulent-elles au sein de la chaîne de production? Au cours de cette sous-section, nous proposons d'analyser comment ce travail est organisé pour maîtriser la volatilité des modèles d'IA.

Sur l'acquisition de données d'entraînement, deux cas de figure typiques émergent de l'enquête. D'abord, l'utilisation de modèles pré-entraînés basés sur des jeux de données ouvertes, puis le recours aux données des consommateurs finaux (les utilisateurs) pour déployer le modèle en production et perfectionner son fonctionnement. On a pu constater ce modèle chez AIMedic avec une phase de ré-annotation des données clientes⁸². C'est aussi le cas chez AIVoice, dont les employés analysent les conversions de leur robot d'appel avec les consommateurs. Les données sont récupérées et annotées afin d'améliorer le fonctionnement des modèles⁸³.

Le modèle de LegalAI est basé sur des modèles déjà existants et entraînés sur des jeux de données bien connus dans le domaine du traitement du langage (notamment Wikipédia). Guillaume explique d'ailleurs bien la course à l'entraînement des grands modèles d'IA génératifs tout en soulignant les limites pour des applications concrètes : *"C'est pour ça que pendant un an, tout le monde s'est battu, en faisant des modèles de langage beaucoup plus gros, pour qu'ils voient encore plus d'informations et avoir des résultats encore plus hallucinants. Mais aujourd'hui on est à la limite de ça, puisque le dernier modèle GPT-3, je crois que c'est 10 millions de dollars de training (...) donc ce n'est pas viable, mais c'est intéressant d'un point de vue recherche."* (Guillaume - CEO LegalAI) Pour cette entreprise, l'enjeu était plutôt d'aboutir à un premier produit déjà utilisable, d'avoir accès à leurs premiers clients afin de collecter, par ce biais de plus en plus de données : *"Évidemment, plus on a de clients qui nous utilisent, plus on a de data et plus le modèle devient intéressant."* (Guillaume - CEO LegalAI)

Un autre cas de figure est lié à l'absence de dispositif de collecte de données et concerne généralement les applications d'IA à destination des entreprises et non des consommateurs. Ces grandes entreprises n'ont pas nécessairement, ni un jeu de données prêt à l'emploi, ni un dispositif de collecte de données déjà utilisé au sein de l'entreprise. Chez Automatik, ce cas s'est présenté lors d'un projet de détection des cartouches de gaz hilarant dans les incinérateurs des déchetteries. L'objectif du modèle est de trier ces cartouches avant qu'elles n'arrivent dans l'incinérateur et perturbent son fonctionnement. Or, le client ne dispose pas d'un dispositif d'acquisition de données (appareil photo) sur la machine de tri et le tapis de tri fonctionne de manière continue ce qui rend plus complexe la constitution du jeu de données. L'entreprise donneuse

82. *"Plus marginalement, de temps en temps, on a besoin d'avoir des données spécifiques. La plupart du temps, on utilise nos bases de données clients, mais parfois il faut plus de choses."* (Cédric - Lead data Team - AIMedic)

83. *"Donc toutes ces interactions, nous avons les audios et nous avons aussi les transcriptions, parce qu'on donne aussi les transcriptions écrites au client. On a donc toutes ces données et cette expertise et c'est pour ça qu'on a mis en place un pôle data qui récupère rapidement les données vocales et textuelles pour les annoter et bien les catégoriser pour entraîner ensuite nos algorithmes."* (Eleonor - CTO AIVoice)

d'ordres a donc demandé à ses employés d'arrêter le tapis pour prendre en photos ces cartouches de gaz⁸⁴. Ces images seront ensuite traitées à Madagascar, par le sous-traitant de l'entreprise d'IA⁸⁵. Automatik met également à disposition du client, une plateforme permettant d'ajouter de nouvelles catégories d'objets à détecter⁸⁶. Pour détecter les nouveaux déchets, il faudra alors repasser par une phase d'entraînement et de collecte des données. Cet exemple montre que le processus d'acquisition de données sert à obtenir des données qui s'appliquent aux cas précis traités par les clients finaux. L'objet est donc de collecter une "vérité terrain" qui est ensuite destinée à être annotée (Jaton, 2017). Comme on l'a vu dans notre premier chapitre, l'enjeu pour les entreprises est d'adapter des modèles pré-entraînés à leurs cas clients (*fine tuning*). C'est au cours de ce processus que vont être définies les caractéristiques importantes du modèle.

Dans le cas de AIVoice, le constat initial provient d'une inadéquation des modèles pré-entraînés du marché avec les besoins de l'entreprise. Par exemple, *"la reconnaissance alphanumérique est assez difficilement compréhensible nativement par ce type de solutions"* (Patricia - AIVoice), alors que leur robot téléphonique est justement censé comprendre les informations données par les clients (par exemple : un numéro de compte...). La solution réside alors dans des solutions d'IA développées en internes.

Dans toutes les entreprises, on retrouve également une phase de problématisation du modèle qui implique plusieurs acteurs de la chaîne. Chez AI Voice, par exemple, une première étape concerne d'abord la partie commerciale. Le client exprime ses besoins et un chef de projet doit détailler ces besoins de manière plus approfondie : *"Un projet type commence généralement chez les commerciaux (...). Ensuite, l'expression de besoins faite par le client passe [par] un chef de projet qui va assurer la communication avec le client, il va faire le premier planning, il va répartir l'expression de besoins en tâches détaillées avec des deadlines, et ensuite on commence à composer l'équipe"* (Eleonor - CTO AIVoice). D'un point de vue technique, chaque projet associe *"un chef de projet, un développeur en conversation (et) un développeur back-end pour tout ce qui est intégration*

84. *"(...)Ils vont nous dire oui, mais attendez, comment ça marche l'IA ? Il faut combien d'images de déchets non conformes dans l'usine pour avoir un détecteur qui fonctionne ? Et si je ne les ai pas, je fais comment ? (...) c'est une question opérationnelle. J'ai besoin de 1 000 fois une image d'une cartouche de gaz hilarant dans mon incinérateur pour construire un dataset ; concrètement, on fait quoi ? (...) la réponse, ça a été comme on ne les voit pas passer et que ça tourne en flux continu, on n'arrivera pas à les prendre en photo, donc on va intentionnellement mettre des cartouches, stopper le tapis roulant au moment où on la voit passer, prendre les photos, et constituer un dataset comme ça."*(CEO - Automatik)

85. *"[Q : Dans un cas comme ça, vous intervenez sur l'annotation par exemple ?] Oui, exactement. Ça dépend, s'il y a 500 images à se faire et que c'est un besoin très ponctuel, ok. Si c'est un truc où on sent que ça va être dur et qu'il faut au moins 5 000 images par catégorie et qu'il y a entre 5 et 10 nouvelles catégories à renseigner, là, personne ne veut le faire chez eux, c'est sûr. Donc dans ce cas, il faut outsourcer et on crée une petite prestation de services et on les aide sur ça (...) On leur dit vous êtes sûrs que pour 5 euros de l'heure, vous n'avez pas envie d'envoyer un petit brief et d'envoyer 50 000 images de cartouches de gaz hilarant, et d'avoir des gens qui ont l'habitude de bosser sur ce genre de thématiques sur notre outil, qui sont consciencieux et qui vont vous faire un dataset aux petits oignons ?"* (Théophile - CEO - Automatik)

86. *"À un moment, il va y avoir une mode cartouches de gaz hilarant. Il faut ajouter cet objet aux objets qui sont reconnus et considérés comme non conformes. (...) Le mieux, c'est d'avoir un datascientist qui bosse en équipe avec le mec du terrain, et ensemble, ils vont bien comprendre le problème métier, et bien faire évoluer leur solution, pour intégrer cette nouvelle composante."* (Théophile - CEO - Automatik)

téléphonique”. Un pôle ”data” intégré à un pôle recherche et développement intègre des annotateurs qui évaluent la qualité des conversations ⁸⁷.

Une fois le modèle créé, une phase de test est réalisée au sein de l’entreprise cliente, dont l’objectif est de faire remonter les problèmes du robot conversationnel. Au cours de cette phase, le client évalue la qualité des réponses données par le modèle en fonction des questions qu’il souhaite traiter par ce biais, c’est-à-dire les ”scénarios de conversation”. Durant cette phase, il s’agit déjà d’un travail de reconceptualisation du modèle qui va évoluer en fonction du test réalisé par le client. Une fois le modèle mis en production, les annotatrices interviennent à nouveau pour superviser les conversations du robot conversationnel ⁸⁸. Cette supervision du modèle se focalise principalement sur *”les conversations qui sont non comprises”, soit quant ”une personne appelle et que (le robot) ne comprend pas sa demande”* (Patricia - AIVoice). L’objectif étant à nouveau de faire évoluer le modèle pour intégrer ces retours. Les annotatrices participent également à prendre en charge l’évolution des conversations : *”(on va) faire intervenir les annotatrices et également les bot trainers dans l’équipe qui prennent en charge le machine learning justement sur l’évolution des conversations, ce qui est très important au lancement d’un bot et tout au long de sa vie aussi.”* (Patricia - AIVoice).

Ces évolutions peuvent également être liées à la transformation du ”problème” que le client essaye d’automatiser. Dans le cas de ce robot conversationnel, les possibilités d’interaction entre un consommateur et le robot peuvent changer lorsque l’objet des conversations évolue. Dans le cas de AIVoice, un modèle a été déployé pour prendre en charge une partie des appels de la plateforme d’information sur le Covid-19. Or, dès que de nouvelles règles s’appliquent, il était nécessaire de repasser par une phase de travail autour des scénarios de conversation : *”Ce qui est intéressant c’est la phase de lancement d’une nouvelle thématique, par exemple si M. Castex devait annoncer un nouveau confinement ce soir, la branche reconfinement est déjà prévue chez nous, mais il suffirait que ce reconfinement ait des conditions particulières et que l’on doive traiter des catégories d’appels spécifiques (...) il va falloir que l’on adapte tout de suite les réponses du bot et que l’on vérifie cette adaptation en passant par des étapes classiques (on fait une proposition, le client accepte la formulation et l’arborescence qu’on lui présente dans le logigramme pour cette nouvelle thématique, après on fait une phase de recette, on teste, et une fois que c’est recetté on met en production)”* (Patricia - AIVoice)

Cette ”problématisation” des modèles se retrouve également chez d’Automatik qui, à propos de la caisse

87. *”Ensuite, on a des intervenants ponctuels sur des projets, qui viennent de la R&D souvent et qui assistent le chef de projet dans son évaluation de la conversation ou d’une version de la conversation et faire des tests, un peu un cahier de recettes pour tout tester. On a une équipe dédiée à ça qui s’appelle l’équipe amélioration continue qui est aussi chargée de faire les recettes, les tests, et voir que tout va bien avant une mise en production.”*(Eleonor - CTO AIVoice)

88. *”Q : Ça veut dire que cette équipe va tester que les scénarios types de conversation sont bien compris par le callbot, c’est ça ? R : Exactement, oui. Un projet dure en moyenne deux à trois mois chez nous. Ensuite, il y a une phase recette avec le client où il nous remonte les problèmes qu’il a rencontrés pendant la phase recette et qu’on améliore. Une fois qu’on a le go du client, on le met en production et on continue à superviser les conversations, les comportements des clients pour signaler s’il y a un problème, s’il y a quelque chose à revoir dans la conversation, dans le wording du bot.”* (Eleonor - CTO AIVoice)

automatique de cantine, nous explique que mettre *”en prod ton AI c’est la partie visible, tu as toute une partie invisible qui est un changement de métier (...) et un projet complexe d’IA tu en as pour deux ans et à chaque fois il faut les réentraîner, redéployer les connecter à de nouvelles données”* (Julien - CTO Automatik). Au fil des projets, il faut intégrer progressivement les nouvelles données, c’est-à-dire les *”images difficiles (qui remontent du terrain)”*, qu’il faut également annoter. Dans le cas de la caisse automatique de cantine, leur CTO nous donne l’exemple de la reconnaissance des contenants qui requiert une évolution des modèles : *”on voulait reconnaître les assiettes, tu te rends (compte) que tu as des cocottes donc il faut prévoir quelque chose, ça ça remonte au fil de l’eau.”* (Julien - CTO Automatik).

Cela explique également que les entretiens réalisés auprès des employés de startup mettent en lumière la longueur des projets d’intelligence artificielle. Entre la signature avec le client et un premier exemple de modèle non déployé, il peut déjà se passer six mois, tout en sachant que ce temps sera corrélé à la complexité du projet et à la masse de données disponibles pour ce cas d’application. Comme dans le cas de AIVoice, Corentin, ingénieur d’Automatik rencontré plusieurs fois au cours de notre travail de terrain, explique le processus de production des modèles en insistant sur les allers et retours entre le client, les développeurs et les *data scientists* de chez Automatik, et le travail d’annotation. Dans le cas d’Automatik, cela conduit d’ailleurs le CEO à se poser la question de l’importance du modèle par rapport aux données, dans la chaîne de valeur. Selon lui, la valeur première du travail de l’IA concerne toutes les activités liées aux données. Le modèle évolue avec le problème qu’une entreprise cherche à automatiser. Mais, pour automatiser, il est nécessaire de s’assurer d’avoir accès à des données annotées pour perfectionner et adapter le modèle.

Encadré n°14 : Le processus incrémental d’adaptation du modèle à la réalité du terrain

(...) On va entraîner un premier réseau de neurones sur les images annotées. On va évaluer l’application sur les images, donc est-ce que les performances sont acceptables. Et là si elles sont acceptables on va directement mettre l’application en production et les clients vont le filer à quelques utilisateurs triés sur le volet et on va avoir les retours terrain. En même temps qu’on a les retours terrain, on va pouvoir comme ça engranger plus d’images et on va réutiliser les réseaux de neurones pour détecter les images problématiques et les réannoter et donc on va pouvoir réentraîner des nouveaux réseaux de neurones qui seront plus performants. Après deux/trois mois on va dire de POV/POC, si nous on a atteint les objectifs qu’on a définis en amont : rapidité, performance (...)

Corentin - Ingenieur - Automatik

Au sein de ModeAI, la question du *fine tuning* est directement adressée dans notre entretien par les deux fondateurs au travers de la tension entre quantité de données et perfectionnement du référentiel d’annotation des données. De manière schématique, poser cette question revient à se demander sur quel aspect d’un

projet d'IA insister le plus : l'acquisition d'une masse importante de données à trier ou la définition préalable des concepts qui s'appliquent à l'annotation de données. Avec des jeux de données très volumineux utilisées dans des modèles pré entraînés vendus par Google ou disponibles dans un contexte de recherche, ils ont *"très vite compris qu'il fallait refaire un nouveau dataset propre, parce que sinon en fait on plafonnait sur du 75, 77, 78% de précision sur les algos avec les datasets publics qui existent aujourd'hui sur la [mode]"* (Jean - CEO - ModeAI). Ce jeu de données est basé sur des images publiques qui sont réannotées en fonction de leurs besoins et notamment la détection des styles vestimentaires.

Encadré n°15 : Créer un référentiel pour standardiser la reconnaissance des vêtements 1

en gros si tu manges de la merde, en sortie tu as de la merde ... tant que tu n'as pas une donnée propre, tu ne peux rien faire en fait. On est parti de l'idée : soit tu faisais tout à la main, et tu créais en fait un référentiel universel, que tout le monde discute, et que tu te tapes tout à la main pour mettre tout d'équerre - c'est un beau rêve, mais ça ne marche pas - soit on part de notre référentiel visuel que on a nous-mêmes normalisé, et c'est l'IA, via la reconnaissance d'images, qui va permettre de labelliser via ce référentiel nos données. Donc ça amène à la suite, c'est là qu'on est parti avec l'IFM (...) l'institut français de la mode

Jean - CEO - ModeAI

Pour créer le référentiel, ils développent une collaboration avec une étudiante de l'Institut Français de la Mode (IFM) : *"l'étude de l'IFM et ça nous a encore plus montré qu'il fallait vraiment avoir un truc hyper carré, le référentiel le plus carré possible, parce que sinon c'était (pas bon)"* (Jean - CEO - ModeAI) Le travail consiste alors à identifier des images sur des sites de vente en ligne et à les catégoriser dans un tableau, en faisant correspondre les catégories des sites et les catégories stylistiques.

Encadré n°16 : Créer un référentiel pour standardiser la reconnaissance des vêtements 2

A : en fait l'intelligence artificielle, le cerveau, on l'a constitué avec cette nana [de l'IFM](...)les dataset existants on les a utilisés avant, et (...) ça plafonne parce que tu as trop d'ambiguïté, la zone d'ambiguïté elle est large (...) en fait l'intelligence artificielle c'est un enfant, il a 2 professeurs, l'un lui dit ça c'est blanc , l'autre lui dit ça c'est noir , donc gars c'est la loterie quoi ... du coup en fait tu avais des niveaux de précision des algos qui étaient extrêmement faibles, du fait que le référentiel, le dataset qui était utilisé n'était pas robuste

Jean - CEO - ModeAI

La difficulté a été de co-construire une base de données de styles vestimentaires respectant des classifi-

cations en matière de mode et qui soient suffisamment homogènes pour être ensuite traitées sous forme d'annotation : *"l'étudiante était très bonne dans son référentiel stylistique, vraiment très bonne là-dedans, mais les [annotateurs] n'ont pas compris (...) la complexité et cette nécessité pour homogénéiser la donnée dans la mode (...)"* (Jean - CEO - ModeAI). Une fois des catégories de vêtements définies, le référentiel a été testé auprès d'utilisateurs de sites de e-commerce dédiés à la mode, puis il a été adapté pour que la catégorisation utilisée pour l'annotation synthétise la vision de l'IFM, la vision des consommateurs et la vision IA. Ce premier référentiel a été utilisé pour annoter une première base d'entraînement de test qui comporte alors 30000 images, avant de lancer l'annotation auprès de leur prestataire à Madagascar. Ils constatent *"que c'est le référentiel qui était invalide"*. La difficulté pointée par Jean est en fait de trouver le bon niveau de détail dans la classification des vêtements et limiter la part de subjectivités. Ainsi, ils souhaitent faire évoluer leur modèle vers une identification des vêtements par style (par exemple : rock, ou années 80), ce qui va nécessairement créer encore plus de besoins en annotation et plus de "zones d'ambiguïtés".

Encadré n°17 : La gestion des ambiguïtés 1

J : mon rêve ça aurait été de dire ça c'est un style rock, ça c'est un style urbain , mais on y est pas encore ! on a essayé, on a tourné tourné tourné, mais pour le moment on n'y est pas ... mais nous en tant que humain on n'est pas capable. On a des idées mais il y a tellement de zones [d'ambiguïtés] en fait sur ce qui est très clair oui, sur ce qui est très négatif oui, mais ce qui est entre les 2, personne ne sera forcément d'accord. Donc pour le moment c'est un sac de nœuds (...), nous notre idée c'est de mettre cette frontière-là plus fine possible. Et à un moment sur le style, sur les tendances, ces choses-là, la limite elle est trop large et on va se perdre, ça ne va rien apporter , ça va être plus déceptif ... donc il vaut mieux ne pas se lancer là-dedans pour le moment.

Jean - CEO - ModeAI

La gestion des ambiguïtés amène alors les deux entrepreneurs à s'interroger quant à la meilleure manière de mesurer le niveau de qualité des données annotées : *"quels indicateurs doit-on mettre en place sur une productivité, sur de la qualité qui est peut-être objective et sur de l'IA avec des volumes énormes (...) parce que même aussi demain notre modèle peut évoluer. Donc les modèles évoluent, donc ça veut dire que les attentes peuvent évoluer ... en fait c'est un mouvement perpétuel en fait"* (Jean - CEO - ModeAI). Ces évolutions du modèle ainsi que la prise en compte du jugement rendent, selon eux, la définition des indicateurs particulièrement épineuse. Dans leur pratique quotidienne de gestion des annotateurs, ils tiennent donc compte des zones d'incertitudes.

Encadré n°18 : La gestion des ambiguïtés 2

J : je pense que ... on a eu une tension là-dessus avec eux, pas une tension ... c'est le flou quoi. C'est que ils veulent que l'on donne une qualité, mais on va leur dire on peut vous donner plusieurs ... la qualité ce n'est pas un chiffre, c'est Il y a de la qualité qualitative, la gestion des ambiguïtés ; il y a la qualité quantitative au sens combien il y a d'erreur au global, et qu'est-ce que une erreur ? une erreur, pour nous, tu t'es trompé d'un skinny VS un slim parce que c'est border, pour nous ce n'est pas une erreur, même si lui il aurait mieux labellisé ... mais pour nous ce n'est pas réellement une erreur de qualité parce qu'il n'y a pas d'impact

Jean - CEO - ModeAI

Les entreprises d'intelligence artificielle ont donc tout intérêt à être attentives à l'organisation du travail des données, car c'est ce qui leur permet de réaliser les modèles les plus performants. Certaines entreprises d'IA sont parfaitement conscientes de ça, à tel point que ces dernières l'intègrent dans leur processus de production. Plus que sur le modèle, il faut donc agir sur les données, comme l'explique Guillaume : *"ça a été majeur pour nous et de dire on pense voir pourquoi la data peut créer ce genre de problème . Ça a marché et c'était un gain gigantesque d'un point de vue data. Évidemment une fois qu'on le dit, d'un point de vue statistique on aurait pu le voir dans le modèle, mais tu n'as pas forcément regardé ça."* (Guillaume - CEO LegalAI)

Très fréquemment, on a par ailleurs constaté à quel point le processus de production de l'IA est dynamique. Les changements dans la finalité du modèle, l'évolution des données à annoter sont des variables qui se retrouvent dans un processus de production associant étroitement les différents acteurs de la chaîne. Comme l'explique Guillaume, cela se traduit par des liens forts entre les modélisateurs et les producteurs de données : *"Quand tu essaies de résoudre un problème d'IA, il faut que ton équipe data voit les résultats du modèle et qu'elle puisse ajuster sa production de data par rapport à ce modèle."* (Guillaume - CEO LegalAI). Dès lors, toujours selon lui : *"idéalement, il faudrait que les gens qui produisent la data, ils voient les résultats du modèle avec l'équipe IA. Ce n'est pas deux mondes séparés comme on l'a fait sur certains domaines (...) l'équipe data, elle fait partie intégrante du truc. (...) Si j'avais juste l'équipe IA et que je leur disais Faites-moi des trucs , ça ne marcherait pas, parce que leur métier n'est pas non plus d'annoter à la main, (...), mais il y a un moment où il faut qu'il y ait une interaction avec une équipe data, que cette data soit versionnée, il faut pouvoir valider la qualité, ensuite tu déploies ton modèle dessus, il faut faire des tests de réversion, être sûr que ton modèle n'a pas baissé ses performances."* (Guillaume - CEO LegalAI)

Les exemples que nous avons cités jusqu'à présent montrent à quel point le travail de l'IA nécessite des allers et retours entre les différents acteurs de la chaîne, que ce soient les clients finaux, les entreprises qui produisent les modèles, ou celles qui s'occupent de la sous-traitance de l'annotation et leurs annotateurs.

Les décisions de modification concernant les modèles sont récurrentes et concernent alors l'ensemble de la chaîne, à tel point qu'il devient nécessaire de mettre en place des outils de coordination de cette dernière.

Cette dimension cyclique s'additionne à une chaîne de valeur dans laquelle le travail de l'IA est distribué ce qui contribue d'autant plus à faire évoluer en permanence la nature du modèle. Deux études de cas permettent de reconstruire les pratiques d'annotation tout au long de la chaîne de valeur. Le premier cas concerne un algorithme a été conçu et entraîné par la startup française Automatik, déjà citée dans notre thèse. Dans ce cas, le produit est développé pour une entreprise internationale de restauration industrielle (CanteenCorp). Cette dernière expérimente des terminaux de paiement automatisés censés gérer la détection des aliments, la facturation et les processus de paiement avec une intervention humaine minimale. À l'aide d'un algorithme de reconnaissance d'images, ce terminal identifie les aliments sur les plateaux et les trie pour les facturer au client. AIView est quant à elle une startup française d'intelligence artificielle qui s'est lancée en 2018 et emploie désormais une cinquantaine de personnes. L'entreprise ajoute un algorithme de vision par ordinateur aux caméras de surveillance des magasins déjà installées dans le but de détecter automatiquement les vols et de prévenir le vol à l'étalage. Les agents de sécurité, les caissiers ou les gérants de magasins sont alors alertés automatiquement via leurs téléphones. D'après sa communication d'AIView sur les réseaux sociaux, l'entreprise fournit près de 5 000 caméras à plus de 100 magasins en France, en Australie, en Belgique et au Royaume-Uni. Supermarchés et pharmacies font appel à la startup pour automatiser la vidéosurveillance. Plusieurs phases sont décrites sur le site de l'entreprise :

Comme dans les autres cas présentés, l'annotation des données est nécessaire pour construire la "vérité de terrain" (*ground truth*) nécessaire à l'entraînement du modèle est déléguée à une entreprise d'annotation de données basée à Madagascar. Dans le cas de CanteenCorp, le CEO d'Automatik déclare qu'il préfère "*garder la main sur les quasi-salariés, même s'ils ne sont pas techniquement des salariés*". Il précise aussi que, selon lui, les plateformes de travail numérique, telles qu'Amazon Mechanical Turk, diffèrent de leur propre modèle de sous-traitance directe, car les annotateurs malgaches et les employés de la startup française communiquent d'ailleurs fréquemment afin d'assurer une compréhension mutuelle du processus d'étiquetage. Les travailleurs des données dédiés à AIView sont également très intégrés et communiquent régulièrement avec la startup. Dans les deux cas, une personne au sein de la startup se charge de cette gestion opérationnelle, car la production de jeux de données se fait sur le temps long en intégrant de nouvelles phases d'annotation. Par exemple, la caisse de cantine automatique nécessite que les plats, couverts, assiettes et autres articles nouvellement introduits dans les restaurants soient annotés. D'autres problèmes découlent du comportement des consommateurs, comme lorsque ces derniers posent leurs téléphones sur les plateaux. Comme cela aurait pu nuire à la vie privée des utilisateurs finaux, la startup française a demandé aux annotateurs de les identifier afin de les flouter.

Les annotateurs de AIView reçoivent des séquences vidéo d'une durée de 6 à 8 secondes sur une plateforme

numérique. Parfois, ces séquences montrent des personnes en train de voler, d'endommager ou de déballer des produits, même si, la majorité des acheteurs agissent bien entendu de façon plus inoffensive. Dans ce cas, annotateurs doivent également s'adapter aux évolutions du modèle. Cela implique par exemple de s'adapter aux préférences des clients finaux, ainsi que d'incorporer dans leurs étiquettes des types de vols inhabituels ou des comportements suspects. Les responsables RH de l'entreprise d'annotation malgache décrivent par exemple comment ils identifient et traitent les gestes de vol non anticipés. Ces cas complexes sont d'abord discutés en interne, soit entre les annotateurs malgaches eux-mêmes, soit entre eux et leurs chefs d'équipe à Madagascar.

Encadré n°19 : Quand la vérité terrain fait évoluer les techniques d'annotation

R : Nous communiquons aussi directement avec le client (...). Je dirais tous les quinze jours, parce que (...), il y a toujours des mises à jour sur la plateforme [d'annotation]. De notre côté, nous faisons de notre mieux pour participer, et pas seulement en tant qu'exécutants. En ce qui concerne les noms d'étiquettes, il y a des acheteurs (...) qui cachent les produits dans leurs vêtements, il n'y a donc pas d'étiquette pour eux.

Q : Et c'est un agent qui l'a fait remarquer au chef d'équipe qui en a parlé au client...

R : Oui, c'est exact. Les techniques de vol à l'étalage évoluent en effet. Nous essayons de les détecter et de les signaler. Le client met alors à jour la [plate-forme d'annotation] et crée des étiquettes pour eux.

Jean - Conseiller RH - Sous-traitant A

L'adaptation permanente du modèle aux changements du phénomène perçu est ici cruciale. Comme les voleurs à l'étalage s'adaptent pour éviter d'être détectés, les annotateurs (et, en fin de compte, le modèle) doivent être en mesure de détecter des vols en constante évolution. Pour ce faire, ils doivent constamment ajuster les outils et les définitions des étiquettes : quels sont les substituts des phénomènes qu'ils déduisent ? Une personne qui met discrètement un objet rectangulaire dans sa poche tente-t-elle de dérober du chocolat ou simplement de ranger son smartphone ? Et comment rendre compte de ces incertitudes ? Faut-il ajouter une nouvelle étiquette ("smartphone") ou simplement l'étiqueter comme "non suspect" ?

Un autre travailleur des données explique que les annotateurs doivent détecter les vols en moins d'une minute et qu'ils sont répartis en plusieurs équipes afin de couvrir les nuits et les week-ends. Ces informations nous ont permis de conclure que ce que les annotateurs visionnent est un flux vidéo en temps réel. Ils agissent principalement en tant qu'agents de sécurité à distance, détectant les vols presque en temps réel ⁸⁹,

89. En fait, quelques secondes s'écoulent entre le moment où la vidéo est enregistrée et le moment où les travailleurs à Madagascar la reçoivent.

ce qui vient confirmer la typologie du travail sur les données dans Tubaro, Casilli, et Coville (2020b), ces annotateurs imitent en fait l'algorithme.

Certaines vidéos ne sont toutefois pas visionnées en temps réel⁹⁰. Ainsi, lorsqu'ils n'annotent pas les images diffusées en direct, les travailleurs de l'information étiquettent les vidéos qui ne sont pas en direct et qui sont stockées sur un serveur en vue d'un traitement asynchrone. Une fois que les annotateurs ont fini d'étiqueter un lot de ces vidéos, ils doivent également revoir les annotations des autres travailleurs en les vérifiant. À la lumière de ces éléments, nous pensons que ce système ne peut pas interpréter automatiquement leurs gestes comme du vol à l'étalage ou comme un comportement normal de consommateur. C'est toute la différence entre l'IA commerciale et non commerciale : le succès du service dépend de la qualité du modèle. Quelle est l'efficacité de l'interaction avec le système "automatisé" dans un environnement de production ? Le succès d'AIView et de son contractant dépend du fait que l'assemblage d'annotations humaines et de modèles de vision par ordinateur alerte les employés du magasin sur les comportements suspects.

En outre, le contrôle de la qualité de l'ensemble des données est réparti tout au long de la chaîne de production. La performance du modèle dépendant fortement de la qualité de l'ensemble de données, les images étiquetées sont vérifiées manuellement par des annotateurs-vérificateurs, puis techniquement par la startup française spécialisée dans l'IA. Les détails et les délais du projet influencent fortement le processus de vérification. Dans les deux cas, l'intervention humaine n'est théoriquement pas nécessaire sauf si le modèle n'est pas en mesure d'aboutir à une conclusion fiable. Dans le cas de CanteenCorp, le système fait d'abord appel au consommateur et, en cas d'échec, à un employé du restaurant, pour scanner les aliments et les étiqueter selon des catégories prédéfinies. Dans le cas de la vidéosurveillance automatique, les employés des magasins peuvent communiquer via l'outil AIView, soit pour confirmer un vol, soit pour signaler des problèmes de "détection automatique". En cas de doute sur un vol, il existe un chat de groupe où "[le personnel du supermarché] peut demander des enregistrements vidéo à la startup". Les faux positifs peuvent perturber le travail des employés du magasin. Le directeur du supermarché a également indiqué avoir soulevé le problème suivant avec la startup : *Nous leur avons demandé de supprimer les alertes pour les sacs à provisions, car cela provoquait trop d'alertes. (...) Oui, nous discutons avec la startup.*

Ce cas illustre la question suivante : qui s'occupe de l'algorithme lorsque celui-ci déraile ? À quel point est-il important pour l'entreprise de prévenir les faux positifs et de s'assurer que les vrais positifs ne sont pas manqués ? La réponse à ces questions détermine l'organisation que prendra la chaîne de production. Automatik compte sur le consommateur pour s'assurer que l'algorithme fonctionne correctement : *"Les clients d'Automatik interagissent avec notre système. En cas de désaccord [avec le résultat de l'algorithme], ils*

90. *"Les alertes sont des vidéos en temps réel et l'analyse consiste à rechercher les vols et à les étiqueter, s'il s'agit d'un vol ou s'il n'y a rien à signaler. Il y a aussi d'autres tâches dans notre équipe (...), notamment les annotations [des vidéos] que nous accumulons sur un serveur. Lorsque nous étiquetons ces vidéos, nous évaluons nos performances".* (Data annotator)

peuvent le corriger”. (Audrey - Solutions architect - Automatik)

Trois points principaux ressortent de cette étude de cas. Tout d’abord, le processus d’annotation nécessite une mise à jour et des modifications constantes. On constate également la division internationale du travail de problématisation du modèle entre les ingénieurs et managers français de l’IA et le sous-traitant malgache. Enfin, les consommateurs et les employés de la cantine évaluent la performance des SIA en dernier recours. Par conséquent, le système d’IA relie les décisions des travailleurs de données aux actions du personnel des supermarchés. Comme l’explique un annotateur de données, les décisions prises par les travailleurs des données malgaches influencent le comportement des employés des grands magasins dans d’autres pays : “Si un enfant casse une bouteille, je le qualifie de suspect et les employés du supermarché reçoivent une alerte et agissent en conséquence”. Selon le magasin, les alertes sont reçues soit par le directeur, soit par les caissiers, soit par la sécurité. Les employés du magasin doivent recevoir les alertes dans un délai ce qui leur permet d’agir, ce qui incite particulièrement les annotateurs à agir rapidement : “Je reçois des alertes toutes les minutes au premier étage du supermarché et toutes les trente secondes au rez-de-chaussée du magasin.” (Directeur de supermarché).

Ces deux exemples nous permettent d’établir deux résultats très importants à ce stade de notre travail, car ils permettent de réarticuler le travail de l’IA dans l’espace et dans le temps en montrant bien deux aspects essentiels de la chaîne de valeur de l’intelligence artificielle : il s’agit d’un processus *incrémental* et *distribué*. Tout d’abord, le processus de développement des modèles d’IA est *incrémental*. Nous confirmons ainsi le propos de Seaver (2017) qui parle de modèles “non stables”. Il sera systématiquement nécessaire d’apporter des corrections tant aux consignes d’annotations qu’aux modèles, notamment dans lors du déploiement de l’outil en production. L’adaptation de modèles pré entraînés aux réalités du terrain nécessite alors des phases de ré-entraînement qui seront en partie prises en charge par les annotateurs de Madagascar (“*fine tuning*”) ⁹¹. Ensuite, le processus de développement de l’IA est *distribué*, c’est-à-dire qu’il implique différents acteurs à différents niveaux de la chaîne de valeur. Ce point nous rappelle les constats de (Burrell, 2016) à propos de l’opacité des algorithmes du fait de l’intégration de nombreux composants techniques (briques informatiques) et organisationnels (services d’une entreprise). Notre travail apporte ainsi des éléments supplémentaires, quant à la compréhension de l’opacité algorithmique de type “organisationnelle”, en soulignant la fragmentation de la chaîne de valeur dans l’espace et dans le temps.

Au-delà de cette dimension, notre analyse illustre la part essentielle des “couches” humaines dans le processus de production des modèles d’IA. L’implication de ce fonctionnement de l’IA par le travail distribué soulève des questionnements quant à la capacité de réguler le processus de production de l’IA. Pour les en-

91. Le *fine tuning* ou “affinage du modèle” désigne “une approche de l’apprentissage par transfert dans laquelle les poids d’un modèle pré-entraîné sont formés sur de nouvelles données. Le réglage fin peut être effectué sur l’ensemble du réseau neuronal ou sur un sous-ensemble de ses couches seulement, auquel cas les couches qui ne font pas l’objet d’un réglage fin sont “gelées” (Wikipédia)

treprises d'IA, comment dès lors contrôler chaque étape de la production ? Et par conséquent, la puissance publique peut-elle outrepasser cette opacité organisationnelle pour vérifier la conformité d'un SIA ?

Cette seconde section nous a d'abord permis d'analyser le processus de formalisation du choix d'un mode d'externalisation. Plus qu'une question de coût, on a pu constater que cela répondait à un objectif d'optimisation du temps de travail des développeurs et *data scientists* des entreprises d'IA. Toutefois, cette phase de la production est tout de même perçue comme suffisamment essentielle pour justifier le recours à des entreprises formelles (BPO) et donc à une force de travail stable. Ainsi, les résultats des modèles sont très sensibles à la qualité des données annotées, ce qui explique l'existence d'un discours centré sur la nécessité de maintenir la qualité de la production de données. Ce choix résulte donc d'une tension entre une activité perçue à la fois comme périphérique au métier de *data scientist*, mais essentiel au processus de production de l'IA. Dès lors, comment expliquer cette tension ? Pour nous, la dimension "qualité" s'analyse en effet au regard du processus de développement de l'IA.

* * *

Dans cette section, nous avons analysé la formalisation des processus d'annotations dans les entreprises d'IA françaises. Un premier élément concerne le caractère *dynamique* de cette activité qui implique de réactualiser en permanence la "vérité terrain" ou "*ground truth*" (Jaton, 2019) ce qui justifie d'intégrer les annotateurs. La nécessité de coller le plus possible au processus que les entreprises souhaitent automatiser crée ainsi le besoin d'assurer une proximité entre les modèles et les données.

Nous avons également constaté l'attention forte portée à la qualité des données pour éviter de créer des biais. Durant ce travail de catégorisation ("*categorical work*") (Bowker & Star, 2000), l'enjeu est aussi de maîtriser la subjectivité lors d'un processus qui s'apparente à la traduction du contexte du client final (utilisateurs de l'IA) dans un outil d'automatisation. De fait, ces groupes "*de pratiques*" au sens de (Bowker & Star, 2000) (p.297) travaillent entre elles au "*maintien de la cohérence*". Le client final possède la connaissance métier, l'entreprise d'IA celle de la modélisation et les annotateurs, une expertise intermédiaire, puisqu'ils développent à la fois une connaissance du projet d'annotation (les données à catégoriser) et des enjeux de l'annotation (modélisation). On constate alors que l'organisation du travail choisie par les entreprises d'IA françaises a pour objectif d'aboutir à un *couplage entre les données et les modèles* ce qui passe par un *couplage entre les développeurs et les annotateurs*.

Nos résultats suggèrent que le marché des données annotées pour l'intelligence artificielle se structure à la fois autour du prix, mais également autour des jugements sur la qualité des données. Ce facteur, bien trop déterminant dans la forme finale du modèle constitue une justification du recours par les entreprises d'IA au modèle BPO. Nos résultats montrent que cette dimension de l'agencement du secteur est à la fois poussée par des impératifs de résultat (*qualité de la prédiction finale*), ainsi que par des enjeux liés à l'éthique

de l'IA et aux menaces de régulation concernant les biais des SIA (*éthique de la qualité*).

Comme nous le verrons d'ailleurs dans notre partie trois, cet agencement du marché de l'IA détermine largement le recours à des travailleurs humains pour perfectionner et compléter les modèles (Tubaro, Casilli, & Coville, 2020b). La criticité et les imperfections des modèles conduisent alors les entreprises qui mettent en place des SIA à recourir à davantage de travail humain, soit pour la production de données, soit pour délivrer le service à la place du modèle. Contrairement à une partie de la littérature en sociologie économique qui étudie les jugements sur la qualité à travers des "professions" (par exemple : les avocats) (Karpik, 1989), nos résultats concernent des travailleurs censés être peu qualifiés. Comment dès lors analyser ce résultat ? Comme nous le verrons dans notre prochaine partie, ces dernières vendent un processus socialisé et collectif de travail des données qui passe la mise à disposition d'une force de travail qualifiée et directement intégré dans le processus de production des entreprises d'IA. Pour les entreprises d'IA, la difficulté consiste ainsi à créer des "standards" qui garantissent un travail de qualité (Sturgeon, 2002). Cette dimension est très présente dans les théories des chaînes de valeur au sein desquelles on retrouve des organisations du travail qui impliquent des interactions fréquentes entre les firmes à travers la forme "hiérarchique" et l'échange de "connaissances tacites" (Gereffi et al., 2005). Les entreprises d'annotations étudiées mettent d'ailleurs en avant leur capacité à maîtriser le processus d'annotation, en comparaison à des "marchés purs" comme Mturk sur lesquels on retrouve des travailleurs indépendants. La capacité de ces entreprises à créer des connaissances sur les manières d'annoter (procédés) ou sur la variété des données va alors justifier ces contacts fréquents. C'est finalement dans la nécessité de coupler connaissance des données (annotateurs) et connaissance de la modélisation (développeurs) que réside la logique d'intégration logistique qui est ici à l'oeuvre.

Par ailleurs, l'analyse de bout en bout de deux systèmes d'IA nous permet de constater que le processus de développement de l'IA est très *incrémental*, notamment du fait de la non-stabilité des modèles d'IA et des procédés de développement (Seaver, 2017). Dès lors, si la relation avec les annotateurs de données s'envisage sur le long terme, c'est aussi qu'elle permet une intégration des travailleurs dans le processus de production, à même de stabiliser les modèles. Comment alors se déroule cette intégration ? Comment analyser le rôle des annotateurs au regard du processus de production des données ? Selon nous, cette intégration permet de mettre en place des réseaux de communication entre travailleurs qui facilitent le travail de l'intelligence artificielle. L'objectif de notre prochaine partie est justement d'approfondir cette question.

4.2 La mise en place de processus de communication entre annotateurs et employés des startups françaises pour gérer la qualité des données

Dans la partie précédente, nous avons constaté deux aspects importants dans les choix des startups françaises en matière d'externalisation du travail des données. Le processus de production des modèles d'IA est cyclique, nécessitant des phases d'aller et de retour entre les équipes de développement et d'annotation. Par ailleurs, il nécessite de disposer de données de qualité ce qui favorise une organisation du travail permettant un *couplage données-modèles*.

Afin d'étudier l'organisation du processus d'annotation par les entreprises d'IA françaises, il nous semble pertinent de revenir sur l'hypothèse de l'existence de firmes "hybrides", autrement qualifiées de firme-réseaux par les économistes de la firme. En effet, nous verrons que les processus de production rencontrés lors de cette thèse se distinguent des modèles des plateformes comme marché tels que Mturk. Ainsi, derrière ce mode fonctionnement basé sur la qualité, nous pensons que l'organisation de ces chaînes de valeur correspond bien à la thèse de la "*hiérarchie étendue*" qui considère la firme-réseau comme une "*organisation intégrée*" (Fréry, 1997) qui : "*introduirait (...) une forme de hiérarchie explicite avec ce que cela suppose de supervision directe, de subordination, d'inégalité entre les partenaires et de centralisation des décisions stratégiques.*" (p.107). Dans ce cas de figure, une firme A exerce une subordination sur les salariés d'un sous-traitant B du fait de plusieurs modes d'intégration qui sont indépendants de la propriété des actifs. *L'intégration culturelle* d'abord, qui passe par des relations amicales ou familiales entre la firme-pivot et les partenaires. *L'intégration médiatique* ensuite, passe par la renommée de la marque de la firme-pivot, comme c'est le cas pour des marques appartenant à de grandes multinationales comme Nike. Ce mode d'intégration permettrait d'obtenir une plus grande loyauté des sous-traitants. Enfin, l'intégration logistique passe par la mise en place d'une infrastructure technologique pour contrôler les firmes du réseau, notamment à l'aide de codes-barres et d'outils de remontée de données sur la production.

Comme l'explique [Baudry \(2003\)](#), cette "*intégration logistique déboucherait même (...) sur de véritables hiérarchies électroniques au sein desquelles la firme-pivot détiendrait un pouvoir sur ses fournisseurs, pouvoir comparable à celui qui serait obtenu par la propriété des actifs non humains des firmes membres du réseau.*" (p. 108) La firme-pivot est alors en mesure de cumuler les fonctions de conception (orientation stratégique), de coordination (minimiser les coûts de transaction) et de contrôle. En insistant sur le pouvoir de contrôle des firmes-pivots sur leurs sous-traitants, cette théorie fournit un cadre pour analyser les rapports de pouvoirs qui s'exercent dans le cadre des relations inter-firmes, mais elle contribue aussi, "*en assimilant les relations interfirmes aux relations intrafirme, à dissoudre en partie les frontières légales de la firme*". Notre hypothèse rejoint alors l'idée que les firmes d'IA mettent en place des hiérarchies électroniques étendues afin de mettre en place des mécanismes "*d'apprentissages*" rendues nécessaires par le caractère incrémental des modèles. L'apprentissage-machine serait donc avant tout basé sur une chaîne d'apprentissages-humains.

Sur ce point, les théories des ressources et la théorie évolutionniste postulent une efficacité propre de la firme par rapport au marché, qui réside dans sa capacité à créer des ressources en mettant en place une organisation du travail qui stimule le capital humain et donc les processus d'apprentissage et d'acquisition de compétences. Ce courant présente la firme comme un "lieu d'apprentissage" à travers trois niveaux d'études. D'abord, le niveau "routinier" (*learning by doing*) par lequel les individus améliorent leurs compétences en répétant une action, et ce, tout en restant dans le cadre établi pour la réalisation de la tâche. Le deuxième niveau correspond à réalisation d'expériences durant la production, qui en modifie le cadre de référence. Cette approche pose la question de la codification des connaissances produites durant l'apprentissage. Ici, Michaël Polanyi distingue les connaissances codifiées des connaissances tacites, qui sont : *"non exprimables hors de l'action de celui qui les détient, et le fait de les posséder est même ignoré ou négligé par leur détenteur. Par là même, la connaissance tacite est un bien qui se prête difficilement à de nombreuses opérations : leur diffusion et leur apprentissage sont coûteux et difficiles à mettre en œuvre; leur stockage et leur mémorisation sont conditionnés par le renouvellement – de génération en génération – des personnes qui détiennent ces connaissances; elles sont difficilement répertoriables et classables.* Baudry (2003) p.33). La codification passe donc par la transformation de ces connaissances en "messages" qui pourront être délivrés, par exemple à un autre salarié via une formation.

Un troisième niveau est celui du développement de nouvelles routines organisationnelles sur la base des apprentissages. Par conséquent, la firme n'est *"plus une forme par défaut, elle possède une efficacité qui lui est propre et qui lui permet de réaliser des opérations inaccessibles à des individus isolés.* Baudry (2003) p.33) Par opposition au marché, elle est un *"moyen d'accroître la capacité de traitement de l'information"* et facilite la coordination en réduisant les conflits. Enfin, elles *"favorisent un comportement adaptatif des individus – donc de la firme –, grâce à l'établissement de règles communes, d'un langage propre et d'objectifs partagés."* La firme est également vue comme un *"panier de compétences"*, car les capacités cognitives limitées des membres de la firme obligent les individus à apprendre. Or, la complexité et le coût de cet apprentissage nécessitent de se concentrer uniquement sur une ou deux compétences centrales (espace au cœur des firmes).

Le principal intérêt de l'approche fondée sur les compétences est de reconnaître la firme comme un espace de développement et de codification des compétences. Dans ce cadre, l'efficacité de la firme comme structure de gouvernance réside dans sa capacité à stimuler, à mobiliser et à inscrire ces compétences pour qu'elles soient transmises parmi les salariés. Ici, l'objet de la firme n'est pas seulement disciplinaire, c'est-à-dire que la finalité n'est pas seulement de contrôler les actions des travailleurs, mais aussi de stimuler ses capacités d'apprentissage et *in fine* d'innovation dans la réalisation de leurs tâches.

Cet état des lieux critique sur les théories des chaînes de valeur et des firmes nous permet de mieux positionner nos hypothèses, au regard du travail d'enquête déjà réalisée avant cette thèse sur le micro-travail en France (A. Casilli et al., 2019a). Dans le contexte, d'une plateforme numérique de travail, nous avons

pu constater que, malgré un travail de codification, la réalisation des tâches passe par un investissement de deux autres acteurs que la firme donneuse d'ordre : la plateforme et le travailleur. Nous avons ainsi pu observer comment les salariés de la plateforme Foule Factory communiquaient avec les micro-travailleurs pour les conseiller sur la réalisation d'une tâche. La plateforme a également mis en place des certifications qui définissent les bonnes manières de réaliser une tâche donnée et qui permettent aux microtravailleurs d'accéder aux tâches réelles.

De manière moins directe, l'interface de Foule Factory, les possibilités de design ou bien l'API qui permet de coder les tâches et de récupérer les données constituent autant de manières de codifier les tâches pour la plateforme. De leur côté, il arrive que les travailleurs créent des tutoriels ou publient des conseils pour réaliser certaines tâches. Par ailleurs, la littérature sur les plateformes et plus spécifiquement sur les plateformes de micro-travail a largement souligné les coûts d'attentes induits par le fonctionnement des plateformes (A. Casilli et al., 2019a). Sur les plateformes de micro-travail, certains travailleurs arpentent le forum en rafraîchissant la page de sélection des tâches, tandis que sur les plateformes de livraison, certains livreurs restent dans la rue à certains endroits clés pour attendre les commandes. De ce point de vue, la plateforme en tant que marché n'optimise pas totalement l'offre et la demande : une partie des coûts de coordination repose sur les travailleurs. En ce sens, la plateforme opère plus comme une firme qui coordonne un réseau "virtuel" de travailleurs qui pourraient tout à fait être des salariés, plutôt que comme une place de marché "pure".

Cette approche a une double conséquence sur la question qui nous occupe. La structure des entreprises d'annotations peut être influencée par la nécessité de mettre en place ces mécanismes. On peut en effet s'interroger sur la capacité des PNT de type marché pur comme Amazon Mechanical Turk, à mettre en place ces mécanismes d'apprentissage et à en tirer des bénéfices. Ensuite, cela questionne également le modèle de travail des données basé sur une sous-traitance classique. On peut alors se questionner sur la réalité d'un travail du clic, complètement déqualifié et donc décorrélé de mécanismes d'apprentissages. Comme on le verra dans notre prochaine partie, la notion de compétence et de qualification du travail sont deux notions sujettes à controverse. S'agit-il de savoirs individuels permettant de réaliser la tâche, de savoir-tacites construits collectivement et codifiés par la suite dans des documents? Par ailleurs, comment se matérialise l'apprentissage au sein des entreprises d'annotations? Enfin, comment tenir compte des relations de pouvoir au sein de ces processus d'apprentissage collectif. Aussi, dans quelles mesures ces mécanismes d'apprentissage sont-ils imposés?

Dans cette section, nous nous concentrons d'abord sur la manière dont les entreprises d'IA mettent en place cette intégration logistique pour arriver à remplir ces objectifs de qualité des données. Au-delà de la complexité liée au projet ou à l'utilisation de certains outils d'annotations, la capacité du client à répondre aux questions des annotateurs et à mettre en place un processus de communication est aussi un élément es-

sentiel. Cet aspect du travail des données est nécessaire pour transmettre de manière efficace les consignes d'annotations et les faire évoluer si besoin. C'est la manière de gérer la relation de sous-traitance que nous allons étudier dans cette section en montrant que c'est par cette relation de proximité que les entreprises d'IA peuvent produire des modèles plus performants. Nous analyserons ainsi deux aspects de l'intégration logistique : la mise en place des outils d'annotation et des consignes et la mise en place des outils de communication qui facilitent les échanges entre le donneur d'ordres et les travailleurs des données.

4.2.1 Des outils d'arrangements pour la standardisation et la stabilisation des concepts de l'intelligence artificielle

Les mécanismes d'intégration logistique des travailleurs des données passent d'abord les outils d'annotation et constituent notre première thématique d'analyse : que permettent-ils de faire ? Comment participent-ils à un mode de fonctionnement proche d'une "hiérarchie électronique" ? Nous évoquons ensuite également la documentation qui formalise les consignes d'annotations en nous demandant comment les entreprises d'IA formalisent-elles ces instructions alors qu'elles sont susceptibles d'évoluer au fur et à mesure des projets ?

En s'arrêtant sur les différents problèmes qui peuvent perturber les projets d'IA, les dirigeants des entreprises de sous-traitance évoquent deux types de problèmes : l'outil d'annotation utilisé par le client final ne fonctionne pas correctement ou est difficile à utiliser et les consignes d'annotation données par le client ne sont pas suffisamment claires⁹², et ce, d'autant plus lorsque ces dernières sont modifiées en cours de projet ou qu'elles ne sont pas suffisamment précises.

Concernant les outils d'annotations, commençons par deux cas de figure : celui de AIMedic et d'Automatik. Ces deux exemples sont particulièrement pertinents, car ces entreprises ont choisi de développer leurs propres outils. Cet aspect du processus de production de l'IA peut surtout s'observer chez les entreprises que nous qualifions d'*entreprises spécialisées* (Table ??), car elles développent un seul produit basé sur l'IA, ainsi que chez certaines *agences d'IA* qui sont des firmes qui développent plusieurs projets d'IA selon leurs différents clients. Dans les deux cas, ce sont plutôt des entreprises de tailles modestes (moins de 100 salariés) et pour lesquelles l'IA est au coeur de la proposition de valeur, contrairement à de plus grandes entreprises ou à des services publics non spécialisés dans cette technologie. Ces dernières passent plutôt par des prestataires qui vendent l'accès à des outils d'annotations. Dans l'encadré qui suit, Cédric, en charge de la gestion des jeux de données chez AIMedic explique que la partie annotation a été développée en même temps que leur produit de détection des ECG. En effet, du fait des spécificités des données traitées -des vidéos d'ECG- il était nécessaire d'avoir une plateforme qui permette de gérer ces formats de données spécifiques au monde

92. "Donc dans quelle mesure les consignes du projet vont être clarifiées et ça va dérouler ou dans quelle mesure on est un peu bloqué dû à des vrais problèmes liés à des soucis d'outil, des soucis de consignes (parfois le client n'est pas très clair et très exigeant) ? Il y a cette notion : dans quelle mesure on arrive assez vite à un rythme de croisière qui fait que c'est confortable pour moi et confortable pour mes chefs d'équipe."(Pierre - CEO - Sous-traitant G)

médical.

Encadré n°20 : Réaliser des tâches sur la version relationnelle de Mturk

Il y a deux autres groupes. Le premier c'est notre plateforme d'annotation interne (...) qui à la base était un dérivé de notre produit et qui (...) permet [maintenant] de réaliser les annotations tout simplement. (...) C'est ce qu'on appelle PLATAN, plateforme d'annotation (...) et Data pipeline, c'est un peu le dernier groupe de missions : une fois que l'on a récolté des données annotées, il y a tout un tas de choses que l'on va faire avec. Par exemple, simplifier les évaluations. Actuellement, souvent, on a autant d'évaluations que l'on a de problèmes spécifiques à résoudre. On se rend compte que l'on n'est pas très bon pour détecter tel truc dans tel cas.

Cédric - Lead data - AIMedic

L'outil d'Automatik a quant à lui été créé au tout début de l'existence de l'entreprise, autour des années 2010, à un moment où *"il n'y avait pas de software d'annotation"* (Augustin Marty - Automatik. La typologie des activités de production de l'IA présentée dans notre chapitre 5.1 illustre d'ailleurs l'émergence d'entreprises dont la proposition de valeur principale concerne la vente d'outils d'annotations, et dans certaines l'accès à une force de travail. Dans le cas d'Automatik comme dans celui de AIMedic, ces outils sont utilisés par une personne en interne, chargée de gérer le processus d'annotation et dans certains cas de faire des retours aux travailleurs⁹³. Cette capture d'écran permet de mieux comprendre comment fonctionne un outil d'annotation. On peut y voir à gauche les différents éléments à détecter sur l'image (les classes) et à droite les différentes techniques d'annotations (encadrement, détournage...). Un onglet message permet d'envoyer des messages directement, dans les cas où l'annotation n'est pas possible ou les travailleurs ont un problème.

93. *"[On a un outil] pour être capables de surveiller et revoir les datasets qui étaient annotés (...) On voyait les résultats arriver en temps réel. Donc on sait qui a fait quoi, ce qui est important. On est capable de dire qui a produit quel travail. Et ensuite, on est capable de dire attention, Max, là on avait dit qu'on s'arrêtait aux camionnettes, or là tu as pris un camion. (Théophile - CEO - Automatik)*

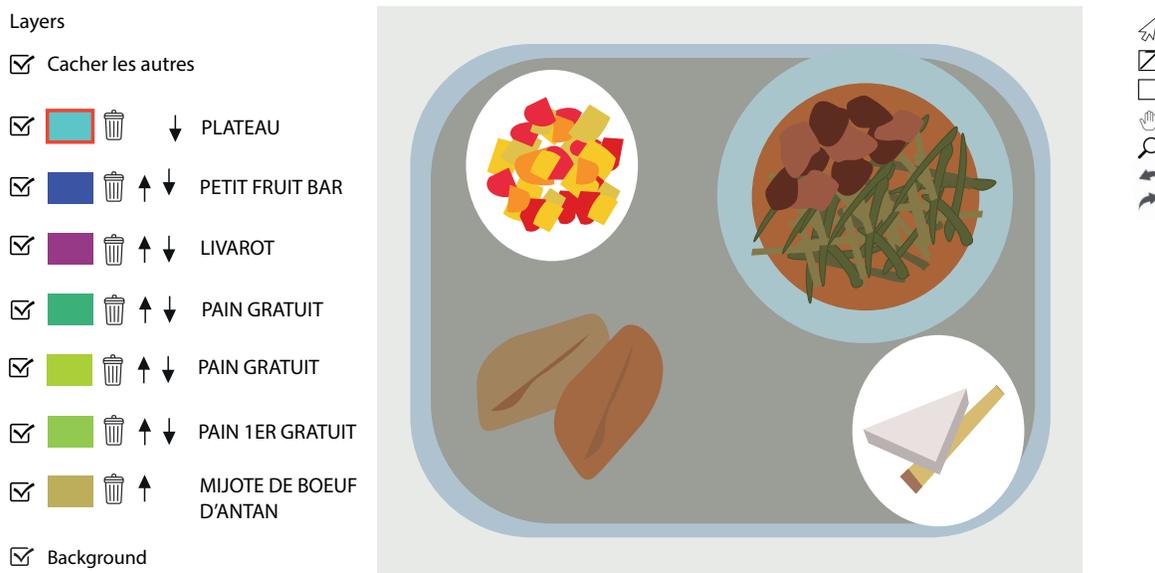


FIGURE 11 – Capture d'écran du logiciel d'annotation de la caisse de cantine développé par Automatik

L'image ci-dessous est une capture d'écran réalisée lors d'une démonstration de l'outil de gestion des annotations par Lin, en charge de cette activité chez Automatik, lors de notre entretien. Il s'agit d'un outil de gestion qui complète l'outil d'annotation en lui-même en permettant de suivre en temps réel le travail réalisé par les annotateurs. En plus de cet outil, elle, ainsi que les développeurs des modèles, échangent avec les travailleurs des données par le biais d'une messagerie d'entreprise (un Slack). Cet aspect de l'annotation se retrouve dans la majorité des entreprises. Victor du sous-traitant expliquant notamment qu'en plus de l'outil d'annotation, il est nécessaire d'intégrer des fonctions de gestion de projet ⁹⁴

94. "Souvent, ce qui est oublié c'est de dire J'ai un outil qui me permet de faire de l'annotation, mais comment est-ce que je peux suivre l'avancement du projet ? Comment est-ce que les personnes qui annotent peuvent accéder à de la documentation support ? (...)" (Victor - CTO - Sous-traitant B)

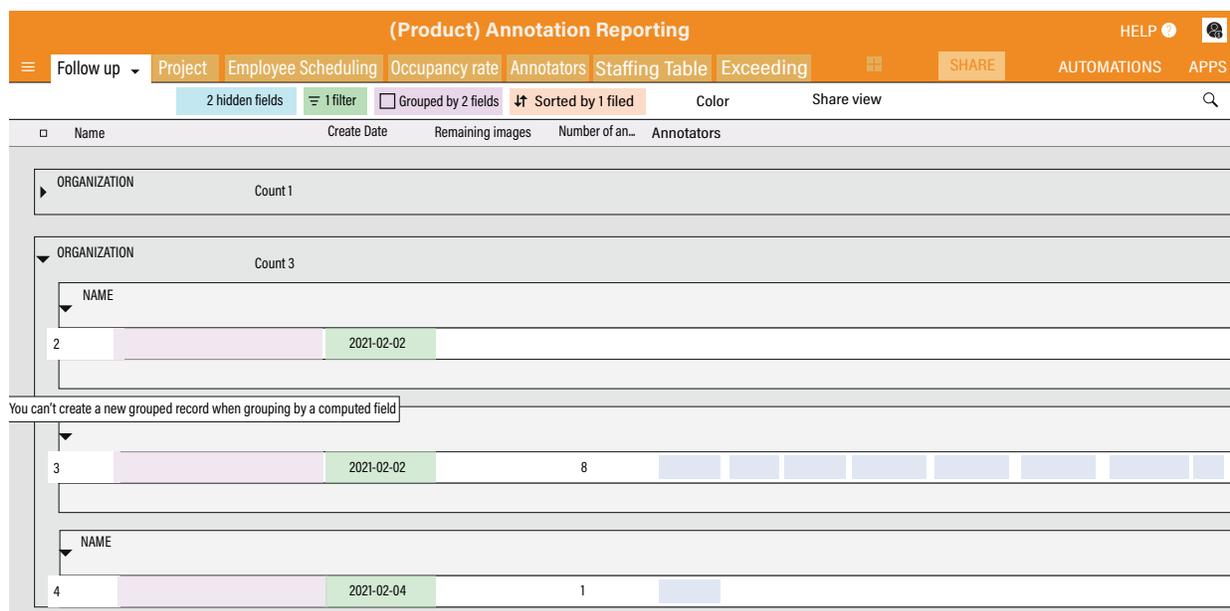


FIGURE 12 – Capture d’écran du logiciel de gestion de l’annotation de l’entreprise Automatik

Toujours sur les outils d’annotations, Pierre, le CEO du sous-traitant G indique qu’il s’agit d’un critère important dans le déroulement des projets d’annotations. Lorsque les clients n’ont pas d’outil, ce dernier fait appel à une entreprise qui propose un service clé en main d’outil d’annotation⁹⁵. Frédéric du sous-traitant F met également en avant cet aspect technique à travers son rapport aux clients finaux. Selon lui, il est rare de tomber sur des clients qui n’ont pas déjà un outil, cependant sa clientèle est plutôt anglophone et composée de grands groupes, ce qui explique ce constat⁹⁶. Dans le cas de BankAI, l’entretien réalisé avec Mathilde montre bien qu’il a fallu tenir compte de la faisabilité technique du projet, et de la capacité du sous-traitant à former ses travailleurs à l’outil. Pour ces deux aspects, Mathilde souligne le travail d’accompagnement de Pierre, également interrogé durant notre enquête, qui, avant de fonder son entreprise d’annotation, a été diplômé de Polytechnique et a débuté sa carrière comme *data scientist*, ce qui constitue un autre apport de ce sous-traitant en termes de compétences⁹⁷.

95. "Donc typiquement mettre en place sur l’outil, montrer au client l’outil, comment ça se ferait, puis faire des premiers tests. Donc, déjà comprendre ce que veut le client. Si c’est l’outil du client, ça va être très rapide parce que ça évite de faire des tests. Si ce n’est pas l’outil du client, il y a une phase un peu de R&D (j’exagère peut-être un peu), en tout cas de mise en place de l’outil"(Pierre - CEO - Sous-traitant G)

96. "Donc on pose des questions [aux clients] : Quel volume ? Quel type d’annotation (image/bounding box [Cadre de contour], vidéo, texte ou audio) ? Avez-vous un logiciel d’annotation ? Ce qui est généralement le cas. Il y a des clients qui n’ont aucun logiciel, mais c’est hyper rare. Ensuite on va calculer le temps de productivité. Soit c’est du process très basique et on a déjà tous les temps en tête (on sait combien de temps on met à faire un bounding box en moyenne) (...)" (Frédéric - CEO - Sous-traitant F)

97. "R : Il est adorable ! Il m’a tellement aidé. On a bien parlé et c’était très agréable (...) Je lui ai expliqué ce dont on avait besoin. J’avais des exigences (...). Donc je lui ai expliqué ce dont on avait besoin. Je lui ai montré le logiciel. Ils ont formé leurs gens sur le logiciel. Il faut quand même le faire ! On l’a installé nous-mêmes, parce que c’était un peu particulier. On l’a installé sur nos serveurs et on leur a donné accès au serveur. Le serveur ne marchait pas. Il y avait des galères comme ça. Et ils ont toujours été hyper réactifs.

Une fois l'outil installé, il faut également définir des procédés d'annotation en formalisant des consignes à destination des travailleurs (*guide*). Ces dernières servent en général à indiquer les exemples sur lesquels les annotateurs sont susceptibles de tomber et à expliquer les concepts propres au projet d'annotation. Par exemple, sur un projet de traitement automatique du langage dont l'objectif est de détecter les éléments importants d'un document juridique, il s'agit de comprendre et d'identifier ce qu'est un titre de niveau un ou le nom d'une entreprise etc. De manière concrète, il peut s'agir d'un document PDF ou d'un élément directement présent dans un outil d'annotation et qui peut faire jusqu'à soixante pages dans les cas d'annotation les plus complexes, notamment lorsque la classification des données comporte des ambiguïtés⁹⁸.

La constitution du guide d'annotation fait également l'objet d'un travail de l'entreprise sous-traitante. Dans le cas de BankAI, il fallait transformer ce que Mathilde avait défini dans le guide, afin que ce soit parfaitement compréhensible par les travailleurs. Par ailleurs, le sous-traitant G a mis en place un fichier permettant aux travailleurs de poser leurs questions et à Mathilde d'y répondre, de manière à ce que les ambiguïtés liées à l'annotation soient traitées au fil de l'eau et qu'elle puisse garder une trace des éventuelles décisions prises dans le cadre du processus d'annotation.

Encadré n°21 : Transformer des guides d'annotations en documents de formation pour les travailleurs malgaches

Ils ont retransformé les guidelines que j'avais créées, de manière beaucoup plus pédagogique pour leurs travailleurs. Ils avaient un lien Google Drive avec toutes leurs questions, et je répondais à leurs questions le plus vite possible (...) Par exemple, page 8, est-ce que c'est un titre ou est-ce que c'est un sous-titre ? je ne me rends pas compte . Je leur répondais oui ce n'est pas facile à voir . Est-ce que c'est une bullet list ? Je ne sais pas si c'est une bullet list . On discutait comme ça.

Q : Est-ce qu'il y avait des choses qu'ils pointaient du doigt ou qu'ils identifiaient, que tu n'avais pas forcément prévues ?

R : Bien sûr !

Mathilde - CTO - BankAI

Un aspect important de l'intégration logistique réside précisément dans le fait de fournir un cahier des charges aux travailleurs du sous-traitant. Dans le cas de l'IA, cet aspect de l'organisation du travail se traduit par la formalisation de guides d'annotations qui regroupent un ensemble de consignes. Il peut s'agir d'un document en format PDF dans lequel l'entreprise va détailler la manière d'annoter, c'est-à-dire s'il s'agit

Donc on a réussi à les faire travailler sur ce logiciel open source, qui donc bug un peu, il faut le dire. (Mathilde - CTO - BankAI)

98. *"La création des guidelines, c'était moi qui la faisais, mais il y avait des moments où... Tu mets des exemples parce qu'il y a différents cas de figure, où tu n'es pas sûr. Il y a toujours un moment où tu ne vas pas savoir et aucune des réponses n'est mauvaise."* (Mathilde - CTO - BankAI)

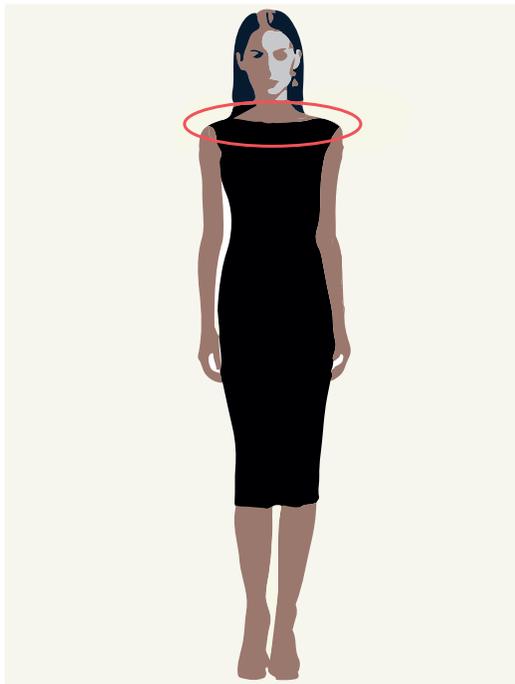
d'encadrer ou de détourer : "avec des Microsoft Word qui dit là (...) vous mettez la boîte comme ça ; la règle c'est vraiment de faire en sorte que chaque paroi de la boîte touche le dernier pixel de la voiture" (Théophile - CEO - Automatik), même si la règle ne va pas nécessairement s'appliquer à toutes les données, car "il y a des subtilités. Parfois, on dit la boîte, on va l'ajuster au contour de l'objet s'il est masqué ou alors vous imaginez l'objet derrière l'arbre, et vous mettez la boîte comme si vous voyiez l'objet en transparence derrière l'arbre (...)" (Théophile - CEO - Automatik).

Dans le guide, se joue également la définition des classes qui s'appliquent aux objets identifiés par les annotateurs. En d'autres termes, il s'agit d'expliquer à la fois le geste que l'annotateur doit réaliser et de définir le concept à annoter. Les deux extraits de guide d'annotation qui suivent permettent de mieux comprendre cette idée de "codification". En l'occurrence, il s'agit d'un guide d'annotation de cinquante-deux pages destiné à entraîner un modèle de reconnaissance de vêtements sur des images et qui concerne uniquement les hauts de vêtement (par exemple : chemise, débardeur...). Sur le premier extrait, on constate qu'il faut reconnaître un vêtement de type "caraco" qui est définie de manière assez vague, puisqu'il est simplement indiqué qu'il s'agit d'un vêtement avec de "très fines bretelles", "souvent similaire à la lingerie"⁹⁹. Dans ce cas, on peut imaginer que le manque de précision dans la définition pourra créer des problèmes au moment de l'annotation. Dans la deuxième image, on constate qu'il y a une indication sur l'image et que la forme de l'élément du vêtement est défini de manière plus précise. Certains guides peuvent être extrêmement complexes selon le nombre de classes à détecter et de la qualité des images.



FIGURE 13 – Capture d'écran du guide d'annotation d'un projet d'IA appliqué aux vêtements

99. Ici, le lecteur doit savoir que nous avons nous-mêmes dû rechercher ce qu'est un caraco



COL BATEAU / BOAT NECK:

- Large, passant au ras du cou et finissant à la pointe des épaules.
- A boat neck, also called a bateau neck is a wide neckline that runs horizontally, front and back almost to the shoulder points, across the collarbone.

FIGURE 14 – Capture d'écran du guide d'annotation d'un projet d'IA appliqué aux vêtements

Ces consignes ont vocation à évoluer tout au long du développement du modèle et de son déploiement à travers les acteurs qui travaillent sur les données et les modèles. Ainsi, dans le cas d'Automatik, le processus de production du guide d'annotation réunit trois acteurs : les clients, les employés de la startup et les annotateurs. Du côté de la startup, les *solutions architect* sont des data scientist qui récupèrent les besoins des clients en les adaptant aux possibilités de l'IA¹⁰⁰. Ce sont eux qui réalisent le travail de modélisation et qui réalisent le guide d'annotation, en interaction avec les clients, la responsable de l'annotation et les annotateurs. Les responsables de l'équipe d'annotation de Madagascar reçoivent ensuite le guide, discutent de sa mise en œuvre et forment les annotateurs¹⁰¹. L'annotation de l'ensemble des données peut alors commencer. Dans ce cas, les annotateurs experts ou les chefs d'équipes reçoivent les commentaires des annotateurs sur les étiquettes et leur pertinence. L'information est ensuite transmise au *solution architect* de la startup. Il s'ensuit un nouveau cycle d'adaptation du guide¹⁰².

100. "Pour l'instant, pour un projet, les solution architects font un document pour expliquer les consignes, soit sur Airtable, soit sur notre outil interne (avant cette fonction n'existait pas, mais maintenant on peut le faire dans notre outil interne)" (Lin - Datasets manager - Automatik)

101. "Le leader lit les instructions et quand tout est clair, il l'explique aux annotateurs qui auront un document de formation pour les accompagner pendant le traitement."(Lin - Datasets manager - Automatik)

102. "Lorsqu'ils rencontrent un cas difficile, ils consultent les responsables. Si [les responsables] ne peuvent pas répondre, ils nous demandent. Notre architecte de solutions et moi-même disposons d'un canal Slack où nous répondons aux questions de manière très concrète. Une ou deux heures après le début du traitement par les annotateurs, les responsables commencent le contrôle de la production." (Lin - Datasets manager - Automatik)

Le CEO du sous-traitant G constate également que le travail de définition des consignes est parfois partagé entre les entreprises d'IA et les entreprises d'annotations. En effet, les clients n'ont pas forcément une vision exhaustive de ce qui est contenu dans les données : *"Les gros enjeux, c'est de comprendre le plus vite possible, de compléter les consignes qui sont souvent partielles - c'est très dur de faire des consignes parfaites parce que de toute façon le client ne voit pas toutes les images. Donc l'objectif c'est d'avoir le plus vite possible une vision de ce que veut exactement le client."* (Pierre - CEO - Sous-traitant G). Par ailleurs, le caractère dynamique de la production des modèles, avec des tests de performance et des tests lors du déploiement, va entraîner des changements dans les consignes.

Dans un autre entretien, Cédric, évoque quant à lui le terme de "convention" soit le fait de se mettre d'accord sur ce que l'entreprise va chercher à annoter et comment cela va être annoté. Ces décisions sont évolutives et dépendront des demandes des clients, autant que des possibilités du modèle : *"le focus de la boîte a un peu shifté et maintenant on n'est plus trop sur ces ECG, mais plutôt sur des ECG de longue durée, qui eux sont revus par des experts, mais sur lesquels les experts passent trop de temps."* Le caractère évolutif et distribué des conventions vient là résonner avec le travail de [Bowker et Star \(2000\)](#).

Encadré n°22 : Construire des conventions d'annotation

Quand je parle de conventions, ça veut dire qu'est-ce qu'on cherche à annoter et comment on va l'annoter . Ce ne sont pas tant des conventions sur la qualité ou des choses comme ça. Au début, on a défini une grosse ontologie [avec un cardiologue]. (...) C'est une liste de labels, avec des liens d'exclusion et d'inclusion. Ça permet d'avoir une vue complète de quand tu regardes un ECG, ce que tu peux dire dessus. On a ça d'une part. Et d'autre part, on a plusieurs autres formats. Par exemple, on a un format qui permet de décrire les différentes ondes dans l'activité cardiaque, par exemple, est-ce que c'est une activité de l'oreillette ou du ventricule, est-ce qu'elle est anormale, si oui, pourquoi ?

Cédric - Lead data - AIMedic

Le travail de formalisation de ces consignes rassemble donc des communautés hétérogènes de pratiques composées des travailleurs reliés par le modèle d'IA : des cardiologues, des médecins utilisateurs de la solution, les développeurs de la startup et les travailleurs de l'annotation, eux-mêmes médecins. Cette formalisation passe par le fait de définir par exemple ce qu'indique tel ou tel état du coeur en fonction de la vidéo ou d'être capable de définir une anomalie cardiaque. Il peut être pertinent de prévoir des formations ou des consignes dédiées sur des aspects spécifiques du modèle. La difficulté réside dans le processus par lequel l'entreprise d'IA fait comprendre son ontologie aux annotateurs, qui, bien qu'étant issus du monde médical, ne sont pas des experts en cardiologie. De ce point de vue, le guide d'annotation est à la fois le reflet d'une

expertise métier dans le domaine médical et dans le domaine de l'IA, ce qui en fait un objet localisé : *"Tout est présent là-dedans. Une bonne partie est expliquée dans des documents que l'on peut partager, mais tout n'est pas expliqué. Il y a des trucs qui actuellement ne sont compréhensibles que par nous et ce n'est pas très grave."* (Cédric - Lead data - AIMedic). La nécessité de déployer le modèle et de le faire valider par les autorités compétentes dans le domaine médical implique cependant de standardiser les résultats du modèle.

L'existence de cette tension rend nécessaire le fait que Cédric, ait dû formaliser des consignes, qui étaient auparavant transmises oralement aux personnes travaillant sur le modèle¹⁰³. Que ce soit pour des projets liés à la reconnaissance de style vestimentaire, ou pour le monde médical, il est alors nécessaire de définir précisément les concepts mobilisés par le modèle pour éviter au maximum les variations dans l'interprétation, et notamment les cas "extrêmes" : *"On a de la documentation aussi de ce que chaque label veut dire et comment est défini chaque seuil. Parce que même une pause du cœur par exemple, certains vont dire que c'est à partir de 3 secondes et d'autres 2,5, etc., donc on a des conventions clairement explicitées."* (Nicolas - CTO - AIMedic) C'est le guide d'annotation qui permet d'inscrire ces concepts en opérant comme un "objet frontière" en tant qu' *"des objets qui appartiennent à plusieurs communautés de pratique et qui répondent aux besoins d'information de chacune d'entre elles"* (Bowker & Star, 1999)(p.297). Finalement, la difficulté de ces guides d'annotations est de trouver un *langage commun* entre toutes les personnes concernées par le travail de catégorisation, clients, utilisateurs, annotateurs, développeurs de solutions d'IA, de codifier les connaissances propres du domaine qu'on cherche à automatiser, elles-mêmes traduites d'un point de vue informatique.

Le cas de l'entreprise de mode ModeAI illustre également cette dimension du travail de l'IA et la nécessaire coordination des travailleurs qui participent au travail de catégorisation. Pour que l'outil soit capable de classer des vêtements dans des catégories, il a fallu construire des classifications notamment avec des étudiantes d'une école de mode parisienne : *"tout ce qu'on a constitué avec l'IFM [Institut Français de la Mode], entre guillemets c'est le gros book de règles, de classification, de gestion des ambiguïtés, qu'on a remis à [sous-traitant B] pour pouvoir dire voici les ambiguïtés, voilà comment on veut faire ça "* (Jean - CTO - ModeAI) Ce guide de classification a d'abord été construit avec de l'expertise sur le domaine ce qui a abouti à une granularité très fine. Puis, ils ont testé ces catégories auprès de consommateurs en les interrogeant sur leur compréhension des catégories. L'objet ainsi construit mélange ces deux points de vue en équilibrant la finesse des catégories, la capacité à définir les images et la nécessité de ne pas surcharger le modèle avec trop de catégories. C'est sur la base de ces règles que sont formés les annotateurs¹⁰⁴.

103. *"Pendant longtemps ça a été de la tradition orale sur comment on les apprend, vu qu'on travaillait en interne ou avec des gens qui eux-mêmes travaillaient sur ces définitions(...) Mais (...) vu que l'on commence à multiplier le nombre de gens qui doivent appliquer ces conventions, on essaie (...) de les décrire de manière sans équivoque, dans des documents, et de fournir des exemples au fur et à mesure qui permettent de bien illustrer les cas limite et les trucs comme ça. Honnêtement, dans notre stade de développement qui est ce qu'il est, (...) on a des définitions vraiment nickel pour nos deux principaux types d'annotations."*(Cédric - Lead data - AIMedic)

104. *"c'est lui qui les a formés en faisant des tests de labellisation. Ils labellent, on voit ce qui ne marche pas très bien. Et donc Alex*

Durant le processus d'annotation, ce guide de règles va également changer, en fonction du travail des annotateurs. Il va d'abord être transformé pour être plus pédagogique pour les travailleurs des données. Ainsi, du côté des entreprises de sous-traitance, la difficulté va être de traduire ce genre de documents en un parcours de formation. Ensuite, il sera également amené à évoluer en fonction des nouvelles ambiguïtés qui émergent des milliers d'images annotés à Madagascar¹⁰⁵ (Jean - CTO - ModeAI). De ce point de vue, il est clair que ce cahier des charges façonne le travail d'annotation, mais qu'il est aussi transformé durant ce processus, notamment du fait des échanges récurrents entre les dirigeants de ModeAI et les annotateurs, au fil de la montée en compétences de ces derniers.

Encadré n°23 : Gérer collectivement le processus d'annotation

J : et eux, le process de gestion de la base de connaissance, pour l'instant il est vraiment bien. Maintenant il reste aussi à valider des petits trucs avec eux, et s'améliorer aussi là-dessus, savoir comment travailler avec la fréquence de mise à jour de la base de connaissance si il y a besoin (...) En fait nous on leur transmet un bouquin, et à priori eux ils ont un bouquin, ils retranscrivent ça de façon encore beaucoup plus simple pour des opérateurs (...) C'est une traduction on va dire chez eux selon leur code de montée en compétences des opérateurs

Jean - CTO - ModeAI

Lors des entretiens avec les dirigeants d'entreprises d'annotations, ces derniers mettent en avant la variabilité des niveaux de maturité des entreprises d'IA sur le processus d'annotation et sur la codification des consignes d'annotations. Dans ces cas, les entreprises d'annotations vont participer à construire les consignes d'annotations, en s'appuyant sur leur connaissance de cette partie du processus de production. Ils savent comment rendre les consignes plus claires, parfois aussi quels sont les aspects des données qui vont provoquer des ambiguïtés. Ils disposent alors de compétences spécifiques utiles pour les entreprises d'IA.

Encadré n°24 : La maturité des processus d'annotation

Parmi les entreprises à qui on parle, déjà il y a différents niveaux de maturité. Il y en a qui sont très matures dans leurs besoins et qui disent J'ai besoin de ça, ça et ça et je vous envoie un document qui exprime ce dont j'ai besoin. et d'autres où c'est beaucoup plus en amont, ils commencent à y réfléchir, ils ont besoin d'être accompagnés pour bien décortiquer leur sujet, etc. C'est donc assez

il leur explique les règles etc. ... et aussi quand on voit des images qui sont un peu limites, on essaie de les instruire pour affiner les règles par rapport à un très gros volume d'images." (Jean - CTO - ModeAI)

105. "Parce que tu trouves toujours des nouvelles petites ambiguïtés liées au fait que tu labellises des dizaines de milliers de photos. C'est comme ça qu'on [réfléchit sur les] sur les règles, pour les préciser, etc. ... "

variable là-dessus. De notre côté, on travaille beaucoup justement pour aider à les guider dans cette démarche de réflexion. On est capable de travailler avec des entreprises qui ne sont pas encore extrêmement matures sur leurs besoins.

Victor - CTO - Sous-traitant B

Il existe encore très peu de littérature scientifique en sciences sociales consacrée au processus d'annotation et de documentation. On peut toutefois repartir du travail de [Miceli, Yang, et al. \(2022\)](#) qui reprend le cadre théorique de [Bowker et Star \(2000\)](#) en considérant ces documents comme des "objets frontières". D'emblée, on peut souligner le fait que le concept est analysé par [Bowker \(1997\)](#) comme le résultat d'un processus de coopération : *"Il s'agit d'accords de travail qui résolvent les anomalies de naturalisation sans imposer une naturalisation des catégories d'une communauté ou d'une source extérieure de normalisation. (Ils sont donc particulièrement utiles pour analyser les situations de coopération et d'égalité relative; les questions relatives à l'imposition impérialiste de normes, à la force et à la tromperie ont une structure quelque peu différente."* (p.297)

Le travail de [Miceli et al. \(2020b\)](#) montre pourtant que ce travail de documentation est imposé aux annotateurs. En effet, malgré le fait que la documentation serve à améliorer les compétences des travailleurs, ces derniers *"ne reçoivent que le strict minimum d'informations leur permettant d'accomplir les tâches conformément aux instructions des clients. Le retour d'information des travailleurs est rarement pris en compte et n'est qu'occasionnellement intégré dans les instructions de travail."* Dans les cas étudiés ci-dessus, on remarque toutefois que les documents d'annotations évoluent au fil des projets, y compris avec les retours des annotateurs, comme on va le voir dans la section suivante, et ce, bien que ce processus se passe dans un contexte hiérarchique. Le cas de ModeAI est à ce titre remarquable, les deux fondateurs confessant en entretien qu'ils ont préféré conserver quelques annotatrices à temps plein à Madagascar, alors même qu'ils n'ont plus suffisamment d'argent pour se payer eux-mêmes, arguant de la nécessité de conserver ces compétences.

Comment analyser ce résultat au regard de la littérature existante ? Nous pensons que la cause de ce résultat en apparence contradictoire est d'abord à chercher dans les modalités méthodologiques de notre thèse, qui intègre la vision des clients des entreprises d'annotation. Ensuite, nous pensons que l'imposition se situe moins dans les pratiques d'annotations en tant que telles, que dans la relation hiérarchique qui existe entre les donneurs d'ordres et les travailleurs des données. Ce résultat a également un impact sur la manière de réguler l'industrie de l'IA. Ainsi, l'AI Act exige des producteurs de systèmes d'IA qu'ils divulguent les processus qui sous-tendent leurs solutions. Nos résultats montrent que ce règlement demeurera inopérant dans le contexte commercial si les pratiques d'annotation des données ne sont pas documentées. Une "éthique de l'IA" passe nécessairement par une éthique du travail des données.

4.2.2 Une forme d'organisation plus proche de la hiérarchie pour prendre en compte les boucles de rétroactions humaines sur les modèles

Nous avons commencé à entrevoir les mécanismes d'interactions qui existent entre les entreprises de la chaîne de production de l'IA au travers des outils d'annotation et de la documentation. Une autre dimension de cette intégration passe par le recours à des outils de communication et de contrôle du travail effectué par les annotateurs. Victor, directeur opérationnel du sous-traitant B, oppose le modèle "BPO" avec le modèle "plateforme", en insistant sur le besoin d'avoir un contact avec le client final : *"on a souvent besoin d'être en contact direct avec le client final et que quand il y a cet intermédiaire (une plateforme), ça ne se fait pas trop. C'est une limitation à ce modèle."* (Victor - CTO - Sous-traitant B) Il souligne ainsi deux aspects de l'organisation du travail d'annotation : la présence d'une chaîne hiérarchique dans les locaux du sous-traitant et la mise en place de canaux de communication avec les clients¹⁰⁶.

Tout au long de la phase d'annotation, le processus associe ainsi les acteurs de toute la chaîne de valeur, de manière à ce que chacun puisse exprimer ses contraintes : *"on a vraiment une démarche partenariale tous les trois où le client final fait part de son besoin, l'entreprise d'IA qui propose des services de data science fait part de ses contraintes sur l'annotation de données, nous on fait part de nos contraintes aussi sur l'annotation de données, et puis après on travaille ensemble tous les trois."* (Victor - CTO - Sous-traitant B) Cette organisation du travail intégrée devient nécessaire afin de préciser les concepts à annoter, surtout lorsqu'ils changent ou qu'ils incluent des éléments de subjectivité. Mathilde évoque par exemple des difficultés dans la définition de ce qu'est une date et les variations qui sont susceptibles de survenir pendant le travail d'annotation¹⁰⁷. L'utilisation d'outils de communication directes avec les annotateurs et leurs managers présents à Madagascar permet d'inscrire les décisions concernant les évolutions de l'annotation dans des documents et des messages, ce qui permet afin que les données soient les plus comparables possible, y compris lorsque de nouvelles consignes émergent des données. Pour ce projet, les travailleurs pouvaient également poser des questions ou faire des remarques sur un tableur afin de faire remonter les doutes et les ambiguïtés.

Encadré n°25 : Résoudre les ambiguïtés

R : (...)Donc soit je tranchais, soit je disais je n'en sais rien, qu'est-ce que vous en pensez ? , et là ils répondaient je ferais plutôt ça, ou ça .

106. *"Déjà, le N+1 direct des annotateurs est avec eux à Madagascar comme ça ils ont vraiment une relation de proximité. Pour ceux qui sont à distance, par exemple moi ou d'autres personnes dans l'entreprise, ça fonctionne comme dans n'importe quelle entreprise dans le sens où on fait des conf calls, on s'envoie des mails et il y a effectivement un Slack et de la gestion de projet avec des outils en ligne."* (Victor - CTO - Sous-traitant B)

107. *"On a eu des tonnes et des tonnes de débats, parce que à votre date de résiliation de contrat , est-ce que c'est une date, ou est-ce que ce n'est pas une entité du tout ? C'est compliqué. (...) Ce n'est pas du tout un truc que tu peux faire sur une plateforme, sans retour de la part de la personne. Il faut pouvoir communiquer avec les gens, parce que ça dépend aussi de ce que tu veux, de tes besoins en tant qu'entreprise qui fait appel aux annotateurs, de tes clients."* (Mathilde - CTO - BankAI)

Q : Et donc il y avait quand même un échange.

R : Oui. (...) Il y avait Simon, Pierre et la fille (...), celle qui a vulgarisé toutes les guidelines que j'ai faites. Après, avec les travailleurs eux-mêmes (...) il y a eu une autre personne qui a travaillé là-[sur des] foires aux questions

Mathilde - CTO - BankAI

Le fondateur de ModeAI considère aussi que l'interaction au fil de l'eau constitue une manière de résoudre les problèmes concernant l'annotation pour produire des données de qualité : *"la qualité elle passait par la relation plus que par les références (...) on travaille toujours en fait avec les équipes, et on dit quels sont les problèmes, et on résout les problèmes ensemble."* (Jean - CTO - ModeAI). Dans ce contexte, la startup met en avant la nécessité de discuter des difficultés posées par certaines images de vêtements, qui passe, encore une fois, cette fois-ci encore par une communication directe et même par des réunions régulières avec leurs annotatrices : *"C'est le meilleur mode de fonctionnement qui permet d'avoir une communication directe. Et les erreurs on les traite, les blocages on les traite c'est ce qui permet d'avoir la meilleure qualité au final."* (Jean - CTO - ModeAI) L'enjeu est également de mieux maîtriser la dimension subjective de l'annotation, certains vêtements pouvant être interprétés différemment selon les annotateurs. Il considère également que les travailleurs ont un rôle de vérification sur ce que contiennent les données, par exemple, concernant de nouveaux éléments, ou des éléments non prévus initialement dans la documentation qui doivent être remontés à la startup : *"(...) Pour nous c'est la meilleure solution, d'avoir une base de connaissance que l'on transmet, et (...) avec le bon niveau d'apprentissage, de détails, d'instruction, et d'avoir un double retour. Parce que si tu n'as pas de double retour, tu ne sais pas t'améliorer (...) c'est à ça que ça sert aussi de ne pas passer par une plateforme, d'avoir des process en place pour faire remonter l'information"* (Jean - CTO - ModeAI).

On constate un fonctionnement similaire chez AIVoice, même si dans ce cas les annotateurs sont internalisés : *"On a maintenu le même fonctionnement avec une réunion hebdomadaire avec l'équipe annotation, où on échange sur la semaine passée"* (Eleonor - CTO AIVoice). Lors de ces échanges, les équipes de recherche et développement discutent avec les annotatrices de certains segments audio difficiles à juger. Certains mots et certaines émotions sont parfois compliqués à déterminer : *"on fait de la réécoute de certains segments où on n'arrive pas trop à juger."* (Eleonor - CTO AIVoice) Enfin, les échanges vont aussi servir à *"enrichir un peu la plateforme d'annotation et le protocole d'annotation pour des besoins précis ou si on se rend compte qu'on est passé à côté de quelque chose."* (Eleonor - CTO AIVoice)

Dans le cas d'Automatik, toutes les discussions se déroulent sur un outil de messagerie à destination des entreprises, auquel nous avons eu accès lors de notre terrain ¹⁰⁸. Au sein de cette messagerie, on retrouve les

108. R : On a une chaîne sur Slack avec les annotateurs et notre solution architect ou moi, pour répondre aux questions très concrètement (...) s'il y a des erreurs, je peux toujours les recontacter. (sur le Slack) (Si Wu - Data lead Automatik)

”solution architect” qui développent les modèles et font l’interface entre le client final et les annotateurs, la personne en charge de l’annotation chez Automatik et les annotateurs et leurs managers sur place sont présents sur ce canal de communication et interagissent entre eux au ”fil de l’eau”. Les captures d’écrans ci-dessous donnent des indications sur la nature des échanges lors du projet de reconnaissance des plats de cantine. Une ”anomalie” a été détectée par un annotateur dans les données, la boisson présente sur le plateau n’est pas prévue dans les classes d’annotation au sein de la documentation et surtout elle n’est pas présente dans le ticket de caisse auquel ont accès les annotateurs. Une discussion s’engage alors entre Eva, la dirigeante de l’entreprise d’annotation A, l’annotateur et le développeur en charge du projet. Dans un premier temps, c’est Eva qui prend la décision de ne pas l’annoter pensant qu’il s’agit d’une ”boisson gratuite” qui sera confirmée par le solution architect (*”ça doit être une boisson gratuite pour le repas de noel”*)

Travailleur sous-traitant A 17 h 37
 Hello @Solution Architect Automatik. On remarque la présence de ce boisson sur nombreux plateaux alors qu’il n’y a pas de tag pour l’annoter dans le ticket
 @Travailleur sous-traitant A uploaded a file : Capture 22_01.PNG



17 h 38 **Solution Architect Automatik**
 Je n’ai pas accès à cette image

CEO sous-traitant A 17 h 40
 ■ peux tu envoyer l’ID stp @faly ? moi aussi je doute q’il faut activer le tag «ticket incomplet» ou non, il peut aussi s’agir d’une boisson gratuite je ne sais pas

17 h 42 **Travailleur sous-traitant A**
 d’accord @CEO sous-traitant A

Solution Architect Automatik 17 h 47
 ça doit être une boisson gratuite pour le repas de noel

17 h 47 **CEO sous-traitant A**
 je suis du même avis donc, pas d’activation de tag ■

Travailleur sous-traitant A 17 h 48
 D’accord, merci beaucoup

17 h 48 **Solution Architect Automatik**
 Est-ce que vous l’annoter du coup ?

FIGURE 15 – Capture d’écran d’un échange entre des annotateurs et des chefs de projet de l’entreprise Automatik

Dans un second temps, le développeur d’Automatik va proposer de quand même annoter cette photo en l’intégrant à une classe qui avait déjà été créée suite à des remontées de la part des annotateurs. Il s’agit de

la classe "pain gratuit" qui sert à annoter les morceaux de pain distribués gratuitement dans les cantines d'entreprises. La consigne va ensuite être intégrée au cas où des situations similaires se présenteraient.



FIGURE 16 – Capture d'écran d'un échange entre des annotateurs et des chefs de projet de l'entreprise Automatik

Le dirigeant de AIMedic nous indique également qu'ils communiquent "par Slack avec [les annotateurs]. On les a ajoutés sur des canaux privés et, du coup, quand ils ont une question ou un doute, on est ouvert à ce qu'ils viennent nous poser des questions." (Nicolas - CTO - AIMedic) Ils ont également mis en place des fichiers de questions sur lesquels les annotateurs en charge de la vérification des annotations peuvent ajouter des commentaires en cas de doutes ou de difficultés. AIMedic a également prévu de mettre en place "des systèmes de shadowing : de temps en temps on va annoter avec eux (...) On s'appelle, ils partagent leur écran et on discute." (Cédric - Lead data - AIMedic). Ce système de travail est d'autant plus important que le travail de catégorisation des éléments présents sur les vidéos d'ECG est source de nombreuses ambiguïtés. Il permet d'aménager des temps d'échanges et de réponses aux questions¹⁰⁹. Au-delà de ces phases d'échanges avec le client final, les annotateurs cherchent également à parvenir à un consensus entre eux, ne gardant que les ECG les plus difficiles pour leurs discussions avec les clients. Cet élément traduit un certain degré d'autonomie dans la décision d'annotation est ressorti de plusieurs autres entretiens. En effet, il y a une tension

109. "ça permet de répondre à leurs questions, de vérifier qu'ils utilisent bien la plateforme, et pour l'instant ils ne sont que quatre donc ça va, mais on imagine qu'à l'avenir ça permettra de s'assurer qu'il n'y ait pas de gens qui n'annotent pas bien ou qui ne connaissent pas bien les outils." (Cédric - Lead data - AIMedic)

entre le fait de ne pas prendre la décision à la place du client et le fait de ne pas le déranger pour "rien", c'est-à-dire pour un cas qui aurait pu être résolu sur place, à Madagascar.

Encadré n°26 : Transférer la résolution des ambiguïtés aux annotateurs

Il y a ça, et ils font des consensus entre eux aussi. On est en train de voir avec [le sous-traitant B], parce qu'ils connaissent un cardiologue là-bas qui avait fait un peu de formation aussi avant que l'on commence à bosser avec Dina, et j'aimerais bien qu'ils le fassent intervenir, parce qu'elles font des consensus entre elles, et elles gardent des ECG difficiles dont on discute ensemble. J'aimerais bien qu'il y ait une première passe avec cette personne pour que tout le monde s'aligne et que seulement les ECG qui restent après revus par ce cardiologue nous arrivent à nous, pour que ça nous allège encore un peu le check.

Cédric - Lead data - AIMedic

Pour Cyril, en charge de l'annotation, il s'agit finalement d'un contact "*direct avec les équipes*" qui passe par des réunions qui ont lieu toutes les semaines. Comme dans le cas de ModeAI, les annotateurs sont encouragés à parler de leurs doutes ou de leurs difficultés de manière à ne pas altérer la qualité de l'annotation et à diffuser les changements éventuels de règles à l'ensemble des travailleurs des données : "*on fait des réunions toutes les semaines au minimum, parfois un peu plus quand elles ont des questions ou pour regarder des cas difficiles. Quand elles rencontrent des cas difficiles aussi, elles peuvent le partager sur le [canal de discussion]. Elles ne le font pas trop spontanément, de temps en temps si, mais je les encourage à le faire parce que ça nous permet de répondre vite fait et ça donne de l'info à toutes les annotatrices.*" (Cédric - Lead data - AIMedic). C'est en ce sens que le dirigeant d'une entreprise d'annotation explique que : "*Pour certains types de projets, il faut que les gens soient tout le temps sur place, soient encadrés, qu'il y ait des chefs d'équipes qui comprennent mieux, qu'ils puissent vraiment leur expliquer ce qu'il faut faire.*" (Pierre - CEO - Sous-traitant G). Ces projets se caractérisent par leur durée et par leur complexité ce qui se traduit souvent par des consignes très longues à comprendre et à assimiler, ainsi que par les spécificités du projet d'annotation et par le risque d'ambiguïtés, qui justifient la mise en place des outils de communication : "*Nous c'est tout de suite aussi la mise en place d'une communication avec le client (...) ça peut être une chaîne mail. Et aussi beaucoup tout de suite ce qu'on appelle un fichier question où les annotateurs peuvent dire cette image, on n'est pas sûr, est-ce qu'il faut faire ça ?, et ensuite le client va répondre.*" (Pierre - CEO - Sous-traitant G). Finalement, l'intégration logistique des travailleurs est particulièrement importante dans la mesure où on assiste à la mise en place d'une véritable chaîne hiérarchique.

Encadré n°27 : Communiquer avec le client pour améliorer la qualité des données 1

En ayant les gens sur place et sur la durée, ils peuvent vraiment apprendre et progresser. Il y a des chefs de projets. On n'est pas du tout en mode boîte noire . Mes chefs de projet vont communiquer directement avec les clients, même si je suis souvent là pour les aider. Ils communiquent avec les clients, sur Slack, sur des Google spreadsheets. Et donc quand on a des clients, ça se passe bien parce qu'il y a ce côté à l'écoute, un peu service premium (même si c'est un peu exagéré), en tout cas avec cette idée de bien comprendre ce qu'ils veulent et d'être au contact pour que ça soit vraiment qualitatif.

Pierre - CEO - Sous-traitant G

Si les clients finaux ne sont pas nécessairement en contact direct avec les annotateurs "simples", c'est le cas pour les managers malgaches qui encadrent ces derniers : *"Très tôt, quand on arrive à avoir un vrai lien avec le client, on le met en contact avec ce qu'on appelle chez nous le chef d'équipe, qui va être le garant de s'assurer que son équipe (ils sont 15) puisse nous donner un feedback."* (Dinasoa - CEO - Sous-traitant E). Cette culture des retours sur les annotations et les données va se traduire par des échanges pendant lesquels les annotateurs demandent des précisions ou à l'inverse pendant lesquels les travailleurs expliquent certaines décisions, notamment sur les données non prévues dans les consignes : *"ils font une capture écran ou ils enregistrent et ils disent Voilà, on a remonté tel cas sur tel lot de travail ; on a appliqué tel concept ; est-ce que vous êtes d'accord ? Ou à l'inverse On a vu tel cas ; on ne sait pas comment y répondre ; est-ce que vous pouvez nous aider ?"* (Dinasoa - CEO - Sous-traitant E). Nous qualifions cette pratique de *"boucle humaine de rétroaction"* qui se superpose aux couches techniques des modèles d'IA.

Cette dimension du travail d'annotation s'est traduite dans la culture de cette entreprise et même dans sa proposition de valeur. En effet, il s'agit d'une pratique qui est mis en avant par les dirigeants comme une valeur ajoutée par rapport aux plateformes concurrentes, car comme l'explique Dinasoa, les clients ne peuvent pas vérifier ce qu'il y a dans l'ensemble des corpus de données. Dans certains cas, il s'agit de projets qui se comptent en millions d'images à faire annoter : *"On essaie aussi de beaucoup renforcer la culture de feedback. Vous pouvez faire tous les samplings que vous voulez, toutes les passes que vous voulez, si dès le départ c'est mal compris, ça ne marche pas."* (Dinasoa - CEO - Sous-traitant E). À travers cette organisation du travail, on retrouve à nouveau l'idée d'un couplage fort entre les données et les modèles qui se traduit de manière concrète par des modalités d'organisation qui facilitent le couplage développeurs d'IA-annotateurs.

Encadré n°28 : Communiquer avec le client pour améliorer la qualité des données 2

Pour faire ça, on essaie d'avoir le moins de hiérarchie possible. D'essayer de faire en sorte que l'on ait un lien très proche et que nos clients aient un lien très proche avec ceux qui font l'annotation. On est en train de travailler sur des mécanismes pour pousser nos équipes à avoir la culture de

feedback, de dire Il y a un truc, je ne suis pas sûr, je ne sais pas si le concept va s'appliquer, il y a un lot de travail qu'on a reçu où tels éléments nous posent problème . On essaie de faire en sorte que le regroupement de tous ces éléments puisse nous aider à travailler contre ces biais cognitifs.

Dinasoa - CEO - Sous-traitant E

Du côté des annotateurs, cette dimension est être complètement intégrée. Dans l'encadré qui suit, Angy, qui travaille comme annotateur chez le sous-traitant G, raconte le cas d'images mal annotées par ses collègues et qu'il a dû corriger. Pendant cette séquence, il a en effet le rôle de vérificateur, c'est-à-dire qu'il contrôle le travail réalisé précédemment. Dans ce cas, il explique les deux manières de procéder, soit sur une chaîne de messagerie (Slack), soit à travers l'outil d'annotation utilisé au sein de son entreprise. Son message va être remonté par sa chef de projet française jusqu'au client final, le CEO de la startup.

Encadré n°29 : Communiquer avec le client pour améliorer la qualité des données 3

Q : c'est peut-être pour ça que vous ne vous souvenez pas du coup de ce qu'il y avait sur l'image ! et vous avez demandé par slack ?

B : parfois on fait par slack, si c'est vraiment urgent on fait par slack.

Q : et là, le dernier exemple dont vous parlez là ?

B : c'est en faisant la review

Q : par slack ?

B : oui j'ai demandé par slack et elle a dit je vais demander [au client].

Q : ok, donc elle, elle a fait un mail du coup [au client] ?

B : elle m'a dit tout simplement qu'elle envoyait le lien, et je peux voir le retour du [client], parce que dans l'interface il y a une petite boîte de dialogue où on peut échanger

Q : donc en fait vous pouvez échanger avec [le client] aussi directement sur [sous-traitant K]

B : mais en fait il faut que ça passe par [la cheffe de projet française] d'abord, c'est elle qui envoie au [client]. Parce que parfois elle connaissait déjà (le projet)

Angy - Annotateur - Sous-traitant G

Les différents cas présentés dans cette sous-section illustrent la question suivante : qui s'occupe de l'algorithme quand il déraile ? Dans quelle mesure est-il important pour les entreprises d'éviter les faux positifs et de s'assurer que les bonnes informations d'identification ne sont pas laissées de côté ? Nous montrons que les décisions de modification des modèles se répercutent donc sur l'ensemble de la chaîne, ce qui explique qu'il est nécessaire d'établir des modes de communication susceptibles de tenir compte de ces décisions. Les cas étudiés jusqu'à présent montrent justement que cette structuration du travail de l'annotation est

susceptible d'éviter ces situations. Par ailleurs, il faut se demander : qui regarde les données ? Nos résultats montrent que ce travail est partagé entre plusieurs acteurs de la chaîne, chez les clients des entreprises d'IA (chef de projet), chez les entreprises d'IA (data scientist, manager de l'annotation), chez les entreprises de sous-traitance (managers, travailleurs des données). Toutefois, ce sont les travailleurs des données qui regardent ce qu'il y a dans l'ensemble des données. Ils sont donc les plus à même de faire remonter les imprévus qui surgissent lors de l'annotation. Par conséquent, bien que les développeurs d'IA ne reconnaissent pas toujours l'expertise des travailleurs des données, nos recherches montrent que dans le contexte de l'IA appliquée aux produits commerciaux, les entreprises valorisent la capacité des travailleurs à comprendre le domaine d'application des modèles et des données, suggérant donc que cette activité est déterminante dans le processus de production.

* * *

Dans ce chapitre, nous avons constaté que les travailleurs communiquent régulièrement avec les employés de la startup afin d'atteindre un niveau d'expertise suffisant pour produire des données de qualité. Des conversations sur Slack, ainsi que des réunions téléphoniques, sont utilisées pour communiquer avec eux sur les guides d'annotation. Les employés de la startup et les chefs d'équipe discutent des doutes et des erreurs courantes en matière d'annotation. Cela nous a également permis de commencer à mieux comprendre l'organisation du travail d'annotation.

Ainsi, nous avons analysé la manière dont se traduit l'intégration des travailleurs. Dans la théorie des firmes, on retrouve différentes formes d'intégrations. Elle peut être organisationnelle lorsqu'il existe des relations poussées entre les travailleurs du sous-traitant et la firme focale, ce que nous avons pu constater au travers des outils d'annotations, d'échanges (système de messagerie et du travail commun sur la documentation). La mise en place de ces outils au sein des entreprises d'IA permet une remontée d'information sur la production des données. Ces informations sont quantitatives et qualitatives et illustrent une dernière forme d'intégration à travers le cahier des charges (documentation). L'utilisation de ces outils de communication constitue donc la pierre angulaire du fonctionnement de cette chaîne de sous-traitance. De manière similaire à certaines plateformes numériques de travail, ces outils servent à mettre en place une "hiérarchie électronique" qui permet une communication en temps réel entre l'ensemble des acteurs de la chaîne.

Au travers de ces outils, nous avons également constaté l'existence de mécanisme de *human feedbacks loop* avec des annotateurs faisant des remontées sur les données qui conduisent à des modifications sur la documentation. Pour nous, l'annotation est donc le résultat d'un processus qui passe par la mise en relation des différentes "couches" humaines des modèles, c'est-à-dire l'ensemble des acteurs de la chaîne à travers les outils de standardisation du modèle. Par conséquent, nous considérons que cette forme d'organisation de la sous-traitance se rapproche de la forme "hiérarchique" (et donc de l'entreprise classique), tout en ob-

servant qu'il existe une marge d'autonomie des travailleurs dans le choix de certaines annotations et dans la capacité à faire remonter les données "aberrantes", au sens statistique. Dans la prochaine section, nous proposons d'analyser des modes d'organisations différents de ce que nous avons abordé jusqu'à présent en soulignant la pertinence de la distinction entre "hiérarchie" et "marché".

Chapitre 5 - La chaîne de valeur de l'intelligence artificielle : comprendre l'intégration des travailleurs dans la production de l'IA

Après avoir consacré notre chapitre quatre aux facteurs qui expliquent l'agencement du marché des données pour l'IA par la qualité et aux modalités qui permettent de l'atteindre, revenons maintenant à la question du fonctionnement des chaînes de productions de l'IA et à celle des activités nécessaires au développement des modèles. D'un point de vue macro, nous analysons, comment est structurée la chaîne de production de l'IA? Quelles sont les activités nécessaires au développement des modèles?

Dans ce chapitre, nous reprenons le concept de firme-réseaux définies comme des entités économiques hybrides spécifiques, entre le marché et la hiérarchie afin de mieux comprendre la structure du marché et les relations de pouvoir structurelles qui en découle. A travers ce concept, Baudry (2003) et Chassagnon (2014) s'inscrivent dans la perspective de Williamson, qui avait déjà identifié une forme hybride d'organisation se traduisant par exemple par des relations de long terme entre un sous-traitant et une firme focale. En insistant sur cette forme-ci, ils se situent pleinement dans la perspective d'étude soulevée par les "chaînes globales de valeurs" et portant sur l'hétérogénéité des structures de gouvernance des chaînes de sous-traitance et la variété des rapports de pouvoir qui existent entre les acteurs.

Cette question est d'ailleurs au cœur de la thèse de Virgile Chassagnon (2010). Il considère en effet que, par rapport au contrat et à la propriété la variable du pouvoir est trop peu prise en compte en économie, alors même qu'elle est au centre de la construction des organisations. S'appuyant sur les théories des organisations, il propose d'analyser les firmes, non pas comme des "noeuds de contrat", mais comme des noeuds de pouvoir et donc des noeuds de relations formelles et informelles qui structurent l'accès des agents économiques aux ressources. Il défend l'hypothèse selon laquelle les firmes font coexister trois types de pouvoirs : juridique de par la nature du contrat, *de jure* de par la propriété des actifs de la firme et *de facto* de par la dépendance des acteurs vis-à-vis de ressources critiques, notamment les ressources humaines¹¹⁰. Si la firme-réseau est spécifique c'est parce qu'elle se fonde sur des liens sociaux intenses entre les divers partenaires et donc sur rapports de pouvoir entre ces mêmes acteurs¹¹¹.

110. Le pouvoir juridique s'exerçant par le contrat, tandis que le pouvoir *de jure* par la propriété des moyens de production (machines...)

111. "La distribution du pouvoir entre les différents acteurs est le principal mécanisme de coordination dans un réseau vertical de

Lorsque des relations économiques s'établissent entre des firmes juridiquement indépendantes, elles se regroupent en réseaux pour définir les frontières d'une même entité de production. Ainsi, la portée économique d'une entreprise de réseau ne peut être limitée à son périmètre juridique. Elle se définit au regard du périmètre de son pouvoir. Les conditions préalables à l'émergence des entreprises de réseau semblent donc reposer avant tout sur de réelles imperfections juridiques, en ce sens que cette organisation complexe va au-delà des fondements juridiques découlant des relations de travail.

Ces différentes approches permettent d'aborder autrement le concept des plateformes numériques de travail évoqué dans notre revue de littérature. Ces dernières ont été considérées comme des entités économiques spécifiques, du fait de leur positionnement d'intermédiaire coordonnant deux faces d'un marché (marché biface). Ce fonctionnement économique est à la fois mis avant par les plateformes elle-mêmes, mais également par certaines recherches qui revendiquent l'existence de modèles de gouvernance spécifiques aux plateformes. Les analyses de Chassagnon (2014) restent d'autant plus pertinentes que le modèle de plateforme est questionné à la fois par le droit du travail et par la subordination économique exercée par ces dernières sur les travailleurs.

Nous pensons qu'il est nécessaire de questionner la variété des formes que prennent ces entreprises, notamment du point de vue de l'intégration des employés (Cornet et al., 2022). La caractérisation des formes de gouvernance de ces firmes réside alors dans le degré d'intégration logistique des employés, que ce soit par la technologie (avec des mécanismes de contrôle de la qualité) ou par l'intégration organisationnelle, par des mécanismes de contrôle (et de délégation du contrôle) et d'incitations par le salaire (à la performance). Dans le contexte de l'IA, on peut alors se demander pourquoi et comment les travailleurs des données sont intégrés ou non à la création du produit final basé sur l'intelligence artificielle? Ici, l'analyse par le niveau des chaînes globales de valeur apparaît pertinente afin d'approfondir la question des modes de gouvernance des relations inter-firmes, dans le contexte de la mondialisation des échanges et du travail.

En faisant du pouvoir la question centrale de l'arbitrage entre "marché" et "firmes" dans les firme-réseaux, Chassagnon (s. d.) rejoint les analyses des théoriciens des GVC qui identifient quant à eux des firmes pilotes capables de déterminer la structure et les conditions de production d'un marché. L'analyse diffère toutefois dans la reconnaissance de l'hétérogénéité des relations inter-firmes. Le concept de firme-réseau est unifié, dans la mesure où il désigne indifféremment des entités, pour peu qu'elles coordonnent des réseaux de sous-traitants.

production. Analyser les relations de pouvoir qui se tissent dans la firme-réseau implique en premier lieu d'admettre un fait saillant de son existence : l'absence de fondements juridiques de la relation de subordination professionnelle. Il ne saurait exister dans la firme-réseau une autorité autre que l'autorité contractuelle dite de résultat émanant de la relation de sous-traitance. Autrement dit, la firme-réseau est strictement empreinte de pouvoirs mais strictement dépourvue d'autorité (...). Les entités légalement autonomes sont intégrées et coordonnées par la firme focale en vertu d'un pouvoir qui ne résulte pas de contrats mais de l'organisation elle-même." (Chassagnon, 2014)

La théorie des chaînes de valeurs insiste quant à elle sur la variété des chaînes de production et sur la variété des mécanismes qui permettent de coordonner les acteurs et de redistribuer la valeur produite. Les deux théories se rejoignent toutefois sur les conséquences en termes de relation d'emploi (Pun, Tse, Shin, & Fan, 2020). Ces nouvelles formes de coordination créent un brouillard organisationnel qui favorise le développement des zones grises de l'emploi : "l'autonomie dans la subordination" pour certains salariés et "l'allégeance dans l'indépendance" pour certains travailleurs indépendants (Chassagnon, 2014)¹¹². Dès lors, on pourrait considérer les PNT comme des firmes-réseaux dont le pouvoir s'exercerait à travers les outils algorithmiques. L'enjeu est alors de questionner le caractère exceptionnel de ces entités économiques, par rapport aux frontières classiques des firmes à travers le cas des plateformes d'annotation de données.

Lors de notre étude de terrain, nous avons constaté que la production de données différait considérablement de ce que la littérature internationale décrit comme du micro-travail standardisé. La sous-traitance à travers des réseaux d'externalisations directes est ainsi plus courante que l'utilisation de portails Internet en libre-service comme Amazon Mechanical Turk pour trouver des travailleurs. Nous avons interrogé des travailleurs qui ont expliqué utiliser des "plateformes", mais il s'agissait le plus souvent de logiciels en ligne fournis par leurs employeurs BPO, plutôt que de véritables places de marché pour le micro-travail.

Ce faisant, nous approfondissons l'hypothèse de l'existence de plateformes de "deep labour", définies par A. Casilli et al. (2019a) comme des entités économiques dont la particularité est d'imposer "un mode d'organisation basé sur le morcellement des processus de gestion de la force de travail." (p.26) Ainsi, ces plateformes mettent en place une chaîne d'intermédiaires s'appuyant sur des plateformes de recrutement, des sous-traitants en charge de l'infrastructure technique, tout ça au service de firmes clientes (lead firms). Comme l'explique Tubaro (2021), cela consiste en : "des accords contractuels entre une entreprise principale (par exemple, Microsoft et son service propriétaire UHRS), ses fournisseurs (tels que Pactera), leurs sous-traitants (chargés, par exemple, de la maintenance de l'infrastructure technique) et tous les fournisseurs de services supplémentaires (tels que les sites web d'emploi)." Dans ce modèle, on peut considérer les plateformes comme des firmes puisque la structure économique repose sur des liens contractuels plus forts avec les travailleurs, proches d'une relation d'emploi, que dans le cas de plateformes comme Mturk¹¹³.

112. "L'analyse théorique de la firme-réseau, qui sépare l'employeur de jure de l'employeur de facto, brouille la relation d'emploi et amène certains juristes à reconsidérer cette relation de subordination professionnelle dans l'intention de redonner des droits aux employés des sous-contractants. Selon le juriste Supiot ([2002], p. 172), l'autonomie dans la subordination et l'allégeance dans l'indépendance sont les deux principales caractéristiques des firmes-réseaux." (Chassagnon, 2014)

113. "le réseau d'affiliation I-F ne comprend pas de relations de travail à proprement parler, mais il lie les travailleurs à l'entreprise ou à la plateforme qui les a sous-traités, au client final qui (espérons-le) remplit les documents fiscaux pour eux, aux sous-traitants qui fournissent les outils techniques (...) Ici, de multiples intermédiaires remplacent les relations de travail, et même les travailleurs non salariés ont accès à des services, à des outils de production ou à une supervision. Les travailleurs renoncent à une partie de la flexibilité et de l'autonomie qu'offrent les plateformes de type marchand comme Mechanical Turk, pour accepter une discipline (la quantité de travail qu'ils s'engagent à effectuer régulièrement) et parfois une supervision par d'autres." (Tubaro, 2021)

L'intérêt de la littérature sur les GCV est quoi qu'il en soit de proposer une analyse des relations inter-firmes sous l'angle des relations de pouvoirs et de la détermination de la structure de partage de la valeur et des conditions de travail par les firmes pilotes. Cet angle nous pousse également à nous intéresser aux modalités de pilotage de la chaîne de valeur dans le secteur de l'IA. Quel sens peut-on alors donner à la manière dont les entreprises externalisent ce type d'activités? Au-delà de la question de la codification des tâches évoquée lors de notre précédent chapitre, on peut également se poser la question de la firme focale dans le cas des plateformes numériques de travail. En effet, si elles se comportent en apparence comme des firme-réseaux coordonnant un réseau de sous-traitants indépendants juridiquement, deux arguments limitent la pertinence de cette analyse.

Le premier argument est de nature juridique. Si les travailleurs voient leur contrat de travail être requalifiés en contrat salariés, c'est que la plateforme concentre les pouvoirs de direction économique, d'organisation de l'activité et de sanction. Par conséquent, on ne serait plus face à un réseau de sous-traitants, mais face à une firme classique. Ainsi, partant des nombreuses décisions de justice qui considèrent que les travailleurs des plateformes sont des salariés déguisés, Uber est bien une firme qui coordonne ses propres salariés, on peut alors se demander s'il ne faudrait pas parler de relation 'intra-firme' plutôt qu'inter-firme? Dans cet optique, l'existence des plateformes apparaît finalement comme une fiction juridique inventée par des entreprises désireuses de s'exonérer des contraintes légales.

Ensuite, le donneur d'ordre final dispose d'un pouvoir d'organisation de la tâche (et non de l'activité), puisque, par exemple sur Mturk, il définit et explique ce que le travailleur doit faire. De ce point de vue, l'organisation platformisée du travail pourrait, à minima, s'apparenter au concept de "*hiérarchie électronique étendue*" (Fréry, 1997) avec, selon les cas, une division des pouvoirs disciplinaires entre le donneur d'ordre et la plateforme et une intégration logistique qui serait réalisée à travers des outils informatiques possédés par la plateforme ou par le donneur d'ordres.

Si nous considérons comme particulièrement pertinent le fait de penser les plateformes comme des "*firmes-pivots*", nous avançons qu'analyser systématiquement les plateformes du monde entier sous cet angle pose problème du point de vue de la frontière conceptuelle entre marché et firmes. Dans le cadre de la théorie des chaînes de valeur, on peut par exemple imaginer que les plateformes comme organisations intermédiaires entre firme et marché constitueraient, non pas des acteurs économiques permettant de créer de la valeur, mais des "*firmes-pilotes secondaires*" qui détermineraient les conditions d'exercices d'une activité à la place des véritables donneurs d'ordres. La question des frontières de la firme se poserait donc plutôt en matière de partage du pouvoir hiérarchique entre les acteurs, qu'en termes de positionnement de ces entités sur un continuum marché-hiérarchie. Les questions qui se posent sont alors de deux ordres : comment est organisée l'activité par les PNT? Quel est le rôle du donneur d'ordre final?

L'évolution de la nature des firmes demeure difficile à percevoir. L'externalisation, comme la plateformes sont des tendances récentes qui questionnent les outils théoriques de la sociologie économique et des sciences économiques. La question de la pertinence de la dichotomie firme-hiérarchie se pose, tout autant que celle de l'émergence d'entités économiques hybrides empruntant à l'une ou à l'autre forme de gouvernance. Nous avançons donc une proposition : la variété des formes de plateformes, de chaînes de valeur et de firme-réseaux implique également qu'il est toujours possible de classer ces entités économiques du côté du marché ou de la hiérarchie, sans les situer dans un entre-deux.

Dans ce chapitre, nous réalisons tout d'abord une typologie des activités du processus de production de l'IA en nous focalisant sur les activités nécessaires aux services basés sur l'IA, selon un axe capital / travail et donc selon les activités les plus intensives en capital et en travail. Dans un second temps, nous montrons comment certaines entreprises hégémoniques pilotent la chaîne de valeur en contrôlant certains actifs spécifiques, notamment l'hébergement de données et la puissance de calcul, ainsi que l'organisation du travail d'annotation. Nous montrons que les entreprises françaises se retrouvent alors dans une position subalterne dans la chaîne de valeur, en vendant principalement des services de développement d'IA qui s'appuient sur des modèles pré entraînés et sur l'infrastructure de ces entreprises hégémoniques (*fine-tuning*). Nous développons le cas de Google, firme hégémonique par excellence et de sa chaîne de sous-traitance en matière d'annotation de données en insistant sur la variété des modèles d'organisation du travail pour le travail des données. Enfin, nous explorons différents modèles organisationnels théoriquement plateformes.

5.1 Construire une typologie des activités de production de l'IA

Derrière la gouvernance de la chaîne de valeur, se joue également une structure de l'appropriation de la valeur économique : les firmes pilotes s'approprient et contrôlent la distribution de la valeur créée pendant le processus de production.¹¹⁴ Un des apports de la théorie des chaînes de valeur est la distinction opérée entre deux types de chaînes de coordination de la production et de gouvernance : les chaînes pilotées par les acheteurs (CPA) et les chaînes pilotées par les producteurs (CPP). Les CPP se caractérisent par des industries intensives en capital et les CPA se caractérisent par des industries ayant recours à de longues chaînes de sous-traitances (par exemple : l'habillement). Dans les CPA, un des apports de Gereffi et al. (1994) a été de montrer comment des firmes pilotes, du fait de leur "capital commercial", peuvent déterminer les conditions de production de nombreuses entreprises de la chaîne de sous-traitance sans avoir de lien capitalistiques¹¹⁵.

114. Ici, nous ne détaillons pas l'ensemble des apports de ces recherches. Ainsi, outre la description des rapports de pouvoir inter-firmes, les deux chercheurs montrent que les chaînes globales se caractérisent par : "une structure entrants-sortants, qui décrit le processus de transformation des matières premières et l'incorporation des composants dans les produits finaux", "une territorialité, ou configuration géographique", et "le contexte socio-institutionnel, qui décrit les règles du jeu s'appliquant à l'organisation et au fonctionnement des chaînes"

115. "même des chaînes caractérisées par des structures de gouvernance relativement proches de celle du marché nécessitent une coordination, et que ces tâches de coordination étaient prises en charge par des firmes qui définissaient l'essentiel de la division du

En mobilisant une base de données d'entreprises d'IA (voir : chapitre 3), ainsi que les entretiens réalisés avec les employés d'entreprises françaises d'IA et les intermédiaires d'annotation français et malgaches, nous analysons la structure de la chaîne de valeur de l'IA. Ces entretiens ont permis de donner du sens à notre base de données et de comprendre plus en détail comment les entreprises d'IA construisent et appliquent leurs modèles aux enjeux de leurs clients finaux.

5.1.1 Un continuum d'activités intensives en travail et en capital divisé entre les pays du Nord et ceux du Sud global. Modéliser d'un côté, produire des données et des matières premières de l'autre

Notre objectif étant de comprendre l'implication des entreprises d'IA dans la chaîne de valeur de cette industrie, nous avons réalisé une analyse des correspondances multiples (ACM) afin de faire une typologie des entreprises d'IA. Nous nous sommes exclusivement appuyés sur les variables qualitatives de notre base de données dans le cadre d'une méthode *data-driven*, qui nous permet de tester la typologie issue de notre *desk analysis*. Ces études de cas nous ont permis d'identifier les principales fonctions des différentes activités de la chaîne de valeur de l'IA et notamment l'hébergement des données, la puissance informatique, l'annotation (c'est-à-dire la collecte des données de l'étiquetage).

Le premier axe représente l'opposition entre les entreprises offrant une main-d'œuvre d'annotation (micro-travailleurs)¹¹⁶, y compris les entreprises dont le seul but est de fournir des annotateurs de données et/ou des outils d'annotation, et les entreprises qui offrent des services tels que des logiciels de développement de modèles, du matériel, le développement de modèles ou l'hébergement de données. Le deuxième axe permet de matérialiser l'opposition entre les entreprises qui fournissent des infrastructures numériques (matériellement parlant), notamment la puissance de calcul et l'hébergement de données, et les entreprises qui offrent soit des services logiciels (développement de modèles, outils de développement), soit des travailleurs d'annotation. L'ACM permet de projeter des observations (dans ce cas, des entreprises) sur les axes.

travail le long de la chaîne et déterminaient les conditions dans lesquelles des acteurs pouvaient y participer.” (Bair, 2010)

116. Nous avons observé une chute rapide de l'inertie relative des axes après les deuxièmes axes (27% pour le premier axe et 15,4% pour le deuxième axe), nous avons donc décidé de concentrer notre analyse sur les deux premières dimensions (Figure 9). Nous avons donc analysé ces deux axes.

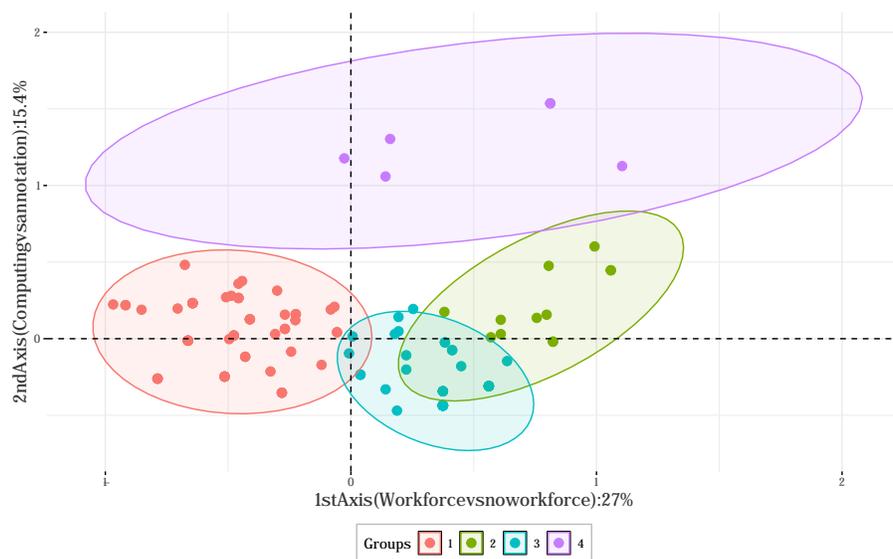


FIGURE 17 – Typologie des entreprises d’intelligence artificielle (source : desk research / Crunchbase.com)

Il est également possible de regrouper nos observations pour créer des groupes homogènes d’entreprises décrivant une typologie d’entreprises d’IA. Nous avons choisi de conserver quatre clusters. Le cluster violet regroupe les entreprises *“full stack”* offrant tous les services. Ce sont de grandes entreprises numériques comme IBM ou Amazon. Le cluster vert est composé d’entreprises offrant de la puissance de calcul, notamment sous forme de *hardware* (par exemple, un processeur spécialisé en IA d’une filiale d’Intel) et de sociétés d’optimisation de modèles. Le cluster bleu représente les entreprises de *“software et de développement”* d’IA qui proposent soit des outils de développement (Rstudio, Matlab) soit des modèles préentraînés qui vont être adaptés à leurs clients par la suite. Le cluster rouge représente les sociétés d’annotation qui proposent deux types de services associés à l’IA, des outils d’annotation (logiciels) et une main-d’œuvre d’annotateurs. Cette classification nous permet d’identifier quatre groupes homogènes d’entreprises qui remplissent différentes fonctions dans la chaîne de valeur de l’IA. Nous pouvons souligner la forte démarcation entre les entreprises offrant des services tels que la puissance de calcul et l’hébergement de données, à forte intensité de capital et les entreprises offrant des services d’annotation, à forte intensité de main-d’œuvre. L’un de ces quatre groupe se distingue par la présence de grandes entreprises du numérique, capables d’offrir à la fois des infrastructures de calcul ou d’hébergement et des services d’annotation (Amazon ou Alibaba), tout en sachant que des entreprises du Cluster 1 telles qu’IBM ou Google, n’ont pas forcément de travailleurs des données, mais font appel à des sous-traitants intégrés à leur service pour l’annotation.

Dans un article, Crawford et Paglen (2019) proposent une visualisation de toutes les opérations d’extraction nécessaires au fonctionnement d’Alexa, l’assistant vocal d’Amazon. Ils montrent ainsi que les différentes étapes de la production impliquent trois types de processus d’extraction : ” les ressources matérielles, le

travail humain et les données”. Ils montrent que : *”La structure requise pour interagir avec un Amazon Echo va bien au-delà de la ”structure technique” multicouche de la modélisation des données, du matériel, des serveurs et des réseaux. La structure complète s’étend bien au-delà du capital, de la main-d’œuvre et de la nature, et exige une quantité énorme de chacun d’entre eux. ”*

En utilisant une perspective plus générale, nous montrons comment cette industrie supposée destructrice d’emplois déplace en fait le travail en dehors des frontières des entreprises traditionnelles, qu’il s’agisse d’activités minières permettant de fabriquer les centres de données, la puissance de calcul ou les activités d’annotation. Pour les premières, Kate Crawford et Vladan Joler précisent d’ailleurs qu’*”Intel possède sa propre chaîne d’approvisionnement à plusieurs niveaux, composée de plus de 19 000 fournisseurs dans plus de 100 pays, qui lui fournissent des matériaux directs pour ses processus de production, des outils et des machines pour ses usines, ainsi que des services de logistique et d’emballage.”*

Dans notre typologie, nous distinguons les entreprises qui ont une forte valeur en capital, puisqu’elles fournissent de l’hébergement de données ou de la puissance de calcul des entreprises qui fournissent une valeur en travail cognitif immatériel, puisqu’elles fournissent de la main-d’œuvre pour l’étiquetage des données aux entreprises d’IA ¹¹⁷. Les premières sont des entreprises telles qu’Amazon Web Services, qui dispose de 175 centres de données dans le monde, ou NTT Communications, qui a 48 centres de données dans le monde. On trouve également des entreprises comme Nvidia, qui fournissent de la puissance de calcul matérielle (*hardware*).

Les entreprises comme Amazon Web Services (AWS) et Google AI constituent des entreprises hégémoniques, par leur capacité à structurer l’industrie de l’IA en maîtrisant toutes les étapes de la production, y compris la partie annotation qui est également largement encadrée par ces entreprises. Ayhan Aytes (2012) ou Lily Irani (2015b) expliquent qu’à l’origine, Mturk répondait au besoin de nettoyer les doublons du catalogue de l’entreprise qui ne trouvaient pas de solution technique automatique. Amazon a donc développé une plateforme interne permettant à ses employés d’effectuer ce travail. Après avoir été ouvert à la clientèle, cet outil a largement été utilisé pour construire les bases de données d’entraînement des algorithmes telles que la célèbre base de données ImageNet. Au-delà du rôle d’intermédiaire, l’intérêt de ce modèle est la gestion automatisée du travail via une interface de programmation d’applications (API). À cette époque, AMT a été intégré à une myriade de services au sein d’Amazon Web Services (créé en 2002) comme l’hébergement de données.

Le 29 novembre 2017, Amazon a lancé Amazon SageMaker qui, comme l’indique la page Wikipedia dédiée *”permet aux développeurs de créer, d’entraîner et de déployer des modèles d’apprentissage automatique (ML) dans le cloud”*. Dans le cadre de ce nouveau service d’apprentissage automatique, Amazon a créé Amazon

117. Même si l’extraction minière nécessite également de nombreux travailleurs

SageMaker Ground Truth, qui est un *”service d’étiquetage de données entièrement géré qui facilite la création d’ensembles de données d’entraînement très précis pour l’apprentissage automatique”* et fournit un accès à un outil d’étiquetage et, par le biais de la *”place de marché AWS”*, à des sociétés de services d’étiquetage de données telles que le sous-traitant C, IMerit ou le sous-traitant E qui sont *”sécurisées”*. Comme le dit l’un de leurs fournisseurs de services à propos de la place de marché AWS : *”ils viennent nous voir pour un audit chaque année”*¹¹⁸. Ce faisant, Amazon est en mesure de se positionner comme un fournisseur de services d’IA pour toutes les activités nécessaires au développement et à l’exploitation d’un modèle d’IA. Que ce soit à travers ses propres installations ou avec l’intégration de partenaires dans leur place de marché (par exemple : pour l’annotation de données).

Les secondes sont des sociétés d’annotation qui fournissent des outils d’annotation permettant aux clients d’étiqueter les données en interne (en utilisant leur propre personnel) ou en mettant à disposition un personnel pour le faire. Il existe également des sociétés qui fournissent la main-d’œuvre pour l’annotation (Teleperformance, Appen ou Samasource).

Parlons maintenant du cas d’Automatik, une entreprise française classée dans le troisième cluster et a une position particulière dans la chaîne de valeur. Même si elle dispose de compétences techniques, elle s’occupe principalement d’adapter les ressources des grandes entreprises technologiques aux besoins de leurs clients, en utilisant des architectures d’IA existantes, hébergés sur les serveurs des grandes entreprises technologie. De plus, elle dispose de ses propres ressources d’annotation avec un sous-traitant de longue date, totalement intégré dans leur processus et leurs outils techniques et dirigé par un *”dataset manager”*.

Dans un entretien avec un ingénieur d’Automatik interviewé plusieurs fois au cours de notre enquête, il soulève l’utilisation des réseaux neuronaux créés par les grandes entreprises technologiques : *”Ce qu’on fait, c’est qu’on doit les implémenter forcément avec notre système interne, mais ici on reste sur des modèles open source, et notamment on n’a que des réseaux enfin, 90% de nos réseaux neuronaux sont des réseaux neuronaux Tensorflow de Google.”* Il précise également que l’hébergement des modèles et des données se fait principalement sur les serveurs de ces entreprises : *”(...) la réalité est que nos clients n’ont pas un grand département informatique, et donc [il] est plus ou moins soumis, il n’a pas son mot à dire si c’est Amazon, Google ou Microsoft.”* Les grandes entreprises du numérique sont donc en position hégémonique sur les segments de la modélisation et de l’hébergement de données. Les entreprises comme Automatik qui se positionne sur le développement de modèles se retrouvent donc dans une position subalterne. Ces entreprises n’ont pas la puissance des firmes-pivots, car elles ne sont pas actives sur des segments de marché à forte intensité capitalistique (puissance de calcul...), mais elles parviennent tout de même à se positionner comme intermédiaires de l’IA en vendant la force de travail de leurs data scientists et de leurs travailleurs des données.

118. Comme le dit l’un de ces fournisseurs : *”Nous avons été référencés sur la place de marché AWS, parce que (...) les équipes d’AWS cherchaient à étendre leur réseau de fournisseurs (...) et c’est ainsi que nous avons été approuvés par Amazon.”*

Comme indiqué précédemment, nous avons identifié l'origine géographique des entreprises d'IA ainsi que la localisation des travailleurs de l'IA. Le résultat nous permet de voir que les entreprises focalisées sur des activités à haute valeur ajoutée se situent dans les pays occidentaux alors que les micro-travailleurs se situent dans les pays en voie de développement. La carte présentée ci-dessous (Figure 2.) est une projection géographique du siège social des entreprises d'IA proposant tous les services à l'exception de ceux impliquant une force de travailleurs de la donnée. on y constate la prédominance de pays tels que les USA, la Grande-Bretagne, le Canada, l'Allemagne, l'Inde et le Japon. Ces résultats sont cohérents avec ceux présentés précédemment et dessinent un paysage de l'intelligence artificiel dominé par les pays les plus développés du Nord.

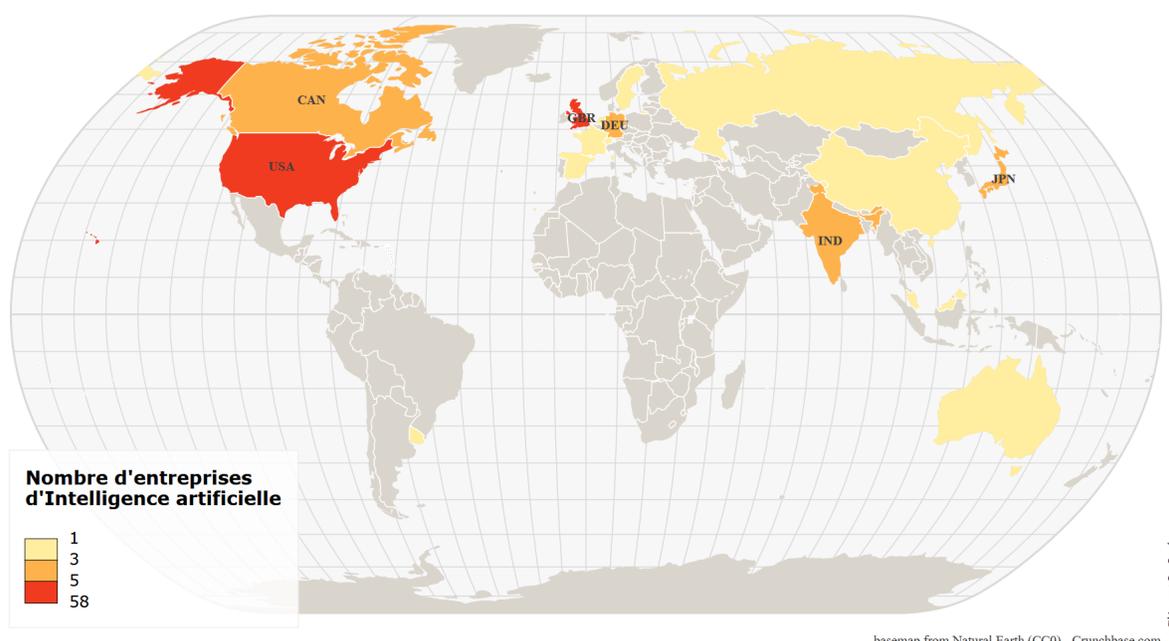


FIGURE 18 – Localisation des entreprises d'IA (source : desk research / Crunchbase.com)

De l'autre côté de la chaîne de valeur, les services d'annotation sont d'autres services que les entreprises du cluster 4 ("Full-stack") et 3 ("développement") encadrent largement. En effet, comme le souligne un salarié d'une entreprise de reconnaissance d'images : *"Le gros problème de notre entreprise, ce sont les données. Lorsque vous venez développer une solution pour certains clients, vous ne pouvez pas [toujours] exploiter leurs données, vous devez donc trouver des données pour alimenter nos algorithmes"*. Il faut pour cela avoir accès à des sous-traitants en matière d'annotation de données.

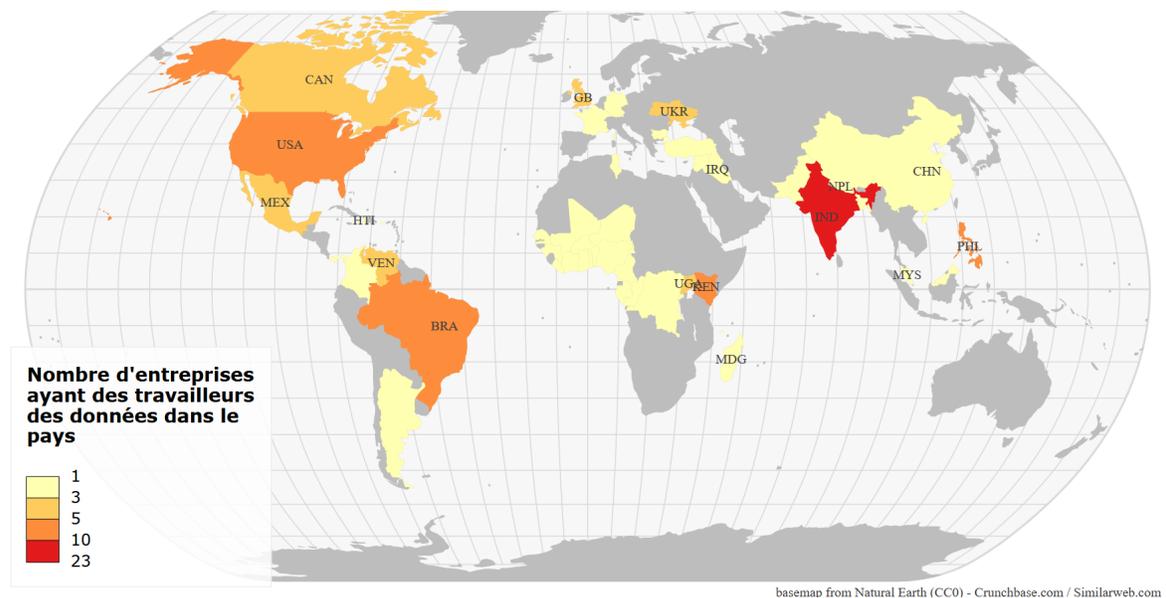


FIGURE 19 – Localisation des entreprises de travail des données (source : desk research / Crunchbase.com)

Si l'on regarde les entreprises ayant recours à des travailleurs des données, on constate qu'une autre géographie de l'IA se dessine. Dans ce cas, nous avons compté la localisation des travailleurs telle qu'elle est déclarée par les entreprises, notamment sur leurs sites Internet. Par exemple, lorsqu'une entreprise déclare avec des travailleurs des données au Kenya, nous avons compté une localisation de travailleur dans notre base de données.

Nous constatons que le pays ayant le plus d'entreprises qui possèdent le plus de travailleurs est l'Inde (23). Suivent des pays comme les USA (8), les Philippines (7), le Kenya (5) et le Brésil (5). Un troisième cercle de pays est composé du Venezuela (4), de l'Ouganda (3), de l'Ukraine (3), du Mexique (3), du Canada (3) et de la Grande-Bretagne (3). Enfin, des pays comme la Chine, la Malaisie, le Népal, Madagascar ou Haïti (2) sont également des lieux d'externalisation du travail des données. Il faut certes tenir compte du manque de représentativité de notre échantillon et les biais culturels associés. On compte par exemple très peu d'occurrences pour un pays comme la Chine, ce qui nécessiterait d'élargir cette base de données.

Cette première section nous a permis d'aborder trois sujets cruciaux des travaux sur les chaînes de valeur : les activités qui les composent, leurs localisations et leurs structures en matière de pilotage (Kano, Tsang, & Yeung, 2020). Concernant le premier point, l'apport de notre travail est d'intégrer les services d'annotations, trop souvent oublié des études sur la chaîne de valeur de l'IA (Attard-Frost & Widder, 2023). Par ailleurs, nous constatons également une différence importante de positionnement entre les services hégémoniques qui proposent l'ensemble des services de la chaîne et les entreprises d'IA françaises qui se positionnent sur des créneaux très précis de la chaîne de valeur, ce qui à nouveau pose la question de qui pilote la firme. Ainsi,

la production de l'IA est largement "contrôlée" par les grandes entreprises du numérique, qui se distinguent par leur capacité à fournir l'ensemble des services nécessaires au reste des acteurs de la chaîne et, si besoin, en intégrant à leurs services les prestataires nécessaires à la production de services basés sur l'IA. Dans ce qui suit, nous nous concentrerons plus particulièrement sur les activités à forte intensité de main d'œuvre : les services d'annotation.

Concernant la localisation, nos résultats sont cohérents avec la littérature sur la géographie des plateformes numériques de travail (Graham, Sabbata, & Zook, 2015), ainsi qu'avec les conclusions de Berg et al. (2018) à propos des plateformes numériques de micro-travail. On constate ainsi que la participation des pays des Suds, ainsi que de certains pays du Nord à bas revenus, comme l'Ukraine, se matérialise par l'hébergement de la force de travail qui réalise les tâches humaines nécessaires au fonctionnement des IA produites dans les pays du Nord. La limite principale de cette analyse géographique est néanmoins qu'elle ne se base que sur des données accessibles en ligne.

5.1.2 La variété des organisations qui fabriquent les données pour l'IA : le cas de la chaîne de sous-traitance de Google

L'analyse de cette base de données, ainsi que les résultats présentés dans notre chapitre précédent nous permettent ensuite de postuler de la variété des modèles organisationnels du travail de la donnée. Si nous nous concentrons sur les trois plus grandes entreprises de notre échantillon et de nos études de cas, nous observons qu'elles ont toutes des modes d'externalisation différents. Tout d'abord, Amazon propose sa propre plateforme de micro-travail, Amazon Mechanical Turk, et un service d'annotation réalisé par des prestataires via une "marketplace" d'entreprises de types BPO. Dans le cas de Google, le travail humain est externalisé, puisque l'entreprise propose à ses clients d'utiliser les services de ses partenaires, GlobalLogic et Teleperformance. Dans ce cas, il s'agit d'un modèle d'externalisation basé sur des centres de services qui sont plus proches des centres d'appels (modèle BPO). Enfin, le cas d'IBM se différencie par la redirection de ses clients vers les services d'Appen et de DefinedCrowd, entreprises extérieures au groupe IBM et qui sont toutes deux considérées comme des plateformes de "deep labour" dans la littérature (Tubaro, 2021).

Nous avons analysé l'exemple de Google plus en profondeur, car l'entreprise californienne a la particularité de déclarer publiquement ses sous-traitants en matière d'annotation de données. Ces dernières sont divisées en deux catégories, les sous-traitants directs et les entreprises affiliées à un sous-traitant de Google. Nous pouvons également les distinguer en fonction de leurs activités. Par exemple, Mediscribe, Teleperformance et Global Logic fournissent des services d'étiquetage des données, tandis que Lionbridge¹¹⁹, Appen, Sigma, E Data Solutions et Productive Playhouse fournissent des services de transcription de la voix. Une dernière façon de catégoriser les sous-traitants de Google est de distinguer les entreprises qui emploient des salariés

119. Lionbridge AI a été racheté par Telus

(BPO) et celles qui emploient des entrepreneurs numériques (freelances), c'est-à-dire des plateformes telles qu'Appen et Lionbridge.

Company name	Type of suppliers	Type of activity	Type of contract	Country of operation
Productive Playhouse	Direct	Audio Transcription	Salaried employees	Global
Sigma	Direct	Audio Transcription	Salaried employees	ESP
Global Logic	Indirect	Data labelling	Salaried employees	IND
Téléperformance	Indirect	Data labelling	Salaried employees	IND
Médiscribe	Indirect	Data labelling	Salaried employees	IND
E Data Solutions	Indirect	Audio Transcription	Salaried employees	KEN
Appen	Direct	Audio Transcription	Self employee and BPO companies	Global
Lionbridge	Direct	Audio Transcription	Self employee	Global

TABLE 4 – Classification des sous-traitants de Google pour les données

Dans cette optique, il est important de considérer que les chaînes de valeur de Lionbridge et Appen s'appuient, au moins en partie, sur des travailleurs indépendants situés dans le monde entier. Par exemple, Appen déclare avoir une main d'œuvre d'un million de travailleurs dans 130 pays et 180 langues. Un troisième niveau de classification apparaît alors et concerne la localisation ou non de la force de travail. Outre ces différences entre les entreprises, nous avons constaté que certaines entreprises d'annotation proposent des forces de travail différenciées, en fonction de critères tels que la complexité des tâches, la gestion du temps et/ou la sécurité des données. Appen, par exemple, propose les services suivants :

- une "foule à la demande", c'est-à-dire des annotateurs disponibles 24h/24 et 7j/7 pour des micro-tâches de faible intensité;
- une "foule à distance", c'est-à-dire des annotateurs qualifiés sélectionnés en fonction des besoins du client;
- une "foule sécurisée", c'est-à-dire des annotateurs certifiés et "identifiés" travaillant pour Appen depuis des "locaux sécurisés";
- une "foule sur site", c'est-à-dire des annotateurs qualifiés qui travaillent dans les locaux du client, mais qui sont gérés par Appen;
- les "ressources internationales du client", qui sont les employés des clients utilisant la plateforme Appen.

Cette étude de cas montre comment une grande entreprise d'IA peut mettre en place une chaîne de valeur autour d'une variété d'entreprises sous-traitantes et d'une variété de travailleurs. Par exemple, Google est connu pour utiliser la main-d'œuvre gratuite de ses utilisateurs pour accomplir certaines tâches¹²⁰. À cette main-d'œuvre gratuite s'ajoute le travail intermédié des plateformes, en l'occurrence Appen et Lionbridge,

120. Comme l'illustrent les Captchas déjà évoqués plus tôt dans cette thèse

ainsi que le travail des données effectué dans des entreprises plus traditionnelles situées dans le secteur de la sous-traitance de services. L'entreprise américaine a en effet recours à des centres de services situés en Inde pour ses activités d'étiquetage de données. C'est le cas de l'entreprise Teleperformance qui fournit historiquement des services de relation client par téléphone. Dans le cas des transcriptions sonores utilisées pour entraîner son service d'assistant vocal "Ok Google", on constate que Google utilise à la fois des centres de services (E Data Solutions au Kenya) et des plateformes de travail numérique (Appen et Lionbridge). Cela peut s'expliquer par la nécessité de d'adapter son produit à différents contextes culturels, et donc de s'appuyer sur une main-d'œuvre mondialisée ce qui est rendu possible par les plateformes numériques de travail. En effet, ces produits sont disponibles dans plusieurs langues, ce qui nécessite d'avoir accès à une variété de locuteurs, justifiant le recours à une plateforme qui dispose de travailleurs dans le monde entier.

Cette diversité de formes de sous-traitance se retrouve à l'autre bout de la chaîne, chez les entreprises d'annotation de données. Dans le cadre de notre terrain à Madagascar, nous avons ainsi pu interviewer des travailleurs et des cadres dirigeants d'entreprises d'annotations dont l'accès aux clients repose dans certains cas sur des plateformes. C'est le cas pour le sous-traitant E, une entreprise de plus de 2000 salariés qui travaille à la fois directement avec certains clients, mais qui a également des partenariats avec plusieurs plateformes numériques de travail. Cette entreprise située fournit ainsi des prestations de "foule sécurisée", qui sont en fait, au moins en partie, des travailleurs employés par des entreprises de *business process outsourcing* de Madagascar qui travaillent au sein de leurs bureaux situés dans la capitale malgache.

Encadré n°30 : Travailler avec comme sous-traitant des plateformes

R : Après, on a aussi des boîtes qui font principalement du crowdsourcing. Il y a Appen, qui est très gros, Lionbridge aussi, qui sont plutôt sur du langage et du NLP, mais il faut se méfier un peu. On a des gens comme Transperfect qui sont un peu des concurrents, mais ils bossent un peu pour nous. Pareil pour Labelbox, on bosse avec eux.

Q : Et avec Appen aussi vous bossez, non ?

R : Oui, on bosse avec eux. Ça pourrait être nos concurrents, mais en même temps l'environnement est tellement vaste qu'on ne se marche pas vraiment dessus.

Dinasoa - CEO - Sous-traitant E

Grâce à nos études de cas, nous avons identifié quatre catégories de fournisseurs associées à différents types de main-d'œuvre. La main-d'œuvre globale, qui rassemble des annotateurs situés dans le monde entier par le biais d'une plateforme dans un mode d'externalisation *orienté plateforme* (c'est-à-dire Appen ou Amazon Mechanical Turk). La main-d'œuvre localisée, qui rassemble des annotateurs d'une zone géographique spécifique et qui peut passer par un fournisseur de centre de services qui correspond au mode

d'externalisation *orienté BPO* (i.e. Téléperformance). Certains clients peuvent accéder à cette main-d'œuvre par le biais de contrats "tout compris" avec de grands fournisseurs d'IA (comme AWS) ou par l'intermédiaire d'une place de marché d'IA. La main-d'œuvre est parfois interne. Dans ce cas, il s'agit d'employés de l'entreprise donneuse d'ordres qui réalisent eux-mêmes l'annotation depuis ses propres locaux en utilisant des interfaces développées par des entreprises qui proposent des outils d'annotation. Ce modèle correspond à l'*externalisation des outils d'annotation*.

Certains fournisseurs se situent, en quelque sorte, entre ces différents types d'externalisation. En effet, ce sont des entreprises de BPO qui passent par la plateforme pour trouver des clients sans même les connaître, ce qui est cohérent avec le phénomène de ré-intermédiation déjà observé sur les plateformes de freelance comme Upwork. D'une certaine manière, cela correspond à un mode d'externalisation *platformized BPO*. Le tableau ci-dessous synthétise cette typologie.

Type d'externalisation	Type de force de travail	Statut légal des travailleurs	Infrastructure d'accès au travail	Caractéristiques
Orienté plateforme	Globale	"Indépendants"	Plateforme en ligne	
Orienté sous traitants	Localisé	"Employés"	Contrats de prestation	Chaîne de production sécurisée
Orienté outils d'annotation	Internalisée	"Employés"	Intégrés dans d'autres services	Outils d'annotations
Sous traitants platformisés	Localisé	"Employés"	Plateformes de "deep labour" et d'annotation (Appen, Labelbox)	Chaîne de production sécurisée

TABLE 5 – Tableau de synthèse des modes d'externalisation de l'annotation de données

* * *

Dans cette partie, nous avons d'abord présenté les différentes étapes du processus productif de l'intelligence artificielle à travers une typologie des entreprises d'IA, avant d'analyser spécifiquement les services d'annotations et leur place dans la chaîne de valeur. Nous abordons trois thèmes importants des études sur les chaînes de valeur : les activités qui la composent, sa localisation et la structure de la chaîne telle que pilotée par (Kano et al., 2020). L'apport de notre travail est d'intégrer des services d'annotation, trop souvent négligés dans la recherche en IA. Par ailleurs, alors que la littérature sur le travail de données s'est principalement focalisée sur des plateformes anglophones ainsi que sur les jeux de données utilisés dans la recherche scientifique, ces premiers résultats permettent d'identifier des différences importantes notamment concernant la participation d'acteurs d'entreprises sous-traitance formelle et pas uniquement de plateformes.

En matière de localisation, nos résultats sont cohérents avec la littérature sur la géographie des plateformes

numériques de travail (Berg et al., 2018; Graham et al., 2015). De manière schématique, les activités intensives en capital (hébergement de données) et intensives en qualification universitaire (modélisation) sont produites les pays du Nord. En revanche, les activités intensives en travail (production de données), sont réalisées dans pays à faible revenu en voie de développement.

En matière de pilotage de la chaîne de valeur, on observe également une différence significative entre les services hégémoniques proposant tous les services de la chaîne et les entreprises françaises de l'IA se positionnant sur des niches bien précises de la chaîne de valeur (modélisation). Nos entretiens laissent penser que ces petites entreprises d'IA sont en réalité des "intégrateurs" de solutions développées par les grandes entreprises du numérique : elles accèdent à de grands modèles neuronaux qu'elles vont adapter à des applications métiers spécifiques. Ces entreprises proposent finalement une "last-mile automation".

La production d'intelligence artificielle a été largement portée par les grandes entreprises du numérique, qui se distinguent par le fait qu'elles peuvent offrir tous les services nécessaires aux autres acteurs de la chaîne et, le cas échéant, intégrer les prestataires d'annotation dans leurs services. Cela pose dès lors les différences entre les grandes entreprises qui produisent l'IA et les "intégrateurs" de solution que nous avons rencontrés en France en matière d'accès à des services d'annotation ? En utilisant le cas de Google, nous avons confirmé l'hypothèse déjà partiellement présente dans la littérature sur les sous-traitants directs (Lehdonvirta et al., 2019; F. A. Schmidt, 2019) et les plateformes qui ne fonctionnent pas sur un modèle de marché "pur" comme Amazon Mechanical Turk. Comment expliquer cette tendance, également visible en dehors du cas de la France ? En particulier, quelles sont les implications de cette évolution ? Après avoir largement évoqué les grandes entreprises du numérique et les entreprises françaises, on pourra se demander si le modèle hégémonique d'organisation du travail des données n'est plus proche d'une forme hiérarchique pour les raisons évoquées dans notre précédent chapitre. Dans notre prochaine section, nous approfondissons cette hypothèse à travers le positionnement des entreprises de sous-traitance malgaches qui sont présentes sur des plateformes qui se présentent en apparence comme des "marketplaces"

5.2 Approfondir l'hypothèse du "travail profond" à travers la comparaison des modèles d'externalisation : une organisation en apparence basée sur le marché qui a pourtant tous les traits de la "hiérarchie électronique"

Le cas de Google, traité dans le 2.1, nous a permis de constater la diversité des sous-traitants et des modes d'organisations de ces derniers. Puis, par le biais de l'étude des relations de sous-traitance directes, nous avons commencé à initier un travail analytique autour du concept de "hiérarchie électronique étendue" propre à inclure plus systématiquement différentes formes de relations contractuelles entre les firmes de la chaîne de production de l'IA. L'enjeu de cette partie est de montrer comment les travailleurs sont intégrés dans

des chaînes logistiques numériques pour produire des données à travers plusieurs types d'organisation du travail qui, en apparence, se rapprochent du "marché pur".

Dans la section qui suit, nous nous intéressons aux mécanismes d'accès à des forces de travail stables et sécurisées par les entreprises qui vendent des outils et de la force de travail pour l'annotation (Cluster 4). Si ce modèle n'est pas celui que nous avons rencontré majoritairement à Madagascar, un nombre non négligeable de travailleurs et d'entreprises sous-traitantes avaient déjà eu recours à ce type de plateforme. En nous appuyant sur nos entretiens, nous souhaitons approfondir notre typologie, en montrant comment s'organisent les relations inter-firmes et quelle y est la place des travailleurs.

5.2.1 Les plateformes d'annotations : une version "relationnelle" des plateformes de type marché "pur"

Afin d'évaluer la pertinence de notre approche conceptuelle sur l'existence des "hiérarchies électroniques", nous proposons dans cette première sous-section d'analyser la plateforme Amazon Mechanical Turk dans sa version "relationnelle", c'est-à-dire à travers un mode d'organisation du travail qui permet des échanges entre les travailleurs et les donneurs d'ordres. Pour ce faire, nous utilisons le cas d'entreprises de sous-traitance malgaches qui interviennent également pour le marché américain à travers l'utilisation de cette plateforme comme outil d'annotation et non comme place de marché.

Le dirigeant d'une grande entreprise d'annotation résume sa position sur la chaîne de valeur en tant qu'entreprise spécialisée dans le traitement des données pour l'intelligence artificielle. Cette dernière *"se découpe un peu de la manière suivante : nous, on est très en amont du process, on annote les données d'input qui vont venir éduquer les modèles, sauf que pour faire de l'annotation [inaudible], il faut faire ça sur des outils qui sont des plateformes d'annotation. Historiquement, en travaillant avec Amazon, ils ont leur propre plateforme, le MTurk, donc on bossait dessus."* (Patrick - CEO - Sous-traitant E). Cette entreprise travaille historiquement pour les services d'IA internes à Amazon via Amazon Mechanical Turk et utilise la "plateforme" dans un mode fermé aux annotateurs de la version ouverte. En effet, l'entreprise américaine maîtrisant toute la chaîne de production de l'IA, son besoin se concentre uniquement sur l'accès à des travailleurs des données ¹²¹.

En plus de l'annotation, la co-directrice du sous-traitant E indique fournir des conseils à Amazon sur la manière d'annoter et sur les consignes à appliquer à des projets d'annotation ¹²². Par ailleurs, on peut s'éton-

121. *"Amazon typiquement, ils s'en foutent, parce qu'ils ont déjà une plateforme qui s'appelle MTurk, ils ont déjà des serveurs via AWS, ils veulent juste un plug-and-play avec des annotateurs. Et c'est ce qu'on leur fournit."* (Dinasoa - CEO - Sous-traitant E)

122. *"Ce qu'on a commencé à leur fournir en plus et ils ne s'y attendaient pas, c'est toute la partie feedback, conseils ; on les aide maintenant à écrire les consignes des projets, mais au début c'était vraiment plug-and-play. Vous avez des gens ? Oui On veut les faire bosser."* (Dinasoa - CEO - Sous-traitant E)

ner du choix de l'entreprise américaine de ne pas avoir directement recours à leur plateforme d'annotation en mode ouvert. Ici, Dinaso analyse cette tendance en opposant les avantages du modèle "platform-as-market" (mode ouvert) et "platform-as-firm" (mode fermé). Dans le premier cas, les équipes d'Amazon ont fait face à deux problèmes associés au recours à des travailleurs indépendants : *"MTurk avait été créé au début parce que les équipes voulaient avoir accès à des annotateurs un peu n'importe où et n'importe quand. Ils se sont rendu compte que sur certains projets, il y avait deux problèmes (qui sont les mêmes problèmes partout). Le premier, c'est qu'ils ne savaient pas qui faisait les annotations donc il n'avait pas une garantie de volume. Ensuite, ils n'avaient pas une garantie de qualité et d'homogénéité du jugement (consistency en anglais), parce qu'un gars au fin fond du Midwest versus une mère de famille en Alaska n'a pas la même façon de voir les choses."* (Dinaso - CEO - Sous-traitant E)

Ce modèle s'avère finalement limité à des *"tâches très basiques"* du fait du manque de contrôle permis par le modèle "platform-as-market". Dans ce dernier, les entreprises donneuses d'ordre transmettent une consigne une seule fois à des milliers de personnes. Si les tâches sont mal faites, elles pourront décider de ne pas les payer. Le contrôle se fait donc par un mécanisme d'incitation économique, par le paiement ou non de la tâche. De leur côté, les travailleurs n'ont aucun moyen de contacter le client final ou la plateforme pour obtenir des explications ou des consignes supplémentaires sur la tâche à accomplir. Enfin, ils ont aussi le droit de refuser les tâches qui leur semblent mal payées ou bien qui proviennent de donneurs d'ordres connus pour ne pas valider les tâches. Deux problèmes particulièrement saillants se posent alors pour le développement de l'IA : un problème de volume et un problème de qualité (homogénéité du jugement).

Dans le modèle "platform-as-firm", les travailleurs utilisent Mturk comme un outil d'annotation, mais sont salariés du sous-traitant E à Madagascar. Ils ont une chaîne hiérarchique (chef d'équipe ou chef de projet) qui gère la production au quotidien et assure l'interface avec le client. En ce qui concerne la production, des employés chez le client final font de la vérification des données produites (contrôle qualité) et donnent des conseils aux annotateurs lorsque ces derniers ont des doutes. En matière de relations avec le client final, des employés organisent des réunions, parfois pour discuter de certaines consignes, mais également pour proposer d'autres pistes de classification de données.

Encadré n°31 : Réaliser des tâches sur la version relationnelle de Mturk 1

R : Ils ont utilisé MTurk, mais ils nous ont créé des statuts particuliers. Sur MTurk, on n'existe pas, vous ne pouvez pas nous trouver ; c'est juste les équipes qui nous ont approuvées. On a créé des identifiants. Nos équipes ne sont pas payées quand elles travaillent sur MTurk parce que ça passe par nous. Et en fait, on est un prestataire indépendant qui travaille via MTurk. (...) On n'a pas de plateforme. On a des outils parce que tout ce qui est contrôle qualité est fait sur des outils qui répliquent exactement la tâche telle que la voit l'annotateur. On a des outils, mais on ne les vend

pas, ça fait partie de la prestation.

Dinasoa - CEO - Sous-traitant E

À l'autre bout de la chaîne, les travailleurs sont donc intégrés à l'outil d'annotation d'Amazon avec des tâches spécialement conçues pour n'être accessibles qu'aux travailleurs de cette entreprise. Au cours de l'enquête, plusieurs travailleurs expliquent avoir réalisé des tâches directement sur l'interface Mturk, pour plusieurs entreprises de Madagascar ¹²³. Un aspect particulièrement marquant de ce mode d'externalisation est que les travailleurs ont des consignes très strictes concernant les informations présentes sur la plateforme. Ainsi, ils ne doivent pas prendre de photos des données ou de l'interface. Cette dimension liée à la sécurité des données est particulièrement importante, au point que, lors d'une visite dans ses locaux, le vice-président d'une entreprise malgache indique en passant devant la salle de travail d'Amazon que nous n'avons pas le droit de voir, car même lui n'a pas le "droit de rentrer" (Louis - CEO - Sous-traitant C).

Encadré n°32 : Réaliser des tâches sur la version relationnelle de Mturk 2

Z : et comment vous les trouvez les tâches ?

H : c'est directement sur l'ordi. Il y a des mots de passe (...) quand on est logué, il faut prendre ... il y a des trucs à entrer, comme quoi ... il y a beaucoup de travail sur M Turk, et on travaille sur les A9

Z : donc les tâches ... Vous tapez A9 dans la barre de recherche ?

H : oui(...)

Z : et il y a d'autres projets d'image, ou il y a juste ces deux-là ?

H : il y a aussi ... il y a ce projet que on a pas eu, mais juste des tests à faire ... parce que on avait passé le test, et l'idée c'est de compter les personnes dans la rue et tout ça (...) il y avait des images de personnes, il faut compter (...) il faut sélectionner avec la souris

Z : (...) d'accord, donc ils ne veulent pas que vous sortiez des informations de la plateforme c'est ça ?

H : oui

Z : et vous vous souvenez du nom du projet ?(...)

H : oui, AWS ! vous connaissez ?

Holy - Annotateur - Sous-traitant C

Nous avons déjà évoqué le cas de Amazon GroundTruth, un service d'étiquetage de données parallèle à Amazon Mechanical Turk qui permet mettre en relation des entreprises d'annotations de données à avec

123. "Q : qui ne s'appellerait pas mechanical Turk ? R : oui c'est ça ! Q : et dessus vous vous identifiez, vous arrivez (...) comment est-ce que vous accédez à l'endroit où vous devez travailler ? R : non c'est juste une tâche(...) il y a des logos donnés par Amazon, et il y a des produits sur l'image. C'est nous qui déterminons que dans les produits donnés il y a des logos d'Amazon" (Anita - Cheffe de projet - Sous-traitant G)

des entreprises clients. C'est le cas du sous-traitant E, qui, après avoir travaillé pendant plusieurs années pour l'entreprise américaine a été un des premiers prestataires à rejoindre cette "marketplace". Cette place de marché leur permet d'être visible pour des clients qui passent par Amazon AWS¹²⁴ pour des services liés à leur modèle (modèle pré-entraîné, hébergement...) et qui souhaitent également faire annoter des données. La contrepartie est qu'ils doivent se soumettre à des audits de sécurité qui ont lieu tous les ans.

Encadré n°33 : Réaliser des tâches sur la version relationnelle de Mturk 3

R : Il faut dissocier Amazon pour qui on travaille de manière historique pour leurs produits internes en Californie, en Inde, etc. et AWS qui vend des services d'algorithmes et d'annotations via des partenaires comme nous. On a été référencé, je crois, en 2018 sur la marketplace AWS pour tout ce qui est service d'annotation parce que les équipes de Californie nous ont branchés avec les gens d'AWS qui cherchaient à étoffer leur réseau de prestataires ou de partenaires. C'est comme ça que l'on est devenu membre de la marketplace AWS et qu'on a été approuvé par Amazon.

Q : Avec tous ces audits sécurité aussi du coup.

R : Tout à fait.

Patrick - CEO - Sous-traitant E

D'après Loik, directeur des opérations du sous-traitant E, la différence réside dans la proposition de valeur des deux modèles, Mturk étant uniquement centré sur la labellisation, alors que GroundTruth permet d'accéder à l'ensemble des services d'IA (hébergement de données...) : *"MTurk permet de faire du crowdsourcing pour la labélisation. Bien sûr, il y a un système de qualification pour s'assurer qu'uniquement certaines personnes, dépendant de la complexité de la tâche, puissent y accéder. Mais ça donne accès à du crowd¹²⁵. GroundTruth par contre, donc SageMaker, c'est une offre complètement intégrée. On peut faire uniquement l'annotation, mais on peut aussi derrière héberger le modèle. Plutôt que d'avoir simplement l'annotation, d'avoir une annotation assistée (human in the loop). L'annotation peut être faite automatiquement, mais un certain pourcentage est envoyé à un annotateur humain pour assurer que la labélisation faite automatiquement est correcte et reste dans certains paramètres, et bien sûr cela peut être intégré à différents services derrière."* (Loik - COO - Sous-traitant E)

De l'aveu même de Loik, : *"Pour la partie IA, c'était zéro client français. On a eu des clients français à travers AWS, mais vraiment de très petits volumes."* La chaîne de valeur que nous reproduisons sous la forme du schéma ci-dessous concerne donc plutôt des clients anglophones. Il résume le fonctionnement de la chaîne

124. Amazon Web Services

125. Une foule de travailleurs

de valeur pour la production de l'intelligence artificielle. Un premier aspect de ce cas est la place très importante d'une entreprise du Cluster 1 (*full-stack AI services*) qui intègre toutes les phases du processus de production à ses services. Un deuxième élément est l'existence de trois chaînes de production potentielles qui ont comme point commun de passer par le même outil d'annotation : C2 - C1 - C4 (1), C1 - C4 (2) et C2 - C1 - Crowd (3). D'après nos enquêtés, ce qui distingue l'utilisation de la plateforme ouverte et de celle fermée sont le volume requis et la complexité des tâches d'annotations. Dans les deux premiers cas, le Client IA et la firme focale ont des contacts avec le sous-traitant, en plus, dans le cas 2, de maîtriser également l'intégration logistique des travailleurs à travers l'outil d'annotation et des critères qualité (notamment sur la sécurité).

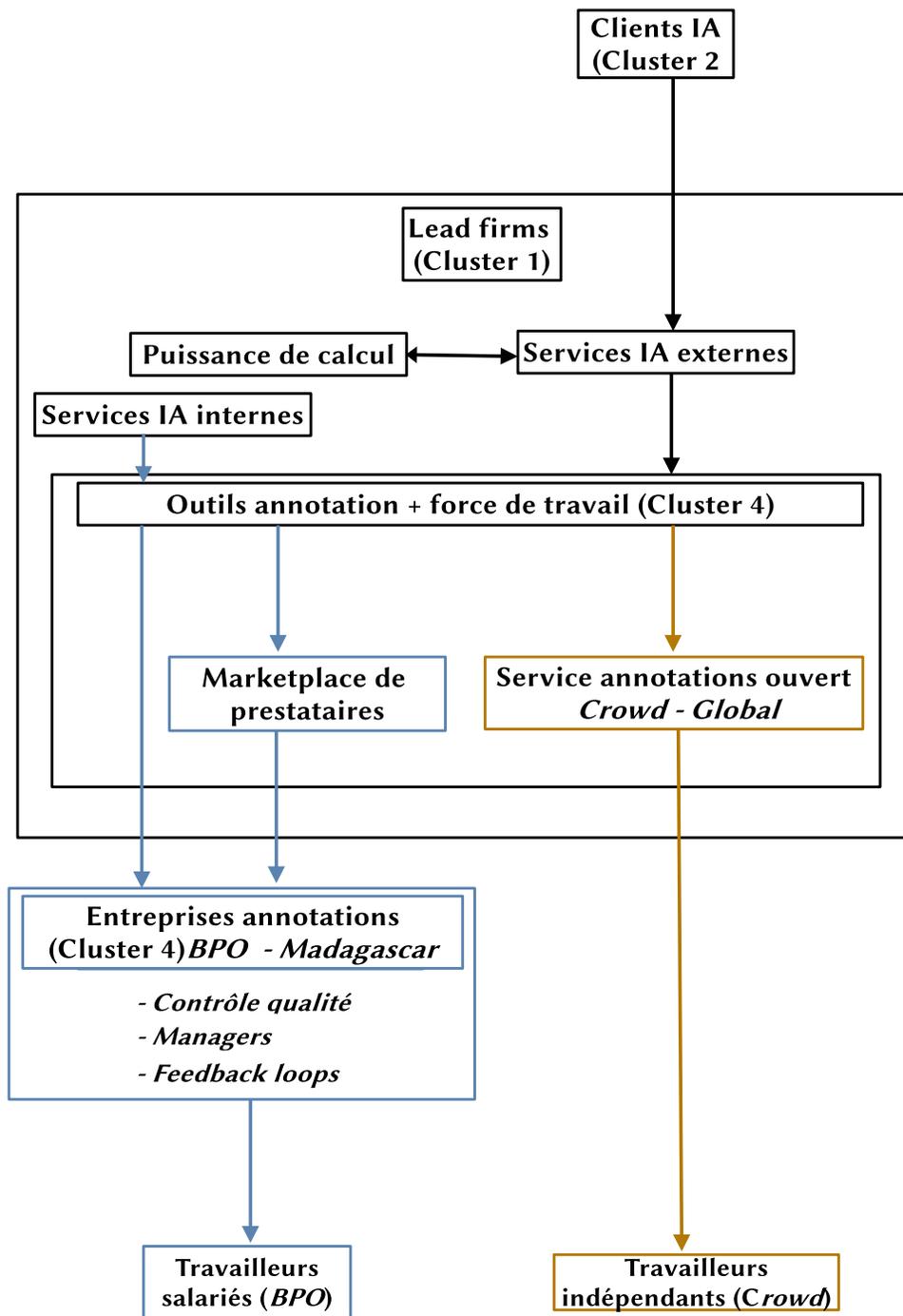


FIGURE 20 – Chaîne de valeur BPO et plateforme ouverte d’une firme focale de l’IA

Ce premier résultat suggère une évolution du modèle *“platform-as-market”*, pourtant largement mis en avant par les entreprises du numérique. On constate qu’y compris dans le cas d’une des firmes hégémoniques de

l'IA, le recours à une force de travail stable dans le temps. Dans ce contexte, la plateforme d'annotation se limite à un outil d'annotation, puisque la gestion des ressources humaines est majoritairement gérée par l'entreprise prestataire de service et partiellement partagée avec le donneur d'ordres. Par ailleurs, contrairement au modèle du marché "pur", ce mode de fonctionnement permet aux clients d'accéder à leurs prestataires afin d'échanger sur les pratiques d'annotations devenant finalement une version "relationnelle" du marché "pur". Par ailleurs, au-delà du seul cas d'Amazon, d'autres entreprises essayent de se positionner sur le créneau des outils d'annotation et travaillent avec les mêmes prestataires à Madagascar¹²⁶.

Le sous-traitant K est un exemple récent de cette forme "relationnelle" de place de marché. Il s'agit d'une entreprise qui vend également un outil d'annotation, et qui propose à ces clients d'accéder à des annotateurs. La différence fondamentale avec le cas précédent est que cette entreprise est uniquement spécialisée sur une seule partie du processus de production (les outils et la force de travail). Comme dans le cas précédent, cet outil va générer plusieurs types de relations commerciales. Dans le cas du sous-traitant E, qui, au moment de notre entretien, essaye de pénétrer le marché français, l'objectif est plutôt d'accéder à des clients. Ces derniers payent une licence d'utilisation de l'outil et peuvent fournir, au besoin, un *personnel, formé et dévoué, (qui) se charge de l'exécution des travaux d'étiquetage ad hoc*. d'après le site de la plateforme. En termes de contractualisation, le client paye un prix à l'heure relativement similaire à leur tarif sur d'autres plateformes. Pour les prestations d'annotations, les clients finaux sont en contact avec les entreprises sous-traitantes, comme c'est le cas dans le modèle platform-as-firm d'Amazon. Ainsi, le client final peut donc interagir directement avec les travailleurs de l'entreprise de sous-traitance.

Encadré n°34 : Réaliser des tâches sur une plateforme relationnelle 1

Q : Et [sous-traitant K], ça marche comment en termes de contractualisation par rapport aux autres plateformes ?

R : (...) Ils vendent une licence d'utilisation de leur plateforme à leur client, et nous on vend à [sous-traitant K]... Si le client décide d'utiliser 1 000 heures d'annotation de [chez nous], il nous paiera 1 000 x X, X étant le prix sur lequel on s'est mis d'accord à l'heure. C'est comme ça que ça marche.

Q : D'accord. Et le prix, c'est à peu près 5\$ de l'heure, c'est ça ?

R : En gros, ça tourne autour de ça. (...) c'est le prix de base. Mais par exemple si on a besoin de recruter des étudiants en médecine pour un projet particulier, on adapte le prix.

Q : Vous l'avez déjà fait ça ?

126. "R : On a des partenaires (...) Depuis 2/3 ans, on voit une catégorie d'acteurs dont le métier est de mettre sur pied des plateformes d'annotations et de dire à leurs clients Nous, on ne va pas annoter la donnée, on vous met à disposition les outils d'annotation, et à vous de voir si vous faites appel à vos équipes internes pour les annoter ou si vous faites appel à des gens, soit comme [nous] qui sont des spécialistes, ou des gens un peu plus opportunistes ou généralistes, comme par exemple Vivetic qui pourrait faire de l'annotation (...) On a des partenariats solides avec des boîtes qui s'appellent Labelbox, Alegion, quelques acteurs comme ça..." (Patrick - CEO - Sous-traitant E)

R : Oui.

Patrick - CEO - Sous-traitant E

Une autre entreprise étudiée a également recours à cet outil pour leurs clients, lorsque ces derniers n'ont en pas. Dans ce contexte, l'entreprise d'annotation est également partenaire de celle qui fournit l'outil, la première payant l'outil lorsque les clients finaux n'ont pas le leur et sont aussi prestataires lorsque le sous-traitant K a des clients qui souhaitent accéder à des travailleurs. Ce mode de fonctionnement implique des liens forts entre l'entreprise qui fournit la force de travail et celle qui fournit l'environnement de travail : l'outil d'annotation.

Encadré n°35 : Réaliser des tâches sur une plateforme relationnelle 2

Q : Je me demandais comment ça marchait, parce que ça me semblait assez différent d'un modèle crowdsourcé classique (dans le sens d'Amazon Mechanical Turk).

R : Ah oui, tout à fait. On est complètement à l'opposé de ça. [sous-traitant K], c'est un outil d'annotation. Ce ne sont pas des gens. Ils vendent un outil d'annotation, par exemple à des grosses sociétés du CAC40 qui ont des données très confidentielles et qui vont même l'utiliser en interne[...]. Par contre, ils ont aussi l'outil cloud. Aujourd'hui c'est l'outil qu'on utilise, que l'on met en place et que l'on maîtrise parfaitement et que l'on utilise pour nos propres clients quand ils n'ont pas leur outil. Avec [sous-traitant K], on a une relation à double sens : à la fois on les paie pour mettre en place l'outil pour nos clients, mais [sous-traitant K] nous apporte aussi des clients, qui sont ses propres clients et pour lesquels ils ont besoin de main-d'œuvre puisque le client peut à la fois avoir besoin de la plateforme pour des besoins d'annotation en interne, mais aussi pour de l'annotation externe. À ce moment-là on fait partie des boîtes (on n'est pas les seuls) qu'ils sollicitent pour annoter pour leurs clients (...)

Pierre - CEO - Sous-traitant G

Toutefois, cette position d'intermédiaire de l'entreprise d'outil d'annotation n'est pas sans conséquence pour les travailleurs. En effet, comme l'indique Simon, un chef de projet du sous-traitant G dans un entretien, les annotateurs font remonter les problèmes de la plateforme d'annotation, contribuant à l'amélioration du produit. Un point très particulier de cette relation de sous-traitance "plateformisée" est que, malgré la présence d'un intermédiaire, les annotateurs et leurs chefs de projet sont en relation directe avec le client final. Dans un entretien, Simon, chef de projet chez le sous-traitant G, évoque l'exemple d'un donneur d'ordre (Cluster 2) qui a souhaité se passer de la plateforme d'annotation considérant que le prix à payer était trop important, par rapport aux problèmes de la plateforme et à l'apport des annotateurs. En effet, il

n'est pas rare que ces plateformes aient des ralentissements ou même des problèmes pour gérer différents types de données ¹²⁷.

127. *on a déjà eu un souci assez cocasse avec (...) une société qui est travaillé dans le système infrarouge et la sécurité. Donc là ils ont remarqué (...), qu'il y avait beaucoup de soucis et des bugs partout (...), ce n'est pas l'idéal pour faire les choses [bien](...) les rares fois où il n'était pas là [les personnes de chez [sous-traitant K]], il nous a dit (...) si on peut continuer à part(...). Là, il a dit carrément les choses : les annotateurs doivent être plus payés que l'interface ... enfin bon une situation assez [inedite] (...) Il a vraiment souligné ça ... c'est très courant, on savait aussi que les annotateurs étaient sous-payés ... c'est tout naturel puisque là, même en France ils allaient payer vraiment des salaires confort, pourquoi pas prendre directement leurs compatriotes ou leurs frères pour leur payer des sommes conséquentes. Donc c'est pour ça qu'ils viennent ici à Madagascar. Donc là on est content sur les 2 aspects parce que là on a gagné aussi, même si c'est maigre ... c'est vraiment très maigre, 500€ en France c'est le salaire des stagiaires, ici c'est un cadre supérieur vraiment ... ! c'est vraiment quelqu'un qui est ... ! c'est à peu près la situation, il y a 2 aspects, on se sent exploité mais pas autant, parce que là on a gagné sur les 2 côtés quand même (Simon - Chef de projet - Sous-traitant G)*

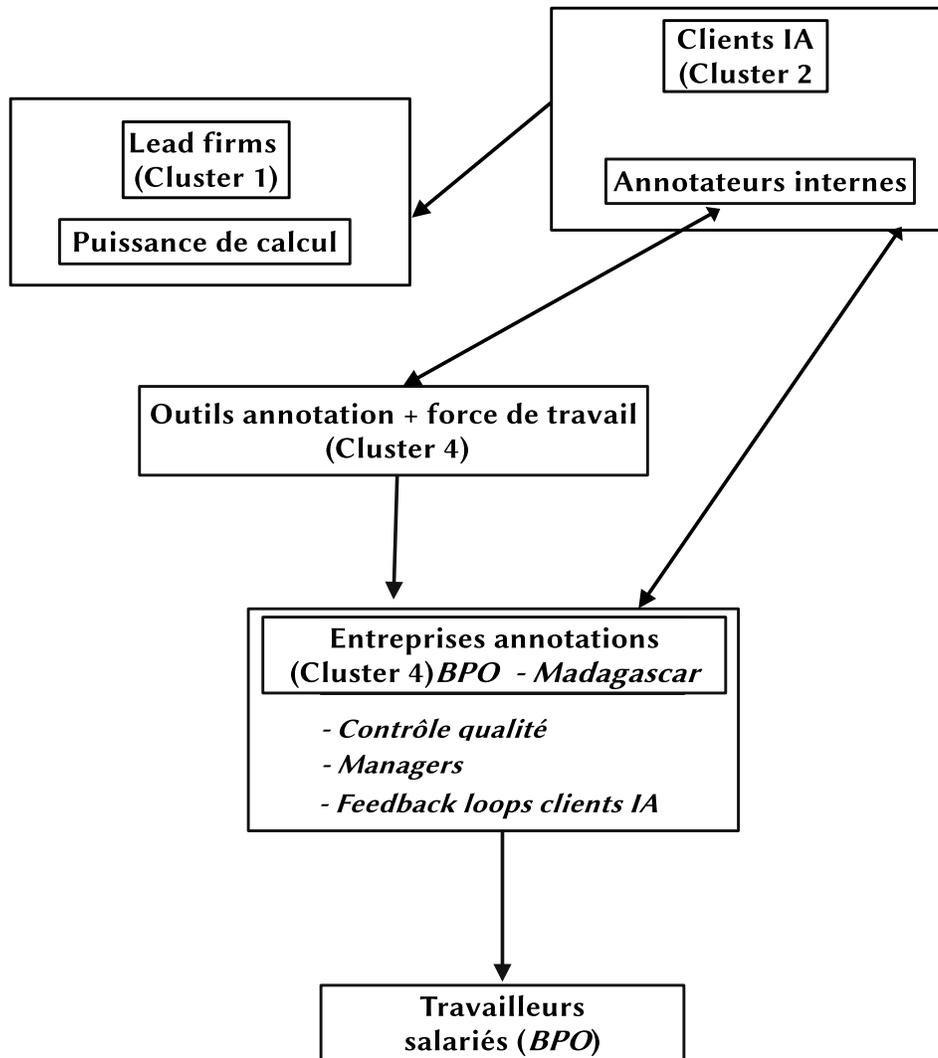


FIGURE 21 – Chaîne de valeur BPO et plateforme ouverte d’une firme focale de l’IA

Comme on l’a vu, les outils d’annotations peuvent également constituer des actifs spécifiques, à la fois recherchés par des clients qui essaient de développer leurs propres produits d’IA, dans les pays occidentaux,

et par les entreprises d'annotation de données situées dans les pays du Sud Global. La création de ces outils implique des développements spécifiques difficilement atteignables pour les entreprises de sous-traitance qui doivent alors nouer des partenariats avec des firmes spécialisées. Les startups spécialisées dans l'IA sont plus nombreuses à développer elles-mêmes ce type d'actifs. Cette tendance répond à la nécessité de maîtriser le processus d'annotation et à des cas particuliers d'annotations limitant alors le recours à ce type d'intermédiaire pour ces acteurs.

Le corollaire d'un investissement de plus en plus massif dans l'intelligence artificielle est la volonté de grands groupes de développer leurs propres projets. On retrouve ainsi des départements R&D de grandes entreprises qui disposent d'équipes de développements en interne (data scientists...) mais qui n'investissent pas nécessairement dans le développement d'un outil d'annotation. Pour ces acteurs, les personnes interrogées soulignent une évolution du secteur de l'annotation qui passe par le développement d'un marché spécifiques aux outils d'annotation et de manière secondaire sur l'accès à des services d'annotations. Ces entreprises d'outil font ensuite elles-mêmes fait appel à des sous-traitants de types BPO dans les pays en voie de développement.

Cette première sous-section nous a permis d'analyser un exemple canonique dans la littérature sur le travail des données : celui d'Amazon Mechanical Turk et plus généralement des plateformes comme place de marché. Les deux exemples analysés nous permettent de mieux comprendre l'organisation du travail d'annotation en identifiant le rôle des outils d'annotations dans l'intégration des travailleurs et complètent les éléments analysés précédemment dans notre précédente section. Dans le cas d'Amazon, la plateforme devient un simple outil d'annotation, la gestion des travailleurs étant partagée entre l'entreprise sous-traitante et le client final. Contrairement aux modèles de marché pur, ce mode de fonctionnement permet aux clients finaux d'échanger sur les processus d'annotation, devenant finalement une version relationnelle d'un marché pur. Notre deuxième cas est une version approfondie de la plateforme comme simple outil de gestion de projet et d'annotation, puisqu'on a constaté qu'il sert à la fois aux entreprises sous-traitantes et aux entreprises d'IA. Plus que l'interface d'annotation, les possibilités relationnelles sur et en dehors de la plateforme sont alors valorisées par les clients finaux.

Pour nous, ces résultats viennent apporter des éléments empiriques à un débat abordé par la littérature sur l'organisation du travail du numérique autour des différences entre les BPO et les plateformes (Lehdonvirta et al., 2019). En effet, nous montrons que les deux modèles coexistent dans le même secteur, voire même au sein de la même entreprise. Plutôt qu'une différenciation des modèles, nous constatons plutôt une forte intrication qui laisse davantage de place à des sous-traitants "encastrés" dans le processus de production (Uzzi, 1997). Les abordés dans le contexte français permettent d'éclairer les facteurs qui expliquent ce résultat. En effet, pour que les modèles correspondent à la "vérité terrain" et qu'ils soient applicables dans le monde réel il est nécessaire que les acteurs échangent constamment, ce qui rend plus pertinent l'accès à

une force de travail stable avec laquelle il est possible d'échanger.

5.2.2 Les BPO intégrés aux plateformes : la forme la plus proche du "marché pur"

Un deuxième exemple d'intermédiation des activités d'annotation mérite d'être évoqué : celui de l'entreprise Appen. Cette dernière se présente comme une "plateforme" qui fournit des outils d'annotations et surtout l'accès à une force de travail. Historiquement, Appen s'est principalement focalisée sur le marché du Natural Language Processing (NLP). Par conséquent, l'entreprise dispose d'un important portefeuille de clients dans le secteur de l'intelligence artificielle. Par ailleurs, Appen a récemment racheté l'entreprise FigureEight, une entreprise pionnière du "crowdworking" qui, *"utilise l'intelligence humaine pour effectuer des tâches simples, telles que la transcription de textes ou l'annotation d'images, afin d'entraîner les algorithmes d'apprentissage automatique."* (Wikipédia). La particularité de FigureEight est d'avoir très longtemps servi comme plateforme de recrutement de travailleurs des données pour Microsoft et pour d'autres "clients historiques"¹²⁸.

Comme évoqué dans notre cinquième chapitre, il faut également noter que Appen se distingue de par la variété de son offre en matière d'annotation de données : "annotation à la demande", "annotation à distance", "annotation sûre"; "équipe sur place"; "annotation par les clients". Le directeur des opérations du sous-traitant E nous explique que le mode de fonctionnement de la "foule sécurisée" proposée par la plateforme Appen diffère du modèle de la plateforme comme place de marché. Les travailleurs du sous-traitant E opèrent en effet depuis travail un espace dédié pour les annotateurs ayant des compétences spécifiques, c'est-à-dire de "l'annotation supervisée" : *"C'est la même chose. C'est-à-dire que ça peut être du crowd comme il y a la notion de qualification qui peut être propre à un fournisseur. Ce sont seulement nos identifiants qui ont ces qualifications et donc le client, quand il publie ses tâches, ce ne sont que les personnes qui ont les qualifications appropriées qui peuvent y accéder et du coup, ce n'est que [nous] qui [avons] accès à ce job"* (Loik - COO - Sous-traitant E). En l'occurrence, il s'agit d'une force de travail plus qualifiée qui opère sur des projets plus techniques et qui nécessitent un plus haut niveau de sécurité, sans toutefois qu'il y ait de contact entre le travailleur et le client final. C'est ce que confirme un chef de projet du sous-traitant F, qui travaille depuis plusieurs années sur un projet d'annotation de factures pour un supermarché américain à travers la plateforme Lui-même remarque qu'il n'a pas de contact avec le client final, et pour lui, son *"client c'est Appen"* (Héry - Chef de projet - Sous-traitant F).

128. "Q (...)Vous diriez que la ventilation des projets entre vous et Appen, ce n'est pas le même type de projets en fait ? R : On a encore une grosse partie de nos projets qui sont du computer vision. Des boîtes comme Appen sont plus sur du NLP, donc de la langue. Il y a aussi la ventilation des clients. Appen a certains clients qui sont des clients historiques ou d'autres clients qu'ils ont signés sur des projets et le nombre de clients dans les boîtes de tech est tellement énorme que ça fait que l'on a de la place." (Dinasoa - CEO - Sous-traitant E)

Encadré n°36 : Une force de travail qualifiée pour les plateformes

R : Pour Appen c'est différent. Quand on a travaillé sur MTurk, c'était vraiment les équipes d'Amazon qui nous envoyaient leurs jobs. C'était eux qui créaient directement et nous dirigeaient les jobs directement. Pour Appen, c'est différent [...] c'est l'offre d'annotateurs qualifiés. Le client final n'est pas nécessairement au courant de l'identité de l'annotateur ou que c'est [chez nous] l'annotation. Ils savent que Appen fait appel à un BPO pour de la main-d'œuvre qualifiée et voilà [c'est] leurs tarifs et c'est eux qui font l'interface. Nous, on parle aux équipes d'Appen et c'est eux qui communiquent avec le client.

Q : Et le client sait qu'il fait appel à un BPO, c'est ça ?

R : Oui. Il le sait à travers Appen.

Q : Appen lui précise et sait ce que les gens recherchent ?

R : Oui.

Loik - COO - Sous-traitant E

Selon un des directeurs du sous-traitant E, il s'agit même d'une évolution de l'entreprise qui, venant du modèle "crowdsourcing" propose maintenant des prestations centrées sur l'outil d'annotation et la mise à disposition de différentes options en matière de force de travail, aspect qu'on a déjà identifié chez d'autres plateformes historiques comme Mturk : *"En fait, Appen s'appelait Figure Eight avant (...). Ils ont deux métiers. La résultante de leur métier historique qui était de faire du crowdsourcing, c'est que maintenant ils vendent aussi des solutions où tu as juste la plateforme"* (Patrick - CEO - Sous-traitant E). Les entreprises peuvent ensuite choisir entre ces différents modèles, en sachant que le recours à une force de travail supervisée est nécessaire pour des projets plus complexes dans lesquels les annotateurs en besoin d'encadrement : *"et à toi de voir si tu fais appel au crowd, donc à n'importe qui, n'importe où, à des acteurs qui ont une supervised workforce ¹²⁹, donc des gens comme nous, qui ont des annotateurs qui sont encadrés, formés, suivis (...) pour les projets un peu plus compliqués et pas les trucs basiques, ou alors, si tu fais les annotations avec des équipes internes."* (Patrick - CEO - Sous-traitant E)

Ce mode de fonctionnement est confirmé par Frédéric, dirigeant du sous-traitant F présente à Madagascar et en Bulgarie qui travaille pour Appen depuis plusieurs années ¹³⁰. Dans cet entretien, nous avons notamment évoqué les différents modèles organisationnels, ce qui permet de mieux comprendre les contrastes entre la sous-traitance par une plateforme comme marché et le modèle proposé par Appen. Frédéric nous explique notamment que le problème des plateformes dans le modèle "crowd" est le manque de disponibi-

129. "Force de travail supervisé"

130. *"ça fait longtemps qu'on travaille avec eux. On les connaît très bien. Eux ont un volume, je ne sais pas comment ils font ! C'est énorme."* (Frédéric - CEO - Sous-traitant F)

lité et de stabilité de la force de travail ce qui limite le recours à cette offre pour des projets récurrents. C'est ce qui a poussé ces deux plateformes à créer des services de type "BPO". Selon lui, *"sur des sujets récurrents, il faut qu'ils aient la garantie d'une stabilité. Et chez les gros comme Appen, MTurk, ils ne l'ont pas. Ils ont des clients qu'ils sont obligés de servir en permanence et ils sont obligés d'assurer la qualité en permanence. Leur grande force, c'est qu'ils vont pouvoir atténuer les pics, mais ils sont obligés en réalité d'avoir du BPO derrière, avec une sécurité et des gens qui connaissent le process, les erreurs, et qui vont garantir une qualité."* (Frédéric - CEO - Sous-traitant F) Par opposition aux travailleurs indépendants qui réalisent les tâches depuis leur domicile ou un cybercafé, ce modèle permettrait, outre la qualité, une meilleure sécurité des données.

Frédéric nous donne également quelques éléments d'explication qui expliquent le recours à son entreprise. Par exemple, lorsqu'un projet nécessite beaucoup d'images à annoter dans un délai restreint, le recours à Appen va être plus aisé, car cela va permettre au client final d'avoir accès à de nombreux prestataires, BPO ou indépendants. Toutefois dès qu'il s'agit d'un projet "récurrent", et qui nécessite de former des annotateurs, il pourra proposer ses services : *"Après, vous pouvez avoir des très gros en direct qui ont besoin d'annotation en permanence, et c'est là où on rentre en jeu. Dès que le projet est très gros en très peu de temps, on ne le fera pas, parce que de toute façon on fera moins bien que des gens comme Appen. Dès qu'il y a un intérêt à former l'annotateur ou que l'annotateur doit avoir une compétence particulière ou un savoir particulier, là ils viennent nous voir."* (Frédéric - CEO - Sous-traitant F). Il explique par exemple travailler pour le compte d'une entreprise de e-commerce qui développe un outil de détection de produits à vendre sur des images à l'aide d'un modèle de reconnaissance d'images. Par exemple, l'outil doit détecter s'il y a un sac sur l'image ou s'il s'agit d'une chaussure ¹³¹. Ce projet s'est déroulé en peu de temps, ce qui explique selon lui que l'entreprise se soit tournée vers le Appen et son service de "foule supervisée". Il cite également le cas d'un fabricant de voitures qui développe un modèle qui détecte les anomalies des produits qui passent sur la chaîne de montage ¹³². Dans ce cas, le projet dure plus longtemps (réurrence) et le modèle a vocation à atteindre un niveau de performance quasi parfait ce qui nécessite d'annoter énormément d'images. Le caractère plus expert du projet nécessite alors de développer les compétences des annotateurs, ce qui justifie le recours à une sous-traitance directe, en comparaison à Appen.

Le dirigeant du sous-traitant E, confirme l'importance du facteur volume, en ajoutant la capacité à standardiser les tâches : *"Quand elles ont beaucoup de volume, très peu de boîtes peuvent faire ça en interne. Donc*

131. *"Par exemple, [Mode-ECommerce] (on bosse pour eux aussi), qui va dire je veux reconnaître en 2 secondes un sac dans une image . Donc ils vont faire quoi ? Ils vont aller voir Appen ou un concurrent. En plus de ça, il y a plein de plateformes maintenant, une quarantaine. On les a tous contactés et on travaille pratiquement avec tous. Donc Mode-ECommerce va dire je veux entraîner mon machine learning à reconnaître un sac parmi des millions d'images . Imaginez que pour le machine learning, il faille lui balancer 1 million d'images donc de les annoter. Et ce qu'ils demandent, c'est le prix et le délai. Donc ils vont contacter toutes les plus grandes plateformes"* (Frédéric - CEO - Sous-traitant F)

132. *"Après, vous allez avoir des gens comme [CarFabric], etc., qui ont à la fois du gros volume et à la fois de la récurrence et qui vont vous dire qu'ils veulent un truc pas à 99,1, mais à 99,999999, et ce n'est pas 1 million d'images, mais 100 millions."*(Frédéric - CEO - Sous-traitant F)

à un moment elles le confient à des gens. Si tu veux plus de volume sur un truc standard (est-ce que tu vois un chien ou un chat ? , captcha basique), c'est totalement adapté pour faire du crowd. Ton critère c'est le coût, point final, et pas vraiment la qualité ni la sensibilité.” (Patrick - CEO - Sous-traitant E). Par le biais des outils d'annotations, son entreprise travaille sur des projets ”où le niveau de sensibilité est hyper importante. On reçoit des clichés d'imagerie médicale (radios, scanners, etc.), c'est le genre de choses que tu ne peux pas balancer sur le crowd, c'est sensible et potentiellement risqué.” (Patrick - CEO - Sous-traitant E). On retrouve le même principe en matière de sécurisation sur Appen que sur Mturk avec la nécessité de remplir des conditions en matière de sécurité des données permettant de gérer ces données sensibles comme : ”Des images aussi pour Amazon où on voyait des internautes ou des données d'internautes/clients, et donc par rapport au crowd, Amazon nous fait un audit sécurité tous les ans, on a un label qui s'appelle le label SOC2 qui est la norme US de sécurité des données qu'on a dû passer avec un cabinet d'audit international.” (Patrick - CEO - Sous-traitant E)

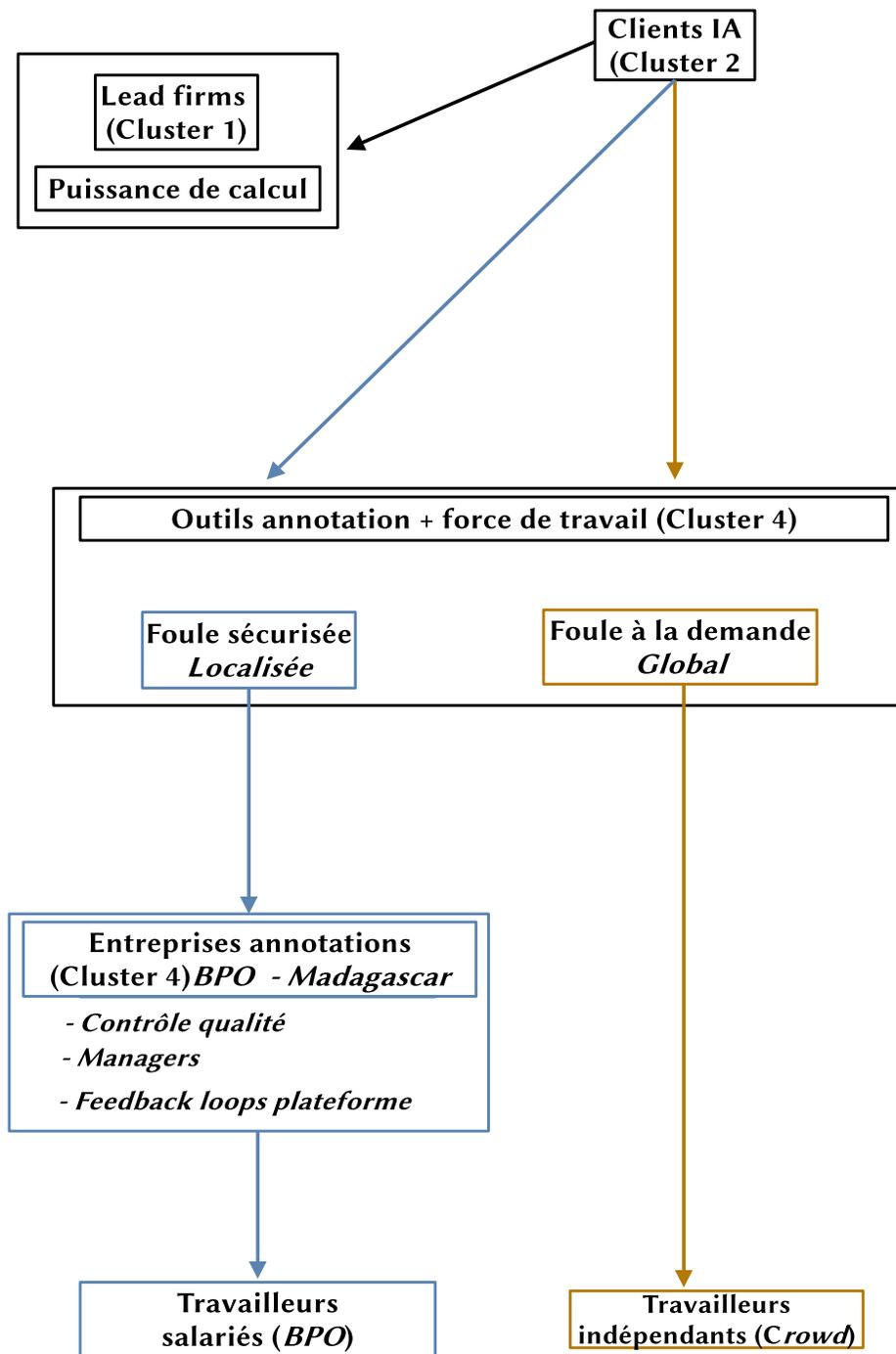


FIGURE 22 – Chaîne de valeur BPO intégré à une plateforme et plateforme ouverte d’une firme focale de l’IA

Enfin, nos entretiens montrent que le recours à une force de travail supervisée est nécessaire pour des projets plus complexes dans lesquels les annotateurs ont besoin d’encadrement ou lorsque les données sont

considérées comme sensibles (par exemple : dans le domaine médical), ce qui justifie d'utiliser le mode d'externalisation "sûr" en passant par Appen. Si le projet nécessite davantage encore d'expertise, par exemple, s'il est nécessaire que le modèle atteigne un haut niveau de performance, cela limite le recours à Appen et favorise la sous-traitance directe. Le manque de disponibilité et de stabilité de la force de travail "à distance" et "sûre" limitent le recours à cette offre pour des projets récurrents.

Type de service	Intégration travailleurs	Statut travailleurs	Catégorie
A distance	Globale	"Indépendants"	"Platform-as-market"
Sûre	Localisé	"Employés"	Mixte
Sur place	Internalisée par la plateforme	"Employés"	"Platform-as-firm"
Clients	Internalisée par le client	"Employés"	"Firm"

TABLE 6 – Tableau de synthèse des modes d'externalisation de l'entreprise Appen

Le cas de Appen confirme l'intrication du modèle BPO dans le modèle plateforme. Dans ce cas, certains travailleurs sont en effet employés et payés par des entreprises formelles à Madagascar, mais opèrent à travers la plateforme, sans qu'il y ait de relation avec le client final. Ce résultat suggère qu'outre la version "marché" de Mturk, ce cas représente la forme plus pure de marché du fait d'une moindre intégration des travailleurs dans le processus de production du client final. Toutefois, plutôt que des relations travailleurs - client final, il s'agit de relations inter-firmes intermédiées par une plateforme de mise en relation (Appen). Ce cas questionne également les relations du pouvoir qui pèsent sur les travailleurs. Sur la plateforme Appen, on constate qu'il s'exerce principalement à travers leurs employeurs à Madagascar plutôt que via la plateforme. Ces éléments viennent confirmer et prolonger les analyses de Tubaro (2021) et A. Casilli et al. (2019a) à propos de l'existence de plateformes intermédiaires et de mécanismes de coordination du travail qui sont partagés entre plusieurs acteurs (*deep labour*)¹³³.

Pour nous cela confirme également le fait qu'il ne faut pas uniquement analyser le phénomène des plateformes à travers le prisme de l'émergence d'un "marché du travail mondialisé" (Anwar, Schäfer, & Golušin, 2023). Il s'agit plutôt du prolongement numérique de la capacité des entreprises occidentales à essayer de standardiser des activités et une organisation du travail. De ce point de vue, nous considérons qu'il ne faut pas voir du marché partout, mais plutôt une variété de mode d'intégration des travailleurs qui passe soit par la hiérarchie (souvent) soit par le marché (plus rarement). Au contraire, nous considérons la forme *marché pur* comme une exception, ce qui limite d'après nous le recours même au concept même de plateforme. Le terme induit en effet une organisation du travail qui fonctionne par la mise en relation de l'offre et de la

133. Schématiquement, une plateforme s'occupe du recrutement, une autre de l'outil de réalisation des tâches et une dernière du paiement.

demande de travail et donc par le marché qu'on retrouve finalement très rarement. La dimension "qualité" joue alors pleinement dans le choix d'un mode d'externalisation plutôt que d'un autre.

Bien sûr, cela ne veut pas dire que nous n'admettons pas le flou organisationnel induit par la mise en place de ces organisations du travail aux marges de la légalité. Nous pensons toutefois qu'il est nécessaire de reconsidérer l'approche conceptuelle de ces objets pour mieux tenir compte, d'une part de l'hétérogénéité de l'organisation du travail (Cornet et al., 2022), et en corollaire des modes d'intégration des travailleurs dans des systèmes hiérarchiques qu'ils soient gérés par le client final, par la plateforme ou par une entreprise sous-traitante intégrée dans une plateforme. Pour nous, il n'y a finalement pas de plateformes, mais uniquement des outils permettant la mise en place d'une "hiérarchie électronique", quelque soit la "face" du marché qui détient le pouvoir sur les travailleurs. Cette approche conceptuelle a des implications concrètes en termes de régulations du secteur. En effet, nous pensons qu'il est alors nécessaire d'analyser ce que Virgile Chassagnon appelle le pouvoir "de facto" qui désigne la dépendance des acteurs dans l'accès à des ressources critiques (par exemple : ressources humaines). Notre dernière sous-section se propose justement de prolonger cette analyse à travers un dernier cas d'études : celui d'une "plateforme" au sein de laquelle on constate qu'un réel pouvoir managérial s'exerce sur les travailleurs.

5.2.3 L'organisation du travail dans une entreprise d'impact sourcing : un pouvoir managérial qui s'exerce directement sur les travailleurs

Dans l'introduction de notre thèse, nous avons évoqué plusieurs formes d'organisation du travail de la donnée, et notamment celui de Sama qui s'est positionnée sur le secteur de l'externalisation "à impact" (*impact sourcing*). Fondée par Leila Janah, une ancienne salariée de la Banque Mondiale, cette entreprise a contribué à forger le concept de micro-work, en s'inspirant du micro-crédit. Comme pour ce dernier, Sama s'adresse aux travailleurs des pays en voie de développement, en leur demandant de réaliser des petites tâches contre une rémunération, censée leur permettre de développer un projet d'entrepreneuriat. Pour ce faire, les travailleurs bénéficient d'un accompagnement de l'entreprise (par exemple : des formations) L'entreprise est particulièrement connue pour avoir trouvé les travailleurs kenyans, faiblement rémunérés, chargés de s'assurer que les données utilisées pour entraîner ChatGPT ne comportaient pas de contenu à caractère discriminatoire.

Ce modèle a été dupliqué par EthicAI, une entreprise européenne qui se présente comme une plateforme éthique et qui propose une main-d'oeuvre qualifiée pour des projets de saisie ou d'annotation de données. Ces travailleurs sont principalement des travailleuses qui sont situées sur plusieurs continents (Amérique latine, Afrique, Asie)¹³⁴. Comme pour d'autres entreprises qui prétendent relever du *marché* ou de la *plateforme*, les travailleuses sont en fait des entrepreneuses. Les travailleuses interviewées indiquent d'ailleurs

134. Dans le cadre de cette enquête, nous avons eu uniquement accès à des travailleuses camerounaises et malgaches.

qu'elles doivent déclarer cette activité sous un statut d'indépendant dans leur pays d'origine pour pouvoir exercer sur la plateforme et ce, même si elles ont un "contrat" de trois ans avec la plateforme.

En termes de rémunérations, elles indiquent toucher 1.5 euros de l'heure. Toutefois, le paiement se fait en fonction de l'heure de travail réalisée et est décidé par le client final. Par exemple, si ce dernier calcule qu'une tâche dure une minute, les travailleuses devront faire 60 tâches, même si chaque tâche leur prend en fait deux minutes. Au moment de nos entretiens, leur salaire a été divisé par deux, passant de trois euros par heure à un euro cinquante. Cette décision a été prise de manière unilatérale par la dirigeante qui l'a justifié par le fait d'adapter le salaire aux niveaux de vie des pays dans lesquels habitent les travailleuses. Comme pour les exemples cités précédemment, nous nous intéressons ici à l'organisation du travail. Dans ce cas, la firme principale (EthicAI) se présente comme une entreprise qui coordonne un réseau d'auto-entrepreneuses. Or, au fil des entretiens, nous avons constaté, que dans les faits, l'intégration des travailleuses se rapproche de la *subordination* plutôt que du "marché pur" à la Mturk.

Cela se traduit notamment par une intégration relationnelle, à travers les formations réalisées par les coachs de EthicAI, auprès des travailleuses. En l'occurrence, Carine raconte son lien avec Cédric, un chargé de projet qui s'occupe de récupérer les projets et les cahiers des charges des clients pour les dispatcher auprès des travailleuses¹³⁵. Comme l'explique également Véronique, une autre travailleuse, le rôle des chargés de projet est centré sur la gestion des ressources humaines¹³⁶. Ils s'occupent également de la gestion des projets, par exemple, la vérification du niveau de qualité et du nombre d'annotations réalisées, ce qui explique que certaines travailleuses indiquent recevoir des pressions de la part de ces chargés de projets : *"La pression que nous recevons régulièrement, ça vient de la part des chargés de projet. Seraphin il a une manie de vous booster. De vous mettre une pression sans vous mettre une pression. Sharika c'est vraiment la pression elle est visible. Elle est très crue. Et parfois même sa pression apporte la démotivation. Après on a passé la main à Gaëlle, c'est une fille, elle vous motive, et elle prend en compte les préoccupations de chaque hiteuse."* (Windy - Annotatrice de données - EthicAI). Ce management du quotidien est une part essentielle du travail des chargés de projet, mais cette fonction "RH" se traduit également par le recrutement et la sélection des travailleuses *"c'est ceux qui sont chargés des projets. Ce sont eux qui nous envoient les tâches (...) comme tout à l'heure on m'a envoyé le message dans mon Gmail que il y a un travail sur la labellisation d'images, il faudrait que j'aille répondre. Quand je réponds présent, on me donne le numéro et le code que je dois intégrer sur la plateforme, je vais entrer et je vais répondre au [chargé de projet]."* (Mélodie - Annotatrice de données - EthicAI). Comme l'explique Mélodie qui synthétise le rôle des chargés de projets : *"C'est elle qui s'occupe du projet, c'est un peu comme la*

135. *"Mon coach, c'était cédric, et on travaillait sur un projet de recherche de [pistes commerciales]. Recherche de contacts, il fallait savoir comment utiliser hunter pour la recherche des adresses mail. C'était vraiment pas facile. C'est la première formation que j'ai eu à faire chez EthicAI, pour pouvoir être opérationnelle sur le projet."* (Carine - Annotatrice de données EthicAI)

136. *Ces chargés de projet étudient comment faire [les tâches], et puis c'est eux qui nous montrent à travers des vidéos et des consignes comment faire les tâches"* (Véronique - Annotatrice EthicAI)

patronne du projet, c'est elle qui va récolter tout, c'est elle qui vérifie les tâches, qui est censée vérifier les tâches au fur et à mesure que l'on travaille. Donc c'est ça le CSM. Et quand elle voit qu'il y a une hiteuse qui travaille plus que les autres, après elle te refile son travail quoi" (Mélodie - Annotatrice de données - EthicAI)

Le travail de contrôle est également facilité par un outil de remontée d'information sur les annotations qui a été mis en place par EthicAI, similaire à ceux que nous avons évoqués précédemment (*intégration logistique*). Ces éléments sont complétés par des échanges réguliers entre les chargés et les travailleuses, comme l'expliquent Mélodie et une de ses collègues qui a rejoint l'entretien en cours. Dans le verbatim qui suit, elles racontent avoir travaillé sur un projet de transcription et traduction de courtes séquences audio, de l'allemand vers le français. Sur ce projet, Mélodie occupe une fonction de vérificatrice de la qualité des transcriptions réalisées par les autres annotatrices ¹³⁷. Or, elle réalise que de nombreuses séquences sonores incluent des passages dans un dialecte allemand différent de l'allemand que les annotatrices ont appris à l'école.

Encadré n°37 : Transcrire du dialecte allemand

C : il y avait le travail sur la traduction, (...) en fait ce sont des gens qui passaient des appels, mais je ne sais pas trop de quel type d'appel il s'agissait parce que en fait ils font tout ... il y avait de l'allemand mélangé au dialecte (...) un peu comme on dit le français des quartiers/ le français appris à l'école. Donc c'était des petits audios, ils avaient demandé que on doit écouter et faire la transcription en allemand sans rien changer. C'était très difficile, parce que l'allemand comme moi j'ai appris à l'école, c'était pas la même chose. Là c'était de l'allemand de dialecte ...

Mélodie - Annotatrice de données EthicAI

En tant que "vérificatrice", elle a réalisé un "rapport" en réécoutant l'ensemble des séquences sonores : *"ça m'a pris pratiquement une semaine, j'ai réécouté et réécouté et réécouté, et après j'ai fait mon rapport et j'ai dit que les hiteuses ont raison de faire des mauvaises traductions parce que l'allemand que vous demandez que nous on fasse la transcription, ce n'est pas l'allemand appris à l'école, ce qu'on nous apprend ici au Cameroun."* (Mélodie - Annotatrice de données EthicAI) Malgré la difficulté de la tâche, le client final ne souhaite pas rémunérer le temps de travail ce qui a conduit Mélodie à contacter le chef de projet : *"ils voulaient que je refasse toutes les tâches ! j'ai donc écrit au CSM, elle m'a répondu sur Messenger (...) pour qu'elle puisse soumettre à la direction que les tâches là ne peuvent être faites"* (Mélodie - Annotatrice de données EthicAI). Le projet finira par disparaître de la plateforme et les tâches demeureront impayées.

L'intégration des travailleuses se traduit également par des canaux de communication invitant les tra-

137. Fonction qui est uniquement symbolique, puisqu'elle ne donne pas lieu à une meilleure rémunération, comme ça a pu être observé dans d'autres entreprises

vailleuses à se connecter pour réaliser des tâches. Comme l'explique Véronique : " (...) ils ont beaucoup de canaux de communication donc quand ils disent qu'il y a des tâches, ils envoient un message et ils disent il y a tel nombre de tâches qui arrivent." (Viviane - Annotatrice de données EthicAI) Elle précise d'ailleurs qu'elle essaye de se connecter, jonglant avec ses autres activités : "Là, vous essayez de vous arranger en fonction de ça, mais si ça vous trouve pendant que vous êtes occupé, vous ne pouvez pas faire autrement ... " (Viviane - Annotatrice de données EthicAI) Une partie du travail relationnel nécessaire à la bonne réalisation des tâches est déléguée aux travailleuses elles-mêmes qui se donnent des conseils par le biais de groupe WhatsApp (messagerie) : "Ce sont des relations d'échange, je peux dire, professionnelles. On a notre groupe whatsapp de hiteuses du Cameroun. Lorsque quelqu'un a un doute, on échange sur le forum." Christiane - Annotatrice de données EthicAI). Certes, ce travail est réalisé en parallèle de projets d'entrepreneuriats, de reconversion ou d'études, mais il faut également souligner que pour ses travailleuses, il s'agit généralement de leur activité principale dont elle tire l'essentiel de leur revenu. Si bien qu'en plus de la subordination juridique qui s'exerce par le contrat et les outils et processus de contrôle, la plateforme exerce également une *subordination économique*.

Ces différents éléments dessinent une organisation du travail par laquelle la plateforme concentre le pouvoir de contrôle sur le travail effectué et un rôle de codification des tâches à effectuer. Cela passe par une intégration *logistique* avec une formation basée sur le cahier des charges donné par le client final et par un contrôle de la qualité des tâches. L'intégration est également *organisationnelle* et s'exerce directement de la firme focale vers les travailleuses par l'intermédiaire des chargés de projet. Ces derniers sont présents sur toutes les phases d'un projet, du recrutement des annotatrices jusqu'à la vérification de leur travail, en passant par leur formation. Finalement, le seul élément qui pourrait pointer vers une forme de relations intra-firmes, est la rémunération à la performance (qualité), qui dépend du contrôle réalisé par les vérificatrices et par le client final et qui ne tient pas compte du temps réel passé à travailler sur la plateforme. En parallèle, une partie du travail de compréhension des tâches se passe à l'intérieur du groupe des travailleuses ¹³⁸.

Dès lors, peut-on dire que cette organisation du travail relève plutôt de la forme pure du marché ou de la "hiérarchie"? De notre point de vue, peu d'éléments empiriques indiquent que ces travailleuses sont réellement indépendantes et qu'il s'agit d'un marché, y compris si l'on prend en compte la forme de la rémunération. Concernant les horaires, les travailleuses ont une marge de manoeuvre pour décider quand elles veulent travailler. Toutefois, cet aspect est limité par le mode de recrutement de la plateforme. Une annotatrice indique même qu'on lui a conseillé de négocier avec les chargés de projets pour obtenir des tâches. Comme l'explique le site d'Internet de EthicAI, ces entreprises prétendent faire exister un modèle hybride en s'appuyant sur le travail indépendant. Nos résultats suggèrent qu'il s'agit davantage d'un narratif utilisé

138. Il s'agit d'ailleurs d'éléments soulignés dans certains travaux consacrés au micro-travail et aux plateformes numériques de travail (A. Casilli et al., 2019a; Lagerie, 2019; Régniez et al., 2020).

permettant d'éviter de salarier les annotatrices, ce qui limite encore une fois l'hypothèse de développement de formes hybrides dans ce type de secteur (firme-réseaux et plateformes).

Dans cette section, notre travail portait sur les autres formes du travail des données rencontrées lors de notre d'enquête en approfondissant les constats de notre section 2.2. Ces modèles alternatifs de l'organisation du travail des données se situent, en théorie, entre la forme hiérarchique et le "marché pur" à la Mturk. Nous avons constaté une évolution dans la manière de produire des données, du moins, de réelles différences avec ce qui a été étudié précédemment dans la littérature et ce que nous désignons comme étant des "plateformes ouvertes". Ce modèle, largement étudié, concerne des plateformes ouvertes aux travailleurs et aux donneurs d'ordres, qui fonctionnent comme un marché dans lequel les clients proposent des tâches qui sont ensuite réalisées par une foule de travailleurs. Au contraire, nous avons constaté l'existence plusieurs modèles qui, à première vue, pourraient être qualifiés d'hybrides, tant ils empruntent à la fois aux modes de fonctionnement du *marché* et de la *hiérarchie* (voir le tableau qui suit).

Outsourcing mode	Workforce type	Legal status of workers	Infrastructure of access to work	Features	GVC type
Open-platform-oriented	Global	Self-employed	Labour platform	non-stable and non-secured workforce	Market
BPO-oriented	Localized workforce	Employee	Direct contracts	Strongly secured premises	Relational
Platformized BPO	Localized workforce	Employee	Deep labour platforms and annotation platforms (Appen, Labelbox)	Strongly secured premises	Modular

TABLE 7 – Typologie des fournisseurs de données

Nous désignons le premier modèle comme étant celui des "BPO" qui se caractérise par le recours par les donneurs d'ordres à des outils d'annotations comme Mturk et à une force de travail qui est située dans une entreprise. Les travailleurs ont un contact direct avec le client final (intégration organisationnelle) et l'intégration logistique va s'opérer via les outils de contrôle qualité des outils d'annotations et par les clients finaux. Les travailleurs sont des employés du sous-traitant.

Nous montrons que cet élément de la chaîne de valeur peut être considéré comme un actif spécifique. Dans ce second cas, nous avons analysé les relations entre une entreprise dont la proposition de valeur est de vendre l'outil et une entreprise de sous-traitance. Dans cette organisation du travail, on se situe dans la

continuité des "BPO", puisque le sous-traitant K vend, en plus de l'outil, l'accès à une force de travail. De leur côté, les travailleurs ont également accès aux clients finaux. Dans une troisième partie, nous avons étudié le cas de Appen qui sous-traite le travail des données à plusieurs entreprises de BPO, y compris à Madagascar. Pour nous, ce cas est la forme la plus pure de "marché" et d'une plateforme qui se comporte comme un intermédiaire. En effet, il n'y a pas de liens entre les travailleurs et le client final. Nous qualifions ce modèle de "BPO *plateformisé*". Ici la plateforme sert d'intermédiaire dans l'échange et d'outil pour l'intégration logistique. Elle se comporte comme le donneur d'ordre final des travailleurs de Madagascar.

Enfin, nous avons analysé le cas d'une entreprise qui se présente comme une plateforme coordonnant un réseau de travailleurs indépendants, autrement dit une firme-réseau. Peut-on pour autant y voir une forme hybride? En effet, l'entreprise utilise des incitations salariales à la performance et a recours au travail indépendant. Toutefois, plusieurs éléments donnent un autre regard sur cette organisation du travail. Les travailleuses ont des contrats sur plusieurs années, elles ont un cadre horaire, elles reçoivent des mails les invitant à réaliser des tâches et la plateforme contrôle leur tâche. Par ailleurs, on constate également qu'il existe une intégration logistique et relationnelle qui passe par les outils de contrôle et de communication.

Notre développement donc suggère que les formes hybrides en tant que catégories conceptuelles se heurtent à la réalité du terrain. Certes, il faut préciser que nous sommes bien en présence de relations intra-firmes, et que, dans chaque cas, on constate des emprunts de mécanismes classiquement associés à l'organisation hiérarchique ou à l'organisation par le marché. Mais, on observe que les organisations se positionnent systématiquement, soit plutôt du côté du marché (*open-platform* et *platformized BPO*), soit plutôt du côté de la firme.

* * *

Notre terrain montre que les organisations se rapprochent plutôt de la hiérarchie, élément organisationnel que nous approfondissons dans notre prochaine partie en analysant le travail des annotateurs comme un travail qualifié résultant d'une organisation collective. Sur ce point, on peut également faire le lien avec la littérature sur les *global value chain* (Gereffi et al., 2005). En effet, on constate la présence de formes de chaînes de valeur qui s'éloignent du marché et se rapprochent plus des formes relationnelles et modulaires, parce qu'il va y avoir des liens d'échanges très forts entre les firmes d'IA et celles de sous-traitance. Nous verrons d'ailleurs dans notre prochain chapitre que cet aspect s'explique par la nature même du travail effectué : les entreprises de sous-traitance vendent finalement plus des travailleurs (le travail des données) que des biens (les données), contrairement à d'autres secteurs (industrie).

Enfin, ces cas permettent d'éclairer le débat sur l'encastrement des travailleurs qui a été soulevé par Tubaro (2021); Wood et al. (2019b). Dans ces deux articles, les auteurs s'appuient sur le concept d'encastrement pour proposer une analyse du travail de plateforme selon laquelle les travailleurs sont " *encastres dans les*

réseaux interpersonnels des travailleurs, (mais qu'ils sont en même temps désencastré des normes culturelles et juridiques qui limiteraient sa marchandisation." (Wood et al., 2019b). Tubaro (2021) propose quant à elle le concept "d'encastrement profond" pour désigner les plateformes comme Appen qui "participent à des structures complexes d'externalisation interentreprises" qui se rapprochent de liens hiérarchiques.

En l'occurrence, notre terrain montre l'importance des formes d'intégrations logistiques et relationnelles dans les liens entre les travailleurs et les firmes donneuses d'ordres (les clients finaux), ainsi qu'avec leur employeur direct. Dans certains cas, ces formes d'intégrations viennent avec des protections proches du salariat, notamment pour les formes proches des BPO. Dans les autres cas, nous retenons une approche normative déjà abordée dans ce chapitre, qui consiste à dire que nous observons des éléments qui se rattachent à la forme hiérarchique, sans les protections qui vont avec. Comme l'explique (Tubaro, 2021) les travailleurs "ne serait pas décrite comme désincarnée ou marchandisée à proprement parler, car les travailleurs conservent certaines des protections qui vont généralement de pair avec l'emploi, mais je souligne des déséquilibres de pouvoir substantiels, par lesquels les entreprises influencent les individus, et non l'inverse." Même dans le cas de EthicAI, les travailleurs doivent répondre à certaines normes du travail en se déclarant auto-entrepreneurs, le problème étant que dans les faits elles ne le sont pas vraiment. Cette approche conceptuelle a finalement des implications dans l'analyse des chaînes de valeurs numériques (Wood et al., 2019b).

Notre approche montre que quelque soit le mode d'externalisation, les travailleurs sont profondément encastres dans des normes locales de travail. Surtout, comme on le verra dans notre dernière partie, cet encastrement se vérifie dans des conditions de travail largement régies par le droit du travail local. Pour conclure là-dessus, nous considérons que malgré ces formes hiérarchiques, la structure du marché dans lequel ces travailleurs opèrent va déterminer leur forme flexible d'emploi, comme nous le verrons dans le chapitre quatre. En effet, qu'il s'agisse de travailleurs des services ou de l'IA, ils sont soumis aux aléas des financements des donneurs d'ordres finaux qui se font souvent auprès des marchés (levée de fonds), dans un contexte où le travail est, en théorie, fortement mobile géographiquement¹³⁹. Ce faisant, nous confirmons également l'idée d'un pouvoir *de facto* dépendant de la structure du marché et qui contribue à façonner l'organisation du travail, en reprenant ici le travail de Chassagnon (2014). Ces éléments nous conduisent à faire deux propositions théoriques concernant les études sur les plateformes, sur les chaînes de valeur et plus généralement sur l'émergence de formes hybrides de firmes : l'existence de "réseau hiérarchique étendu" et l'application d'un pouvoir hiérarchique "de facto" basé à la fois sur les outils de coordination numériques et sur la structure post-coloniale de la chaîne de valeur.

139. Nous verrons également que ce n'est pas aussi simple que cela dans le chapitre 4

Conclusion de la partie 2 - La mise en place d'une division internationale du travail de la donnée

Dans cette partie, nos questionnements étaient initialement ancrés dans des interrogations empiriques portant sur comment les entreprises françaises externalisent le travail des données. Se comportent-elles comme les chercheurs en sciences informatiques ou en psychologie lorsqu'il s'agit d'utiliser les plateformes ? Observe-t-on des tendances similaires dans le monde francophone ainsi que celles observées dans le monde anglophone ? Deux hypothèses ont ensuite guidés notre cadrage théorique : *l'usage des plateformes est contraint par des enjeux liés à la qualité des données, plus déterminants dans le cas des entreprises d'IA et les entreprises d'intelligence artificielle emploieront des travailleurs plus intégrés dans leurs processus de production, ce qui leur donnera un plus grand contrôle sur la qualité du processus et les compétences des travailleurs.*

Nous avons proposé un cadrage fondé sur l'analyse des organisations hybrides entre le marché "pur" et la hiérarchie ancrée dans la sociologie économique. Pour nous, ce débat académique est essentiel de par sa contribution à la compréhension du phénomène des plateformes numériques de travail et des débats sous-jacents sur le statut et les conditions de travail des travailleurs des plateformes (Régniez et al., 2020). Au-delà de la situation des travailleurs, ce débat sur la structure de la chaîne de valeur de l'IA est également pertinent pour mieux saisir les mécanismes qui guident la fabrication des modèles d'IA, un thème essentiel de la sociologie de l'IA portant sur les biais algorithmiques et l'opacité algorithmique. Dans ce chapitre, nous avons donc montré comment se structure le processus de production de l'intelligence artificielle en nous appuyant sur trois thématiques de recherche : la localisation de la chaîne de valeur, son organisation et le degré d'intégration des travailleurs des données.

Des firmes pour les gouverner toutes : localisation de la chaîne et structure du pouvoir entre les acteurs

En termes de localisation de la chaîne de valeur, nos résultats permettent d'identifier d'autres flux géographiques que ceux mis en avant dans la littérature. Ces flux d'échanges intègrent à la fois des entreprises de la Silicon Valley, parfois parmi les plus importantes, des entreprises françaises d'IA (startup et grands groupes) et des entreprises d'annotations situées à Madagascar. Alors que les liens entre les États-Unis et l'Inde de l'autre ont été largement mis en avant dans la littérature, nous confirmons le travail de (Anwar & Graham, 2019) qui indique que les pays du continent africain ont une place importante dans le *freelancing* et le travail des données. Nous montrons par ailleurs qu'il existe d'autres organisations du travail que les plateformes numériques de travail, autrement appelé *travail profond* dans la littérature.

L'enjeu de l'équilibre des pouvoirs qui a été souligné par Tubaro (2021) est ici lié à la fois à la structure du marché et à l'organisation du travail de la donnée. Nous montrons ainsi comment le marché de l'IA se

structure autour de firmes hégémoniques (*lead firms*) qui maîtrisent l'ensemble de la chaîne en concentrant des actifs spécifiques déterminants. L'infrastructure matérielle permet de proposer de l'hébergement de données et de la puissance de calcul. Quant à l'infrastructure logicielle, elle consiste à vendre des modèles pré entraînés, des outils d'annotations et l'accès à des services d'annotations¹⁴⁰. Ces services sont ensuite vendus aux startup françaises, qui, quant à elles, font de l'intégration de modèle en réentraînant les modèles sur les cas d'usages de leurs clients (*fine tuning*), en faisant appel à des sous-traitants situés à Madagascar. Dès lors, le déséquilibre des pouvoirs se joue sur la dépendance des entreprises de sous-traitance vis-à-vis des clients français, notamment sur les projets à plus long terme. *In fine*, les remous de la *French Tech* et de la *Silicon Valley* se font ressentir avec force jusqu'à Antananarivo, ce que nous verrons dans notre dernière partie.

Une première implication de ces résultats réside dans la difficulté à dévoiler l'opacité "organisationnelle" de la chaîne de valeur de l'IA. En effet, parler d'une IA véritablement éthique impliquerait de pouvoir tenir compte de l'ensemble des processus de développement qui contribuent au modèle final, notamment lors des phases de production de données. Il faut aussi préciser qu'il est compliqué de véritablement auditer l'ensemble de la chaîne de production : comment avoir accès aux modèles et architectures pré-entraînées développées par les grandes entreprises du numérique et utilisées par les entreprises d'IA françaises ? Comment accéder aux décisions prises lors des phases d'annotations ? Comment documenter ces décisions qu'elles soient prises lors de la création des modèles pré-entraînés des grandes entreprises numériques ou bien lors de la création des modèles finaux ?

Ces résultats illustrent également la position subalterne des entreprises françaises dans la chaîne de valeur. De manière schématique, ces dernières s'appuient d'un côté sur les technologies des grandes entreprises numériques et de l'autre sur la force de travail des travailleurs d'une ancienne colonie. Ce résultat a de fortes implications quant à la construction d'une souveraineté numérique européenne et notamment sur l'IA aujourd'hui portée par les régulateurs. En effet, la position des entreprises françaises leur demande encore davantage d'efforts de traçabilité, puisqu'ils ne maîtrisent pas la production des modèles pré entraînés sur lesquels ils basent leurs modèles. Par ailleurs, l'externalisation du travail de la donnée constitue également un facteur de risque en matière de biais algorithmique.

De la distribution du travail de standardisation du travail d'annotation à l'impossibilité de "débiaiser" des décisions humaines

Dans cette partie, nous avons également étudié la manière dont se constituent les standards d'annotation en reprenant le concept d'objet-frontières (Bowker & Star, 1999). Ces objets émergent de coopérations du-

140. C'est notamment le cas pour Google et Amazon qui proposent d'accéder à des services d'annotations de données directement depuis leurs services.

rables entre différentes communautés de pratiques. Dans deux articles, [Miceli et al. \(2020b\)](#); [Miceli, Yang, et al. \(2022\)](#) évoquent également les guides d'annotations comme le résultat d'une "imposition" de la part des donneurs d'ordres. Nous avons approfondi cet aspect dans notre chapitre 5 en montrant comment les travailleurs des données participent à la définition de ces objet-frontières en contribuant à problématiser le modèle. Pour [Bowker et Star \(1999\)](#), les "*objets frontières naissent au fil (...) d'une coopération durable entre les communautés de pratique. Il s'agit d'accords de travail qui résolvent les anomalies de standardisation sans imposer une standardisation des catégories provenant d'une communauté ou d'une source extérieure de normalisation.*" Le fait de considérer les guides d'annotations et les modèles comme des objet-frontières prend donc tout son sens, dans la mesure où l'on constate qu'ils agrègent différentes groupes professionnels. Les clients finaux qui ont une expertise du terrain, les entreprises d'IA qui ont une expertise informatique et les annotateurs qui ont une expertise sur l'annotation, sur le terrain et sur la manière dont le modèle fonctionne à travers les guides d'annotations. Dans ce processus de "*classification*", le défi des entreprises d'IA est de maîtriser la subjectivité dans un processus de traduction du contexte du client final (utilisateurs de l'IA), en un SIA. Ces groupes "*de pratique*" décrites par [\(Bowker & Star, 2000\)](#) (p.297) travaillent en fait ensemble pour "*maintenir la cohérence*" du modèle et de la classification sous-jacente. C'est alors la maîtrise de ce processus de standardisation, et, d'une certaine manière *la maîtrise de la subjectivité des travailleurs* qui justifient aussi de les intégrer logistiquement et rationnellement à travers différents outils de communication et de contrôle.

Nous avons également abordé cette question du maintien de la cohérence à travers les variations des modèles (sa volatilité) en fonction de l'évolution des décisions de développement (utiliser une approche du problème plutôt qu'une autre), des remontées des annotateurs sur les données (les ambiguïtés et les imprévus) et les évolutions de la vérité terrain (lorsque le modèle est déployé en production). Nous montrons ainsi, que ce processus est très *incrémental*, compte tenu de l'instabilité des modèles d'IA et donc du processus de développement [\(Seaver, 2017\)](#). D'après nous, c'est ce qui explique que la relation avec les annotateurs de données est envisagée à long terme. Les travailleurs des données opèrent comme des *stabilisateurs des modèles d'IA*. Par ailleurs, la compréhension des modèles de *machine learning* passe par l'étude de la "*composition des données d'entrée, [de] l'architecture choisit par les différents systèmes et [des] objectifs qui ont été donnés à la supervision de l'apprentissage*" [\(Cardon et al., 2018\)](#). Dès lors, il paraît logique pour les développeurs d'IA de maîtriser le plus possible la chaîne de production des données annotées, c'est ce que nous avons appelé le *couplage modèles-données* qui se traduit du point de vue de l'organisation du travail par le *couplage développeurs-travailleurs des données* et donc par l'intégration des travailleurs.

Nous contribuons ainsi à la littérature en sociologie en ajoutant plusieurs éléments "d'opacités" et surtout plusieurs phases "non-techniques" de la production qui méritent d'être analysées. Ainsi, la "volatilité algorithmique" ne se matérialise pas uniquement par l'architecture propre au *machine learning* et à des boîtes

noires, mais plutôt par l'implication de nombreux acteurs qui façonnent à différents degrés le modèle final et ce de trois manières : la *volatilité des décisions algorithmiques*, la *volatilité des données existantes* et la *volatilité de la vérité terrain*. La nature de ce processus implique alors une organisation basée sur la qualité. Le corollaire de ce mode de fonctionnement de l'IA est, comme on a pu l'observer à travers l'exemple de la surveillance des parkings, qu'il paraît peu probable de produire des modèles non biaisés tant ils sont le résultat d'une chaîne de production impliquant de nombreuses couches de travail humain (F. A. Schmidt, 2019). L'enjeu de notre prochain chapitre sera justement d'analyser les pratiques de travail de ces différentes couches de travailleurs, en insistant sur le rôle des travailleurs des données.

Derrière le caractère dynamique du processus de production de l'IA, la forte intégration des travailleurs des données

Nous avons également analysé le processus de production de l'IA au sein des entreprises françaises et constaté son caractère *incrémental* avec des allers-retours réguliers entre les différents acteurs de la chaîne de valeur, ce qui justifie le recours à une force de travail stable. Nous avons également proposé une analyse des choix d'organisation du travail de la donnée de ces entreprises, au-delà de la dimension *dynamique* de la production des modèles basés sur la *machine learning*. La nature du projet d'intelligence artificielle constitue un élément décisif dans le choix de l'organisation à mettre en place : est-il complexe de par le type d'annotation ou le type de données à annoter nécessitant alors plus d'expertise ? A-t-il vocation à durer dans le temps du fait du volume des données à annoter ? Ce dernier point étant lié à la variété des données utilisées pour le modèle (c'est l'exemple de la nourriture) et aux objectifs de performance du modèle (c'est le par exemple : appliqué à la santé).

Un autre aspect concerne le type d'entreprise qui développe le modèle. S'agit-il d'une petite startup dont l'IA est au coeur de la proposition de valeur et qui réalise des produits déployés en production ou bien d'une grande entreprise qui n'est pas forcément issue du secteur technologique et qui se lance dans des projets d'IA dans le cadre de ses activités de recherche et développement ? Dans cette perspective, nous proposons des pistes de travail à travers le tableau ci-dessous.

Type d'entreprise	Type de données	Financement	Maturité produit
Agences d'IA	Variées	Levée de fonds	Moyenne
Entreprises IA spécialisées	Spécifiques	Levée de fonds	Moyenne
Grandes entreprises	Spécifiques	Autofinancement	Faible
Grandes entreprises numériques	Variées	Autofinancement	Importante

TABLE 8 – Typologie des entreprises d'IA selon leur besoin en matière d'annotation (d'après l'auteur)

Au travers de ces exemples, nous avons également constaté la manière dont les travailleurs sont intégrés d'un point de vue relationnel et logistique au processus de production des firmes, à tel point que certains donneurs d'ordres ont tendance à considérer ces travailleurs comme leurs propres employés. Par rapport au cas de l'industrie manufacturière, il est important de préciser qu'ici, ce ne sont pas les données qui sont vendues (biens), mais bien le travail des données (services). Cet aspect sera d'ailleurs approfondi dans notre prochain chapitre.

D'un point de vue théorique, nous contribuons donc à l'analyse des relations inter-firmes et intra-firmes dans le secteur de l'externalisation de services numériques et plus particulièrement dans celui de l'intelligence artificielle. Notre terrain d'étude illustre le concept d'*intégration* des travailleurs dans des réseaux de sous-traitance. Ce faisant, nous participons à l'étude de la plateformes du travail. En effet, comme on l'a vu dans notre première section, les études abordant le freelancing et le micro-travail s'appuient sur les théories des chaînes globales de valeur afin de questionner les frontières des firmes et les effets de ces nouveaux objets "hybrides" sur les travailleurs. Or, ces résultats illustrent le recours à des formes hiérarchiques de chaîne de valeurs, ce qui questionne le caractère *exceptionnel* d'objets comme les plateformes et les firme-réseaux. Quel sens donner à ces résultats qui diffèrent de ce qui a pu être observé dans la littérature existante ? Nous constatons ainsi qu'un certain nombre de facteurs conduisent aujourd'hui les acteurs de l'IA à transformer le modèle de travail platformisé qui existe depuis le début des années 2000 en un modèle qui intègre des sous-traitants classiques dans la boucle. Ainsi, en mettant en place des dispositifs de communication directs entre entreprise d'IA et annotateurs, ces dernières créent les conditions pour que s'établissent des échanges sur le design des tâches ou bien sur les concepts mobilisés par les data scientist dans leurs modèles. Il peut également arriver que certains types de projets nécessitent différents types d'annotations. Une hypothèse est que certains systèmes transitoires peuvent très bien fonctionner avec des ensembles de données étiquetés et anonymisés, tandis que les systèmes complexes à long terme nécessitent une main-d'œuvre stable et bien établie. Nos résultats peuvent également être lu comme l'émergence de "hiérarchies électroniques" de par les "plateformes" comme outils de coordination et non de mise en marché. Les entreprises dépendent alors encore davantage de "*ce mode d'intégration, qui offre une nouvelle architecture pour les stratégies transnationales de gouvernance organisationnelle*" (Aneesh, 2006) (p.75).

Ni marché, ni hiérarchie ? Les plateformes de travail comme firmes-hybrides ?

L'analyse des "réseaux virtuels de production" (*virtual production network*) réalisée par Wood et al. (2019b) apporte une autre vision du débat sur l'intégration des travailleurs, sous l'angle de l'*encastrement* des travailleurs des plateformes. Ils reprennent les apports de Graham et Anwar (2019), qui montrent que les travailleurs des plateformes sont placés en situation de compétition, ce qui nécessite une adaptation aux besoins des clients, notamment en termes d'horaires de travail. Ils s'appuient également sur une division de la notion d'*encastrement*. Ainsi, ils constatent l'inscription des travailleurs dans des réseaux d'entraide

locaux (amis, familles) (*soft embeddedness*). Les réseaux de production plateformisés ont toutefois la spécificité d'être désencastrés, puisque les travailleurs ne sont pas intégrés dans les normes de travail de leur pays de résidence (droit du travail) et que la répartition des tâches est basée sur l'offre et demande.

Par ailleurs, dans un article sur l'organisation de la plateforme Uber en Allemagne, [Howson, Ferrari, et al. \(2022\)](#) suggèrent d'ailleurs d'analyser les plateformes numériques de travail comme des "firme-pivots" utilisant la technologie et une organisation en réseau pour capter la valeur produite par les chauffeurs plus efficacement. En contrôlant, en standardisant et en coordonnant toute une "myriade" de firmes (les chauffeurs), elles contribuent à la distribution inégale des ressources produites tout en obscurcissant toujours plus la frontière entre le "marché" et la "firme", comme en témoignent les débats autour de la qualification du contrat de travail des chauffeurs Uber ([Howson, Ferrari, et al., 2022](#)).

Dans le cas allemand, il est toutefois important de noter qu'Uber a ajusté son modèle d'affaires en fonction d'une décision de la Cour fédérale de justice allemande, en mettant en place une sous-traitance en cascade comprenant une firme focale qui gère un réseau de firmes régionales qui ensuite externalisent à des petites entreprises privées qui salarient leurs chauffeurs et à des chauffeurs indépendants. Selon l'environnement réglementaire des pays et les décisions de justice, les plateformes ont dû adapter leurs modèles d'affaires, tout en conservant ce principe de standardisation de la prestation par la technologie et des règles managériales. Ce qui traduit dans la littérature une forte inquiétude concernant la forme de contractualisation de ces travailleurs et l'accès à des conditions de travail protectrices alors que ces derniers exercent en tant d'indépendants ([Berg et al., 2018](#); [Graham et al., 2017a](#)).

Dans notre travail, nous proposons une analyse des relations entre les acteurs de la chaîne et nous montrons que, malgré des formes d'organisations qui se présentent comme relevant du marché, les relations de pouvoirs entre les acteurs pointent plutôt vers des relations "hiérarchiques", dont le degré va différer selon les modèles d'organisation. Pour nous, cette idée va dans le sens de la proposition théorique de [Tubaro \(2021\)](#), qui analyse la structure organisationnelle des plateformes de micro-travail sous l'angle du concept "d'encastrement". Dans le secteur du micro-travail, elle postule l'existence de deux types de plateformes : les "*plateformes comme places de marché*" et les "*plateformes comme firmes*". Au sein des premières, les travailleurs ont des liens économiques "minces" et ils sont très autonomes "*dans une forme de désencastrement*". Les secondes "*participent à des structures complexes d'externalisation interentreprises*" qui impliquent un "encastrement profond" pour les travailleurs (Appen). Loin d'être un "objet unique", on retrouve l'idée que les plateformes sont très variables dans leurs modes d'organisation du travail. Par ailleurs, en parlant d'encastrement profond pour caractériser les multiples couches d'intermédiation qui caractérisent les plateformes du second types, elle rejoint également certains arguments avancés par les sociologues de l'IA sur l'opacité organisationnelle qui entraîne une opacité du fonctionnement des algorithmes d'IA ¹⁴¹

141. "As in deep learning, some layers are 'hidden', meaning that the whole structure is hard to see, especially in corporate nets that

Les concepts théoriques abordés dans l'introduction de cette partie pointent vers le développement de firme-réseaux hybrides qui se distinguent par leur capacité à coordonner un réseau de sous-traitant. Notre terrain illustre pourtant, qu'à travers les différentes organisations du travail de la donnée, on perçoit plutôt le développement de formes de "hiérarchie électronique étendue" (Fréry, 1997, Malone, 1987). Ainsi, la numérisation des relations intra-firmes ajoute des outils supplémentaires de contrôle de la force de travail qui vont bénéficier à la firme focale, qui, dès lors, pourra se comporter comme si les travailleurs étaient ses propres employés.

Ce résultat prolonge un argument que nous avons déjà défendu dans un article consacré à l'hétérogénéité des plateformes et des modes d'externalisation dans l'IA (Cornet et al., 2022). De notre point de vue, cet intérêt pour ce nouvel "objet", en tant que modèle typique d'organisation masque le fait que les plateformes ne constituent pas un groupe homogène d'organisations, que ce soit en termes de secteur d'activité, de définition du statut juridique et plus généralement de structures organisationnelles. Dans cet article, nous montrons donc l'intérêt de l'analyse de l'objet "plateforme" à un niveau sectoriel, en tenant compte des spécificités propres au secteur d'activités. C'est dans ce sens que, dans un article (Le Ludec, Cornet et Casilli, 2022), nous analysons l'IA comme un outil de coordination de forces de travail variées. L'IA serait le résultat de la coordination de ce travail humain. De ce point de vue, la forme hiérarchique est la forme optimum pour favoriser la codification de savoir-tacites¹⁴².

Derrière les faux intermédiaires de marché : des "hiérarchies électroniques" organisant les interactions et le contrôle sur les travailleurs des données

Le résultat détaillé précédemment introduit des éléments empiriques dans le débat très discuté sur l'organisation du travail numérique et qui porte sur la différence entre les plateformes et les entreprises formelles (BPO) (Lehdonvirta et al., 2019). Nous montrons que ces deux modèles cohabitent dans un même secteur d'activité, voire au sein d'une même entreprise. Au lieu d'une distinction entre les modèles, nous montrons qu'il existe plutôt un enchevêtrement avec une primauté aux modèles "hiérarchiques". Les facteurs évoqués dans le contexte français mettent en lumière des éléments d'explications. En effet, pour tenir compte de la vérité terrain et créer des modèles applicables dans le monde réel il est nécessaire de mettre en place des outils de communication et d'avoir accès à une main-d'œuvre stable.

stretch globally. Varied arrangements to source labour may coexist : if platform systems with non-employed, formally independent workers are the most common, some layers may include salaried temporary workers managed in-house or through dedicated suppliers (for example, Appen offers all these options depending on clients' needs). The latter solutions are little visible because they mostly serve to deal with confidential data – so that workers have to sign stringent non-disclosure agreements – or to discreetly outsource part of production to lower-cost locations overseas (...) In this sense, deep embeddedness means that these multi-layer structures integrate labourers and control their activities, while obfuscating their role" (Iubaro, 2021)

142. C'est précisément l'objet de notre prochain chapitre, puisque nous y abordons l'hypothèse de la déqualification du travail sous le prisme du travail réalisé par les travailleurs de Madagascar

Dans le cas d'autres entreprises traditionnellement présentées comme des plateformes numériques de travail, nous constatons également la présence d'entreprises formelles intégrées aux plateformes (Appen) et de travailleurs indépendants fortement intégrés dans le processus de production de la plateforme (EthicAI). D'un point de vue purement empirique, cela confirme les analyses de Tubard (2021) et A. Casilli et al. (2019a) suggérant l'existence de plateformes intermédiaires entre la hiérarchie et le marché et de mécanismes de coordination du travail partagés entre plusieurs entreprises.

Pour nous, cela confirme aussi que le phénomène de fond ne doit pas être analysé uniquement au prisme de l'émergence d'un "marché mondial du travail" (Anwar et al., 2023). Nous y voyons plutôt une extension numérique de la capacité des entreprises occidentales à standardiser leurs opérations et leur travail. De ce point de vue, nous pensons qu'il ne faut pas voir des marchés partout mais plutôt une variété de formes d'intégration des travailleurs entre la hiérarchie et le marché. Néanmoins, l'analyse de nos résultats questionnent ces formes hybrides, notamment si l'on tient compte de la structure du pouvoir dans les entités présentées comme relevant du marché. Nous considérons la forme purement marchande comme une exception, plutôt que comme une tendance de la plateformesisation.

Cela ne veut bien sûr pas dire que nous n'acceptons pas l'ambiguïté organisationnelle que la création de ces organisations de travail provoque aux marges juridiques. Il nous semble cependant nécessaire de reconsidérer l'approche conceptuelle de ces sujets afin de mieux considérer l'hétérogénéité de l'organisation du travail numérique (Cornet et al., 2022), d'une part et par conséquent la fatalité des modes d'intégration des salariés dans les hiérarchies qu'elles soient managées par le client final, par la plateforme ou par une société sous-traitante intégrée à la plateforme. Pour nous, en fin de compte, il n'existe pas de plateformes, seulement des outils qui permettent d'établir une hiérarchie électronique, quelque soit la face du marché qui détient le pouvoir sur les travailleurs.

Ensuite, l'étude de la mise en place des relations entre acteurs à l'intérieur des firmes de la chaîne de valeur montre que même dans les cas qui paraissent se rapprocher du *marché pur*, les entreprises intègrent aussi les travailleurs. Nous nous appuyons pour cela sur l'analyse de Chassagnon (2014) qui indique que les firmes font coexister trois types de pouvoirs : juridique de par la nature du contrat, *de jure* de par la propriété des actifs de la firme et *de facto* de par la dépendance des acteurs vis-à-vis de ressources critiques, notamment les ressources humaines.

Ainsi, les plateformes sont soit de simples outils d'annotations ou de réels outils de coordination du travail. Dans le premier cas, elles délèguent donc le pouvoir juridique et *de jure* aux entreprises de sous-traitance qui partagent le pouvoir *de facto* avec le donneur d'ordre final. Dans le second, elles détiennent le pouvoir juridique et *de facto* et partagent la propriété des actifs avec les travailleurs, ces derniers possédant leur outil de production (un ordinateur) et la plateforme l'outil d'annotation. De notre point de vue, cela

limite l'utilisation même du concept, puisque le terme désigne une organisation du travail qui fonctionne en reliant l'offre et la demande de travail et donc comme un marché. Dans l'une ou l'autre position, les plateformes fonctionnent comme des "catalyseurs de hiérarchie", qui il faut le dire contribue à redistribuer le pouvoir sur les travailleurs.

Quelque soit le mode d'externalisation, les travailleurs sont alors profondément encastrés dans des normes locales de travail. Surtout, comme on le verra dans notre dernière partie, cet encastrement se vérifie dans des conditions de travail largement régies par le droit du travail local. Pour conclure là-dessus, nous considérons que malgré ces formes hiérarchiques, la structure du marché dans lequel ces travailleurs opèrent va déterminer une forme d'emploi flexible, comme nous le verrons dans le chapitre 4. En effet, qu'il s'agisse de travailleurs des plateformes ou de l'IA, ils sont soumis aux aléas des financements des donneurs d'ordres finaux qui se font souvent auprès des marchés (levées de fonds), dans un contexte où le travail est, en théorie, fortement mobile géographiquement¹⁴³. Ce faisant, nous confirmons également l'idée d'un pouvoir *de facto* partagée entre les entreprises intermédiaires du secteur de l'annotation et le donneur d'ordre final Chassagnon (2014). Ces éléments nous conduisent à faire deux propositions théoriques concernant les études sur les plateformes, sur les chaînes de valeur et plus généralement sur l'émergence de formes hybrides de firmes : l'existence de "réseaux hiérarchiques étendu" et l'application d'un pouvoir hiérarchique "de facto" basé à la fois sur les outils de coordination numérique et sur la structure post-coloniale de la chaîne de valeur et non uniquement sur le code source derrière les outils numériques, ce que nous détaillons dans notre dernière partie¹⁴⁴.

143. Nous verrons également que ce n'est pas aussi simple que cela dans le chapitre 4

144. "How can these global teams, with no single bureaucracy to organize them in a unified work space establish a tight work regimen? What managerial hierarchy does the job of integrating or coordinating work teams? These questions about the global regime of virtual migration are relevant to a large transformation of how work is organized in general, even within the national boundary, indeed within the same organization (...) If power and authority relations organized around bureaucratic hierarchies are on the decline, where do we locate new relations of power presupposed in organized work, which is, as Kunda (1992) maintains, fundamentally shaped by processes of control? My answer is quite simple : code" (Aneesh, 2006) (p.102)

Partie 3 - L'activité réelle des *petites mains* de l'IA : vers la solidification d'un groupe professionnel de travailleurs de l'IA

Dans notre précédente partie, nous avons montré que les entreprises d'IA organisent la chaîne de production en intégrant davantage les travailleurs pour des raisons de qualité. Cette dimension organisationnelle se retrouve également au niveau de la qualification des travailleurs des données. Sur ce point la, littérature en économie présente l'intelligence artificielle (IA) comme une technologie de changement structurel pour le travail. La capacité de cette technologie à transformer et à supprimer des emplois fait l'objet d'une controverse toujours vivace. Cependant, il faut se garder de l'écueil du déterminisme technologique et d'attribuer aux technologies des effets inéluctables en matière d'organisation du travail (Benedetto-Meyer & Boboc, 2021) (p. 26). La tension entre la dimension technique et celle sociale et organisationnelle est d'ailleurs au coeur des premières controverses sur les effets de l'automatisation sur le travail¹⁴⁵. Il en ressort un débat jamais vraiment clos sur les "transformations du travail ouvrier et l'organisation du travail en lien avec les techniques" (Benedetto-Meyer & Boboc, 2021) (p. 21).

Dans cette troisième partie, nous proposons d'analyser le développement du travail des données à Madagascar au regard d'une dynamique plus générale de transformation du travail liée à l'IA. Ce faisant, nous souhaitons réinscrire notre objet de recherche dans un ensemble de débats scientifiques sur les effets des nouvelles technologies sur le travail. Plus précisément, ce sont les effets sur la qualité des tâches au travail et, en retour, sur la qualification des travailleurs qui seront examinés.

Partant du constat que les travailleurs interrogés lors de cette enquête ont des expériences professionnelles qui les conduisant à exercer de nombreuses activités liées à l'externalisation de services, nous nous appuyons sur le travail de SHESTAKOVSKY (2015), afin d'analyser ces activités externalisées sous l'angle de la continuité du travail humain lors du déploiement des technologies numériques. Nous nous intéressons alors aux différentes activités de l'externalisation à travers l'étude de la continuité du travail humain dans des activités "automatiques". Nous y examinons le travail de l'externalisation en nous demandant : est-il si "facile" que "tout le monde peut le faire" ? Nous montrons d'abord que la massification des usages numériques, loin de supprimer le travail humain, crée de nouveaux besoins. Les travailleurs contribuent en effet de plusieurs manières à l'automatisation en réalisant de nombreuses *tâches palliatives* aux services numériques automatiques. Ce faisant nous montrons que le travail d'automatisation s'inscrit dans un continuum

145. Nous précisons ici au lecteur que nous utilisons automatisation et automatisme indifféremment

de tâches apparues dans le sillage de la numérisation.

Par ailleurs, ces tâches sont parfois centrales dans la production de services basés sur l'IA, notamment dans le cas de l'annotation de données. Ces tâches requièrent un travail de résolution de problèmes et d'interprétations. De ce point de vue, notre travail se situe également dans la continuité des travaux sur les algorithmes. Le travail de [Muniesa \(2003\)](#) sur les *configurations algorithmiques* apporte une deuxième piste de questionnement. Selon lui, le fonctionnement des algorithmes est localisé, car les règles qui régissent leur fonctionnement reposent sur des conventions entre les acteurs qui participent à la production du dispositif. Appliqué à notre thèse, cela revient à se demander quels sont les travailleurs qui contribuent à produire les règles de production de la signification des modèles?

Ces conventions ne seraient pas possibles sans un travail de production de sens. [Bowker et Star \(1999\)](#) définissent la classification comme *"une segmentation spatiale, temporelle ou spatio-temporelle du monde"* (p.10) et considèrent qu'il ne doit pas exister d'ambiguïté dans la classification. Dans le contexte de l'IA, ce sont les annotateurs qui réalisent ce travail. Ainsi, un travailleur qui traite des images d'animaux doit pouvoir les ranger dans des catégories qui ne se recoupent pas. Mais, ce qui paraît évident pour ce thème, l'est peut-être déjà moins pour des images de vêtements : comment distinguer indubitablement certains types de cols, ou certains types de pantalons? À partir de cette l'analyse de l'annotation comme opération de classification du réel, nous proposons d'analyser les pratiques de travail et ce qu'elles racontent des qualifications nécessaires. Se pose donc la question de comment sont façonnées ces catégories. Toujours selon [Bowker et Star \(2000\)](#), il est nécessaire que le système soit complet, ce qui signifie que la classification fournit une *"couverture totale du monde qu'il décrit"*. Comme on le verra, il est fréquent que les classifications utilisées par les annotateurs pour réaliser leur travail contiennent des ambiguïtés ou des manques. Dès lors, que font ces travailleurs avec ce qui est *"laissé dans l'ombre"* par les classifications de ceux qui produisent les modèles? Que faire avec des données qui sont bien souvent abandonnées dans la catégorie "autre"? ¹⁴⁶

Dans ce chapitre, nous proposons donc d'articuler le problème de la qualification du travail des données avec une réflexion sur la construction des classifications qui sous-tendent l'intelligence artificielle. Ainsi, à l'aide des entretiens réalisés auprès d'employés d'entreprises d'IA, de clients de ces entreprises et de travailleurs des données, nous approfondissons le processus de production de l'intelligence artificielle, les interactions entre les acteurs de la chaîne et les pratiques d'annotation. *L'hypothèse principale derrière ce chapitre est que les travailleurs contribuent de manière substantielle à la problématisation du modèle durant leur travail de classification des données.* En effet, en développant des savoir-tacites sur les données, ils participent à

146. *"No matter how good the scheme, its scope is limited by the fact that data entry is never an easy task, and there are never enough resources or trained personnel to make it happen. Not only will there inevitably be mistakes with respect to the internal structure of whatever classification one is representing, there will also inevitably be cultural variations with respect to how it is interpreted as well as culturally biased omissions."* ([Bowker & Star, 2000](#)) (p.107)

leur interprétation et dépassent le travail "prescrit" par la chaîne hiérarchique. Le modèle d'IA est alors le résultat de ce travail collectif. Ce sera l'objet du dernier chapitre de cette partie.

Questionner le développement de l'intelligence artificielle revient par conséquent à analyser son processus de production, son organisation et les structures dans lesquels ce marché s'inscrit. Le travail d'annotation n'est pas seulement le reliquat humain des technologies automatiques, il est aussi et surtout partie intégrante d'un processus de production dont la finalité est la vente d'un produit ou d'un service. Plus que les consommateurs soumis aux mécanismes d'extraction de valeur des entreprises du numérique, ces travailleurs façonnent, fabriquent et complètent les processus automatisés. L'enjeu de cette partie consiste à analyser les effets de la phase actuelle d'application des technologies automatiques numériques. Cela passe par l'étude d'un autre terrain d'enquête, qui nous a conduit à Madagascar : au sein des entreprises de services qui vendent des prestations en annotation de données. En dehors des tâches d'annotations, si souvent étudiées dans la littérature sur le micro-travail, notre matériau fait entrer le lecteur dans les organisations du travail, les chaînes hiérarchiques ainsi que les processus de production structurés qui permettent de fabriquer l'intelligence artificielle.

Chapitre 6 - La délocalisation de services pour pallier les limites de l'automatisation algorithmique

Avant de présenter nos résultats, revenons sur les travaux relatifs aux effets de l'automatisation. D'abord, la sociologie du travail française qui oppose, dès son origine, deux tendances, respectivement portées par Georges Friedmann et Pierre Naville : la disparition du travail taylorisé (requalification) et le développement d'emplois non qualifiés (déqualification).

Georges **Friedmann** (**1955**) imagine la disparition de ce qu'il appelle le "travail en miettes", l'automatisation favorisant l'émergence d'emplois plus qualifiés ¹⁴⁷. Michel Freyssenet reprend cette hypothèse à travers l'étude de la comptabilité. Selon lui, l'apparition de machines de calcul opérées par des travailleurs moins qualifiés (par exemple : les secrétaires) renforce la division verticale du travail. Il observe une spécialisation des opérateurs qui manipulent ces machines à calculer ¹⁴⁸. Pour lui, l'arrivée des systèmes informatiques permettrait de regrouper certains emplois autrefois divisés. Par conséquent, les postes de travail nécessitent de maîtriser davantage de tâches et s'accompagnent donc d'un plus haut niveau de compétences (*requalification*). Néanmoins, Georges Friedmann reviendra plus tard sur cette hypothèse en déclarant que : *"l'observation des faits ne permet pas de partager l'optimisme de ceux qui (...) voient dans un proche avenir la suppression de toutes les tâches éclatées, dégradées, et leur revalorisation"*. Pour lui, *"le jugement, l'adaptation à des circonstances variables, le réglage, sont en effet de plus en plus complexes. Mais, au-delà d'un certain point de flexion, c'est la machine qui absorbe ces qualités dans son automatisme et en dépouille l'opérateur"*.

Les études de Pierre Naville considèrent à l'inverse que l'automatisation risque d'entraîner le développement d'emplois non qualifiés, renforçant alors la distinction entre les ouvriers qualifiés et les opérateurs "simples". Les premiers sont des ouvriers professionnels (OP) formés à leur métier et capable de comprendre le processus de production dans son entièreté, tandis que les seconds sont des ouvriers spécialisés (OS), ayant des compétences plus spécifiques et mono-tâches comme la surveillance du bon fonctionnement d'une machine. Dans le cas du secteur textile, cela se traduit par une hausse du nombre d'ouvriers professionnels du fait de l'augmentation de l'intensité du travail permise par l'automatisation. En revanche, les métiers initialement les moins qualifiés comme les "bobineuses" qui consistaient jusqu'alors à mener et surveiller le bobinage des fils, se rapprochent davantage d'un travail d'approvisionnement et de déchargement des machines textiles (*déqualification*). Il observe donc, d'un côté la *déqualification partielle* de certains emplois, et de l'autre une *polarisation globale des emplois* avec l'émergence d'emplois requalifiés. Ainsi, le travail ne

147. *"Si... l'automation était destinée à se propager dans l'industrie et dans les bureaux à un rythme foudroyant, si son emploi généralisé était pour demain, les tâches en miettes seraient, dans leur quasi-totalité, supprimées à bref délai et les problèmes que nous examinerions à leur propos se trouveraient d'ores et déjà périmés..."* (**Friedmann, 1955**) (pp15-16)

148. *"les uns ne traitent que les chèques, les autres que les factures clients, d'autres encore que les factures fournisseurs ou la paye, etc. La variété relative de leur travail disparaît."* (p. 64)

disparaît pas, il ne se transforme pas nécessairement du fait de la technique, mais il se déplace progressivement vers l'intégration d'autres qualifications et compétences.

Les études sur les industries de process telles que l'industrie du papier, de l'agro-alimentaire ou des produits pharmaceutiques viennent nuancer ces résultats. On y retrouve une division des tâches en trois types : le lancement et l'arrêt des cycles de production, la surveillance de ce cycle par des appareils de mesures et la régulation du processus. Il s'agit donc d'un travail effectué à distance, puisque les manipulations doivent se faire dans des conditions de sécurité suffisantes étant donné la nature des matières premières utilisées. La dangerosité et la continuité du processus de production vont également nécessiter autant de tâches de lancement, d'arrêt de cycle et de surveillance de la production. Le niveau de production dépend donc, non pas du temps passé par les ouvriers sur les machines, mais de la qualité et la quantité des produits fabriqués, ce qui rend le travail de surveillance particulièrement déterminant. Les collectifs de travailleurs ouvriers doivent éviter au maximum les interruptions de production, tandis que les travailleurs qui conçoivent les machines doivent les rendre le plus fiables possible¹⁴⁹. Deux phénomènes sont alors constatés : une hausse des qualifications, observée par l'équipe de Pierre Naville et un "*nivellement des qualifications*" observées par Claude Durand (1966).

Dans ce contexte, Pierre Naville observe l'émergence de "*nouveaux postes*" pour "*l'interprétation des signaux*." (Du Tertre & Santilli, 1992). Ces postes nécessitent de surveiller les signaux liés aux défaillances éventuelles de la chaîne de montage et d'adopter la bonne réaction en cas d'anomalie¹⁵⁰. Cet aspect du travail explique que ces postes soient considérés comme étant à forte responsabilité et qu'ils soient donc occupés par des ouvriers très qualifiés. Il observe également le développement d'une force de travail plus qualifiée, dévolue à l'entretien de machines devenues de plus en plus complexes avec l'automatisation et nécessitant parfois des compétences plus sophistiquées (par exemple : en optique ou en électronique).

Claude Durand (1966) observe quant à lui l'émergence de deux types d'opérateurs : les opérateurs de commande et les opérateurs de contrôle. Les premiers travaillent directement à la fabrication des produits. Leurs profils sont aussi variés que les machines à utiliser pour la fabrication et peu d'entre eux sont polyvalents. Outre l'habileté à manier les commandes de la machine, ils sont confrontés aux différences entre les schémas de travail conçus par les bureaux de méthodes et la réalité du terrain, ce qui les oblige à être autonomes et très réactifs. Ils développent ainsi des "*savoirs tacites*" qui sont parfois mobilisés durant la conception de nouvelles machines, à laquelle certains contribuent, en collaboration avec les ingénieurs. Les opérateurs de contrôle interprètent les appareils de mesure depuis les salles de contrôle pour éviter les défaillances, ce qui

149. Finalement, "*les capacités productives des collectifs ouvriers de production qui dépendent de leur capacité à faire face aux événements aléatoires, d'interpréter les pannes, d'anticiper les réactions du processus.*" ((Du Tertre & Santilli, 1992), p.31)

150. Ils nécessitent toute une série de tâches complexes : "*respect des consignes, contrôles répétés et eux-mêmes soumis à contrôles, surveillance continue, vigilance pour déceler toute anomalie ou accident dans la marche du process, liaison et coordination avec les ouvriers d'entretien*" (Du Tertre & Santilli, 1992)

leur donne d'importantes responsabilités. Selon lui, le remplacement des ouvriers spécialisés par des opérateurs plus qualifiés va permettre à ces derniers des progressions de carrière, notamment lorsqu'ils rejoignent les bureaux de méthode afin de participer à la conception des processus de travail. Les opérateurs les moins qualifiés sont souvent délaissés par leurs entreprises, finissant même parfois par quitter l'entreprise. Claude Durand parle donc d'un *nivellement des compétences*.

Beaucoup plus critique des conséquences de l'automatisation en cours à cette époque, Serge Mallet (1969) analyse le développement d'une "nouvelle classe ouvrière", capable d'occuper des emplois automatisés, qui s'oppose à un sous-prolétariat, chassé de l'emploi par les machines¹⁵¹. Tout d'abord, son analyse de l'état de l'automatisation en France fait ressortir les limites de la substitution capital/travail. Les maigres salaires des populations colonisées ou anciennement colonisées d'alors permettent au patronat de bénéficier d'une main-d'oeuvre abondante au prix le plus bas possible. Ainsi, l'automatisation est dépendante d'un arbitrage entre le coût des machines et le coût des travailleurs disponibles. Serge Mallet différencie également l'automatisation de la parcellisation du travail, selon les transformations du rôle des ouvriers. Ces derniers deviennent des "opérateurs", car leur travail consiste principalement à surveiller des installations de production et non plus à produire. D'un côté, on retrouve les *"surveillants, chargeurs, opérateurs, préparateurs affectés aux unités de production automatisées"*. De l'autre, il s'agit des ouvriers d'entretien *"chargés de la réparation et de la surveillance des mécanismes, de l'outillage."* Les premiers se caractérisent par de fortes capacités d'attention : *"Ils sont les correcteurs humains des possibles défaillances de la machine."*, donc *"leur poste exige une connaissance assez complète du processus de production dans lequel ils s'insèrent"*. Les seconds sont des ouvriers plus traditionnels comme les mécaniciens ou les réparateurs qui ont plus *"plus de responsabilités que les opérateurs"*, car ils doivent s'assurer du bon fonctionnement de la chaîne de production.

Dans un article, Paul Attewell (1987) revient sur l'hypothèse formulée par (Braverman, 1998) à propos de la tendance du *"capitalisme à continuellement abaisser les niveaux de compétences dans l'ensemble de l'économie américaine"*. Pour (Braverman, 1998), il existe plusieurs implications organisationnelles du capitalisme qui ont pour conséquence la déqualification progressive de la force de travail. La première reprend les analyses d'Adam Smith, qui explique la division du travail par la nécessité de gagner en productivité se traduisant par une distinction des tâches effectuées par des travailleurs séparés. Mais, selon Braverman, cette spécialisation des travailleurs n'est pas nécessaire. Une deuxième implication, formulée par Charles Babbage, *"suggère qu'il existe un impératif pour le capitalisme de diviser des tâches complexes en étapes simples et standardisées"*

151. Il considère toutefois que l'automatisation est retardée, en France, par « un patronat peu assuré de ses arrières financiers et commerciaux (qui en) retarde au maximum la généralisation » et, que son « introduction en France sera toujours freinée tant que (le patronat) trouvera des masses de colonisés de l'intérieur ou de l'extérieur, des Portugais ou des Bretons, des Algériens ou des Flamands, des laissés pour compte de l'enseignement technique et des paysans chassés par la concentration des terres agricoles, pour (fournir) une "réserve" d'esclaves sans qualification, toujours assez bons pour le travail à la chaîne et qui ont d'autant plus de mal à défendre leurs maigres salaires qu'ils se savent menacés en permanence par les éternels dangers d'une "modernisation" dont ils feront les frais. » (S. Mallet, La nouvelle classe ouvrière, Ed. du Seuil, 1969, p. 11).

et d'embaucher une force de travail non-qualifiée et bon marché" pour réaliser ce travail Attewell (1987).

Braverman apporte une explication supplémentaire relative au contrôle de la force de travail. Les managers ayant pour objectif de maximiser le contrôle sur les travailleurs en renforçant leur connaissance de la production réduisent "les travailleurs à de simples exécutants du travail". Cela accroît la dépendance des travailleurs vis-à-vis des managers, qui se sont arrogés l'expertise sur le fonctionnement de la production. Dès lors, le travail se "vide de sens" et conduit les travailleurs à l'aliénation¹⁵². Cela entraîne un processus de déqualification des travailleurs, progressif et général. Les travailleurs des secteurs touchés par l'automatisation rejoignent progressivement d'autres secteurs qui requièrent des emplois moins qualifiés. Cette séparation entre le travail de conception et d'exécution permet toutefois de faire émerger des nouveaux emplois qualifiés, dont le nombre reste toutefois trop faible pour intégrer les travailleurs déqualifiés par l'automatisation. Paul Attewell critique cette hypothèse en s'appuyant notamment sur la tendance à la délocalisation des tâches à faible valeur ajoutée. Cette tendance aurait pour conséquences de maintenir les compétences dans le pays donneur d'ordres et d'éventuellement permettre au travailleur recruté pour réaliser les tâches à plus faible valeur ajoutée de gagner en compétences ("*redistribution of work*")¹⁵³.

Ainsi, ce qui ressort des études qui portent sur l'automatisation, c'est davantage une forme de polarisation des compétences, davantage qu'une disparition du travail matériel ou une déqualification uniforme. De nouvelles tâches d'interprétations des signaux et de compréhension des "remous" de la production sont alors nécessaires.

Cette controverse est marquée par deux visions des emplois qui perdurent à l'heure d'étudier l'automatisation algorithmique. D'un côté, une vision "substantialiste" dans laquelle le travail est analysé par essence dont on pourrait "*appréhender, caractériser et hiérarchiser les propriétés directement, en vue de leur conférer, ensuite, une valeur*" (p.73) Stroobants et Singly (2016). Dans cette approche, il s'agit alors d'analyser le "poste" de travail au regard de sa dimension technique, de sa position dans l'échelle de l'entreprise et de la responsabilité dans la production et de "la fréquence relative des qualités requises". De l'autre, Pierre Naville considère plutôt que les emplois et leurs descriptions sont "*déjà marquées par des valeurs relatives, indissociables de la qualification*" (p.73). Ces valeurs renvoient alors au "*jugement social qui s'exerce globalement et classe les travailleurs les uns par rapport aux autres*" (p.72).

152. "à mesure que le travail devient vide de sens, les travailleurs sont de plus en plus aliénés." Attewell (1987)

153. Pour appuyer cette proposition, l'auteur prend l'exemple de tâches de saisie de données informatiques dans le domaine des assurances. Il y observe la réorganisation du travail entre les opérateurs de saisie de données liées à des dossiers d'assurances et les "examineurs" de dossiers qui doivent valider la requête d'un assuré. Cette réorganisation consiste à transférer des tâches de codification de factures et d'identification d'informations, perçues comme à plus faible valeur ajoutée, des "examineurs" vers les opérateurs de saisie de données. Les premiers peuvent se recentrer sur les tâches à plus haute valeur ajoutée tandis que les seconds ont désormais un travail plus élargi incluant des tâches de saisie simple et des tâches d'organisation et d'identification d'informations dans le dossier des assurés.

La vague actuelle d'automatisation a fait l'objet de travaux analysant ses effets sur l'emploi, dans une perspective macro-économique, en s'appuyant sur la décomposition des emplois en tâches dites routinières automatisables. Par exemple, le travail de Frey et Osborne (2017), montre que jusqu'à 50% des emplois sont susceptibles d'être automatisés. En s'appuyant sur la décomposition des tâches qui composent un emploi, ils identifient également des tâches susceptibles de ralentir l'automatisation. Certaines activités nécessitent ainsi des compétences sociales de négociation, de créativité qui sont, selon eux, peu propices à l'automatisation. Largement critiqué, leur travail a été nuancé par d'autres études. On peut citer celle de Arntz, Gregory et Zierahn (2018) pour l'OCDE qui évoque plutôt 13% des emplois susceptibles d'être automatisés et 25% des emplois risquant d'être profondément modifiés¹⁵⁴.

Les controverses portant sur la mesure des effets de l'automatisation montrent la difficulté pour la littérature économique à apporter des preuves concrètes de ces derniers d'un point de vue macro-économique. La difficulté principale réside dans le choix de l'approche permettant de définir un métier comme automatisable. Comme dans l'analyse de Frey et Osborne (2017), de nombreux travaux se basent sur la distinction entre les tâches et les emplois. Ainsi, plus un emploi sera composé de tâches routinières, prévisibles, et se déroule dans un endroit stable, plus celui-ci sera théoriquement aisément remplaçable par une machine (Autor et Dorn, 2013; Le Ru, 2016).

L'hypothèse de la fin du travail a également été critiquée à partir du postulat d'une différence entre le niveau des tâches et celui des emplois (Autor & Salomons, 2018). Or, si certaines tâches au sein des emplois peu ou moyennement qualifiés sont "susceptibles d'être automatisées", *"de nombreux emplois moyennement qualifiés continueront d'exiger un mélange de tâches provenant de l'ensemble du spectre des compétences"*. Certes, certaines tâches peuvent être automatisées, mais de nombreux emplois sont composés de tâches interpersonnelles (c'est le cas des caissières et des caissiers). Une contribution récente tente d'analyser cette question du point de vue des effets potentiels de l'IA sur la qualité de l'emploi (Benanav, 2020). De son point de vue, aucun effet significatif de l'IA sur la productivité ne peut réellement être observé, de la même manière que l'automatisation n'a pas eu d'impact significatif sur la productivité dans le secteur industriel. Il s'agirait plutôt d'une crise de surproduction qui aurait pour effet de ralentir le dynamisme du marché du travail. La question n'est donc pas de savoir si le travail a un avenir, mais plutôt de savoir à quoi ressemblera l'avenir du travail, l'auteur estimant que nous assisterons plus probablement à l'émergence d'un précarat et d'un sous-emploi qu'à la disparition du travail.

Une dernière thèse évoque plutôt une *polarisation des compétences*. Elle est à la fois soutenue par des études

154. Toujours dans la perspective d'analyser le caractère routinier des tâches, Gregory, Salomons et Zierahn (2016) montrent que, même s'il peut y avoir une destruction d'emplois à fortes tâches routinières, il s'agit d'une période d'ajustement, qui se terminera par une tendance à la création de nouveaux emplois induits par la technologie, une fois la phase d'automatisation stabilisée.

macro-économiques (Goose et Manning, 2014) sur l'évolution de la part de l'emploi selon les qualifications, mais également par des études comme celle de Acemoglu et Restrepo (2019) qui montrent que l'automatisation a comme effet un déplacement du travail. Des tâches anciennes disparaissent au profit de nouvelles pour lesquelles le travail est plus compétitif que les machines. Or, ce déplacement du travail pourrait finalement favoriser les emplois les plus qualifiés d'un côté et ceux les moins qualifiés de l'autre. Il existe ainsi un consensus autour de la polarisation des qualifications identifiables au niveau macro-économique ¹⁵⁵.

Ces modèles interprétatifs paraissent cependant insuffisants pour décrire les évolutions du travail liées à l'IA. L'analyse par les tâches, en se basant sur le descriptif des tâches par emploi, se heurte à la réalité du travail. Dans la zone d'incertitude entre le travail prescrit et le travail réel, notre hypothèse est qu'il subsisterait des tâches intermédiaires qui ne sont pas forcément prises en compte dans les classifications d'emplois utilisées dans les calculs macro-économiques. Un deuxième argument concerne plutôt la redistribution géographique du travail qui faciliterait cette polarisation des tâches en permettant à des acteurs institutionnels de déqualifier certaines tâches en minimisant, voire même en invisibilisant, le rôle des travailleurs peu qualifiés (*misrecognition*). Dès lors, c'est la relation de travail, dont découle la catégorisation des travailleurs, qui serait un facteur déterminant de cette polarisation et non la mise en place de l'automatisation algorithmique ¹⁵⁶.

Le travail de Shestakofsky (2017) montre que l'application de technologies automatiques crée "de nouveaux problèmes" qui nécessitent d'être réglés par travailleurs humains. Contre l'hypothèse d'une fin inéluctable de certaines tâches qui va de pair avec celle de la déqualification du travail, il montre que des entreprises du secteur numérique s'appuient continuellement sur différents types de travail délocalisés dans des pays en voie de développement. À partir du cas français, nous montrons comment les travailleurs de Madagascar participent à la production de services automatisés sans nécessairement contribuer directement à la production des modèles d'IA. Cette dynamique de délégation de tâches à ces travailleurs vient alors pallier aux imperfections des machines automatiques. Ce travail humain est alors "palliatif".

6.1 Le travail de l'optimisation algorithmique réalisé depuis Madagascar

En France, la numérisation de nombreuses activités induite à la fois par la transformation de l'organisation de certains services, mais aussi par la mise à disposition de documents papier en version numérique conduit les organisations à se restructurer. Entreprises et services publics proposent dorénavant de nombreux services numériques qui nécessitent la structuration des services de relations clients. Dans le même temps, la

155. Charnoz et Orand (2016) vont dans ce sens en montrant par exemple qu'en France, la part des métiers routiniers a baissé et qu'il y a une redistribution des travailleurs vers des emplois de service peu qualifiés.

156. Un dernier argument que nous ne mobilisons pas dans notre analyse empirique limite également la portée de ces analyses et réside dans l'émergence éventuelle de nouvelles compétences qui ne seraient pas identifiables actuellement, mais qui seraient consubstantielles au déploiement de l'IA.

numérisation des documents papier oblige les entreprises et administrations à mettre en place une chaîne de production dédiée.

Dans les pays comme Madagascar, cette transformation numérique se traduit par l'émergence de secteurs entiers dédiés à l'externalisation de ces tâches. L'enquête réalisée auprès des travailleurs malgaches a ainsi permis d'identifier de nombreuses activités externalisées liées au numérique et à l'automatisation de services. L'analyse de ces activités permet de mieux comprendre la position des travailleurs dans la production de services automatisés ainsi que la continuité entre les projets de numérisation et d'automatisation plus anciens et les nouveaux projets utilisant le *machine learning*.

Pourtant, à l'exception de l'enquête de [Tine \(2021\)](#), peu de recherches francophones ont étudié l'externalisation de services dans des pays en voie de développement. Si cet objet de recherche a fait l'objet de nombreuses discussions entre le début des années 2000 et les années 2010, c'est principalement sur des centres internes à la France que portent les enquêtes. C'est sur la base du travail de [Shestakofsky \(2017\)](#) et de notre enquête de terrain que nous suggérons qu'il existe une continuité entre les pratiques qui supportent l'automatisation algorithmique ¹⁵⁷. Dans cette première sous-section nous abordons deux catégories d'activités qui complètent l'automatisation algorithmique. Nous abordons d'abord le travail de numérisation des textes. Ce phénomène qui émerge dès les années 1990 va s'appuyer à la fois sur du travail humain et sur l'utilisation de logiciel automatique de reconnaissance optique de caractères (OCR). En plus de la numérisation de textes, deux activités vont également se développer, autour des années 2010, en lien avec l'essor du e-commerce afin d'améliorer la découvrabilité des produits et des sites Internet : l'enrichissement des catalogues produits publiés en ligne et la rédaction web. Cette activité émerge alors en écho au développement des algorithmes qui ordonnent le web.

6.1.1 Le travail algorithmique direct pour la numérisation des livres et des documents administratifs

La numérisation des procédés de composition des documents est un des cas d'usages des technologies d'automatisation. En effet, si les chaînes de production évoluent avec l'introduction de l'ordinateur, cela ne simplifie pas nécessairement la division des tâches. Ancien chef de projet chargé de la "numérisation" à la Bibliothèque Nationale de France (BNF), Pierre évoque son parcours d'imprimeur en abordant les évolutions du métier au regard de l'émergence du numérique. Après un Brevet de Technicien Supérieur (BTS), il travaille dans plusieurs imprimeries dans lesquelles il constate au début des années 2000 la disparition des techniques argentiques au profit des technologies numériques. Ces évolutions se traduisent par une redis-

157. Des entreprises comme le sous-traitant C ont commencé par la numérisation, continué avec les activités de la voix et s'occupent désormais de tâches liées à l'automatisation. C'est également le cas dans le parcours des travailleurs des données interrogés. Par ailleurs, nous défendons l'hypothèse d'une continuité entre la numérisation et l'automatisation, et la présence toujours plus importante de services d'aide aux clients délocalisés dans des pays en voie de développement.

tribution du travail de numérisation de documents textes. Ainsi la mise en page des catalogues de produits des supermarchés est externalisée à l'étranger, d'abord en Chine puis à Madagascar, où des travailleurs effectuent le détournage des photographies ¹⁵⁸.

Il évoque également la numérisation de certains éléments moins nobles des titres de presse, en l'occurrence les petites annonces. Il raconte son expérience au journal *Le Parisien*, qui passe par un prestataire à Madagascar pour publier ces éléments sur son site web : *"j'ai travaillé à la régie publicitaire du Parisien, ils publiaient des petites annonces (...) sur leur site internet (...) [ils] les envoyaient à Madagascar, mais je me souviens qu'on envoyait un mail à cette entreprise pour sortir tout ça."* (Pierre - Chef de projet - Gallica.fr). Les travailleurs devaient saisir les petites annonces à partir du document imprimé pour que ces dernières soient publiées sur le site ¹⁵⁹. Cette reconfiguration de la production de la presse se traduit par trois tendances : la disparition des imprimeries et de leurs travailleurs, l'élargissement des tâches de certains membres des rédactions (maquettistes et journalistes) et l'externalisation de certaines activités liées à la numérisation (photographie, intégration web) à Madagascar. Selon lui, ce sont la crise de la presse, la massification de l'information en ligne et l'instabilité des modèles économiques liés au Web, qui poussent ces entreprises à améliorer leurs marges en externalisant certaines activités ¹⁶⁰.

Après cette expérience au journal *Le Parisien*, il rejoint la Bibliothèque Nationale de France (BnF) pour travailler sur le projet de *"numérisation de masse"* des fonds patrimoniaux. Depuis 2007, la France s'est en effet dotée d'une politique publique de numérisation des documents et livres anciens en réponse à Google qui crée en 2004 sa bibliothèque numérique à travers le service Gallica. L'objectif consiste à permettre à n'importe quel citoyen de consulter les fonds de livres anciens français et européens. Plus de 15 ans plus tard, et après des périodes de tâtonnements et des investissements importants, 58 millions de documents numérisés sont disponibles au niveau européen et 9 millions au niveau français. Cette première tentative pour construire une alternative française et européenne à une grande plateforme n'a pas été sans difficulté, notamment en matière de financement.

En 2007, l'institution publique commence à *"sous-traiter en masse"* des tâches de numérisation de livres anciens. Le *scan* des livres est réalisé en France et les tâches de saisie à Madagascar : saisie de la table des

158. *"l'opération de détournage, c'est on a une photo et on va la coller sur un fond, c'est une sorte de photo-montage, c'est l'opération qui est la moins complexe, et c'est des chinois qui faisaient ça en masse, notamment pour des catalogues, sur un catalogue produit en France on était à 100 euros la page et c'était 10 euros (à l'étranger). En fait, ils produisaient le détournage (des photographies). Les opérateurs français faisaient la mise en page. Et ça ils continuent de le faire à Madagascar."* (Pierre - Chef de projet - Gallica.fr)

159. *On produisait en numérique, les maquettes, eux ce qu'ils recevaient c'est le document type imprimé avec les petites annonces et saisi dans une base de données pour que ce soit sur le site web."*(Pierre - Chef de projet - Gallica.fr)

160. *"Dans les redacs ils ont fait des économies. Dans les rédactions, tout le monde accepte de faire plus de choses. Et dans certains cas, ça se passe à Madagascar, et à la fin ça fait des effectifs qui se réduisent parce que les sites web n'arrivaient pas à trouver de modèle. Tout ce qui se produisait en Web ne rapportait rien."* (Pierre - Chef de projet - Gallica.fr)

matières, indexation des pages et correction du moteur d'Optical Character Recognition (OCR)¹⁶¹. Le travail humain est important, car il permet d'obtenir un taux de qualité proche des 100% difficile à obtenir sur des livres anciens, car les logiciels d'OCR sont basés sur des dictionnaires modernes. La BnF distingue l'OCR brut, qui est le résultat du logiciel, de l'OCR "haute qualité", qui implique des travailleurs malgaches. En effet, la BnF avait *"des exigences de qualité importante donc c'était corrigé à Madagascar en double saisie, on regarde si les deux saisies sont les mêmes. On arrivait à 99,9% de qualité ce qui était important parce qu'après ça allait dans Gallica."* (Pierre - Chef de projet - Gallica.fr).

Concrètement, le logiciel utilisé par la BnF, estime un score de confiance entre 0 et 1 pour chaque mot d'un livre : *"Disons que l'OCR met une note de détection, et tous les mots qui n'ont pas 1 on va demander de les détecter. On demande (aux travailleurs) de vérifier et de corriger."* Au-delà de l'orthographe du français ancien, les techniques d'imprimerie au plomb, la qualité de conservation des livres et du scan sont des éléments qui affectent la qualité de détection de l'OCR. Les mots sont parfois trop anciens par rapport aux dictionnaires utilisés pour les modèles d'OCR. Dans d'autres cas, les mots sont aussi effacés du fait de la conservation des livres. Les travailleurs sont là pour combler l'écart de qualité entre le logiciel pur et un taux proche de 100%.

Encadré n°38 : Le travail de numérisation des livres anciens

"ce qu'on va leur demander c'est d'identifier les pages et vérifier qu'il y ait pas de doublons. On lui demande de vérifier les doublons et il va typer les pages, il va dire ça c'est le sommaire, ça c'est la page de titre. On va lui demande de folioter, donc de mettre le numéro de page. On va aussi lui demander de recadrer la page(...) ça c'est pour la partie préparation du document pour Internet. La deuxième opération, c'est (...)la correction d'OCR et là il va aller voir dire là y a une photo et il va corriger l'OCR, et dans certains cas la saisie des tables des matières avec le lien vers la bonne page."

Pierre - Chef de projet - Gallica.fr

Le plan de numérisation de masse trouve finalement ses limites sur le plan du financement, ce qui conduit les institutions publiques à d'une part réduire la voilure et de l'autre à chercher d'autres manières d'opérer cette numérisation très coûteuse. Désormais, la BnF a choisi de se focaliser sur une solution entièrement automatisée pour des raisons de coûts : *"on passe plus par Madagascar pour la correction (...) on a juste l'OCR brut."* Cette évolution se fait néanmoins au détriment de la qualité. Le site de Gallica explique très

161. *"Tout ce qu'on ne peut pas faire autrement qu'en France [étaient faits à Madagascar]. Donc la numérisation des livres était faite sur les sites de SAFIG et à la BnF à côté de Chateauroux, à La Châtre. Donc il y avait un centre pour la BnF et les plans pour EDF, dont les plans des centrales nucléaires. Et puis, une fois que la numérisation était faite (c'est-à-dire) faire une photo de la page avec des scanners performants (...) tout était délocalisé à Madagascar : le détournement des images, la numérotation des pages. (Pour les mots), on numérise l'image et le logiciel d'OCR il va dire c'est un « C » et il va dire c'est « Carrefour."* (Pierre - Chef de projet - Gallica.fr)

clairement cette distinction entre d'un côté le travail du logiciel ("sans intervention humaine") et de l'autre l'OCR réalisée avec une "montée en qualité du texte" qui implique des travailleurs humains : "Au fil des programmes de numérisation de la BnF, deux types d'OCR ont été produits : OCR brut : la conversion se fait sans aucune intervention humaine [et] OCR avec montée en qualité du texte (...) par une correction manuelle afin d'atteindre un taux qualité cible (généralement 96%, 98% ou 99,9%)" ([Site web de Gallica](#)).

Lors d'un entretien, Haino qui travaille chez le sous-traitant F à Madagascar, me raconte sa première expérience comme opérateur de saisie en 2006. S'il évoque un travail "facile", Matthieu souligne tout de même qu'il est nécessaire de pouvoir comprendre et interpréter des mots qui sont parfois en vieux français. Ces sont présentes depuis la fin des années 1990 à Madagascar, où l'entreprise C opère une activité de numérisation de livres pour de nombreux clients. À l'époque, les livres sont acheminés par avion puis numérisés dans les locaux de l'entreprise. Toutefois, contrairement au cas de l'OCR, il s'agit alors de taper par ordinateur des documents écrits ou tapuscrits qui sont ensuite renvoyés par disquettes ¹⁶².

Naro est informaticienne à l'Agence Universitaire Francophone (AUF) et travaille dans le secteur depuis une vingtaine d'années. Lors de notre entretien, elle m'explique l'émergence du secteur de l'externalisation à Madagascar "a coïncidé avec tout ce qui est, les transformations numériques dans les bibliothèques, dans les grandes entreprises, quand ils ont commencé à numériser toutes leurs archives.". Pendant dix ans, cette numérisation s'est appuyée sur l'assemblage entre les logiciels d'OCR et le travail humain réalisé à Madagascar, qui eux ont fini par disparaître de la chaîne. La responsable de production d'une grande entreprise analyse cette évolution en évoquant l'amélioration des moteurs d'OCR : "sur la numérisation patrimoniale, les moteurs se sont améliorés, la boîte quand je suis arrivée, y avait 750 personnes et quand je suis parti y en avait 350 en 10 ans, donc forcément y a des hauts y a des bas. Là où avant, c'était de la saisie, petite main, trois personnes, deux qui font la saisie et une qui compare la saisie, là une seule est nécessaire." (Caroline - Responsable - Sous-traitant C). Ce qu'elle attribue aux progrès des logiciels d'OCR et de l'automatisation correspond en fait, au moins en partie, à une baisse de l'exigence en matière de qualité, consécutive à la baisse des fonds dédiés à la numérisation patrimoniale : "les marchés publics ont quand même baissé, y a quinze ans, y avait des gros appels d'offres sur des marchés publics, plus ça va moins y a de choses à numériser, et les budgets sont plutôt en baisse dans les marchés de ce type" (Pierre - Chef de projet BnF). Pour pallier à cela, l'institution publique lance plusieurs expérimentations pour remplacer le travail sous-traité à Madagascar. Tout d'abord, la BnF a repris de Google l'idée de faire un "captcha à la française". Les utilisateurs des services de la BnF seraient susceptibles de fournir du travail humain gratuitement pour compléter la numérisation, en saisissant les

162. "il y a 26 ans, c'était de la saisie (...) au kilomètre, d'ailleurs, ça s'appelait saisie.com. Les choses à saisir venaient par les stewards d'Air France, les livres ou les choses comme ça, il y a encore des gens – Andrea, une directrice de la data qui a encore le pupitre de l'époque, elle l'a gardé, en bois, qui avait été fait pour le poser (le livre) c'était de la saisie de livres pour commencer la dématérialisation d'informations." (Louis - Vice-président - Sous-traitant C)

mots non détectés par les logiciels d'OCR¹⁶³. Une deuxième manière de "sous-traiter" ce travail a été expérimentée sur des titres de presse anciens. L'idée de ce projet est alors de faire contribuer les internautes sur le modèle du "crowdsourcing", où, via un site Internet dédié, ces derniers peuvent proposer des corrections. Cette idée est rapidement abandonnée à cause des résistances internes et externes face à ce "travail gratuit" demandé aux utilisateurs de Galica. L'institution publique a toutefois conservé ce principe pour un projet de géolocalisation des documents, avec une interface qui permet à n'importe qui d'identifier la géolocalisation des dessins, images et autre.

Encadré n°39 : Le travail de numérisation des livres anciens

"En 2016, y avait le modèle Wikipédia(...), la BNF s'est engouffré là-dedans, a fait des projets de correction d'OCR dans le cadre de projets avec Orange (...). On a fait pas mal de recherches et puis quand même à la BNF, on est attaché au droit d'auteur, et des gens nous ont dit, « wikipédia c'est très bien, mais c'est un peu du travail gratuit », et on s'est demandé : est-ce qu'il faut faire travailler les gens gratuitement ? Et d'ailleurs (...) on s'est fait attraper par la CGT des correcteurs de presse qui disait « vous nous volez notre travail », ce qu'on constate c'est quelques utilisateurs (experts) qui se mettent dedans, donc c'est pas forcément des micro-tâches."

Pierre - Chef de projet - Gallica.fr

Après le texte viennent les images, pour lesquelles la BnF lance un vaste projet d'étiquetage dont l'objectif est d'attribuer des mots-clés à chacune des 200 millions d'images présentes dans le fonds patrimonial de l'institution publique¹⁶⁴.

De l'introduction de l'ordinateur racontée par (Maruani & Nicole-Drancourt, 1989) dans le secteur de l'imprimerie aux projets de numérisation et d'automatisation, les technologies changent, mais pas nécessairement la vision de ces activités, perçues comme résiduelles par rapport au travail de la machine, ce qui s'accompagne d'une faible valorisation sociale des qualifications. Pourtant derrière l'illusion du "travail facile", il existe un réel enjeu de maîtrise de ces tâches. Dans le cas des travailleurs de Madagascar, leur travail est supposé "facile", car ils ne doivent que corriger le logiciel d'OCR. Pourtant, c'est précisément dans cet

163. "On regardait ça de manière attentive, on a fait une expérimentation avec les archives de France en faisant un captcha à la française pour se débarrasser de Google (...) ça a pris je sais pas, c'est pas complètement terminé, normalement il devait y avoir FranceConnect, mais ils sont pas venus, mais c'est pas définitif." (Pierre - Chef de projet - Gallica.fr)

164. "On estime entre 100 et 200 millions d'images dans le fonds et on s'est dit qu'on allait lancer un fonds de développement qui fera appel à l'IA pour aller étiqueter ces documents. Y a des moteurs qui savent ça c'est une citrouille, y a des moteurs qui savent le faire, mais sur des fonds contemporains, mais finalement a ressurgi cette histoire de Madagascar (...) Sur des images contemporaines, avec Google Lens c'est quand même pas mal, on peut récupérer tout un tas de choses. Nous, on a bon espoir d'avoir des bons résultats, en se demandant comment c'est fait. En France, y a pas d'entreprises, on a pas eu d'autres choix. Le marché n'est pas lancé. Mais les répondants ça peut être des gens qui font de la sous-traitance à Madagascar, mais ça peut être des gros groupes comme Google ou Amazon." (Pierre - Chef de projet - Gallica.fr)

interstice que se situent leurs qualifications réelles. Il s'agit en effet de corriger des mots en ancien français, orthographiés de manière non conventionnelle et publiés dans les livres qui datent d'avant la normalisation du français au 19ème siècle. Lorsque je soulève ce paradoxe à Pierre, il me répond après un silence que ce n'est pas un problème car *"comme vous l'avez constaté en allant là-bas, nous on est dans des entreprises avec des niveaux d'études assez élevés donc corriger du français ça leur faisait pas peur."*

Au-delà de ces qualifications linguistiques, les travailleurs tombent dans l'exigence de perfection, corollaire d'un travail dit non-qualifié comme l'analysent (Maruani & Nicole-Drancourt, 1989). Pierre insiste d'ailleurs sur la nécessité de vérifier le travail réalisé à Madagascar et sur les difficultés de la BnF à imposer des standards de qualité respectés par leurs prestataires : *"In fine, c'est quand même eux qui produisaient l'indicateur final de qualité et on a essayé de mettre des contrôles croisés, c'est pour ça qu'il y avait les audits pour qu'ils nous montrent comment ils se contrôlaient. Parce que la BNF avait pas les moyens [et] le temps de contrôler grand chose et donc ce taux de qualité était à la main du prestataire. (...) on pouvait pas être certains [de] que ce qu'ils nous livraient"* (Pierre - Chef de projet numérisation de masse BnF) C'est d'ailleurs le fait d'atteindre ce taux qui nécessitait de mettre en place un processus de production où deux saisisseurs vérifient et notent les mots manquants de l'OCR avant qu'un autre travailleur ne contrôle ces deux saisies. Du côté de la BnF, un système d'échantillonnage des livres est mis en place pour vérifier si le niveau de qualité attendu est suffisamment bon.

Derrière ce travail facile, c'est donc le fait d'obtenir un taux de qualité proche de 100% qui rend la tâche plus complexe. Du point de vue des travailleurs d'abord, car ces derniers ne doivent pas tromper sous peine de subir des conséquences de la part de leurs managers (baisse des primes, licenciements...) ¹⁶⁵. Du côté de la chaîne de production, cela se traduit par une distribution du travail entre des travailleurs plus ou moins compétents. Cet exemple confirme l'analyse de (Maruani & Nicole-Drancourt, 1989) à propos du travail de secrétariat transformé par l'apparition de la bureautique : *"La qualification devrait renvoyer à cette maîtrise que possède - ou non - l'employée. Or, elle est niée au nom de la simplification (...) du travail de l'employée. Ces observations ne tiennent pas compte des surcharges imposées au nom du "travail facile" : c'est tellement simple de pianoter que l'on peut multiplier les brouillons de manuscrits et exiger une perfection toujours plus grande"* (p.61). A l'envers de l'image d'un travail facile, les compétences réellement nécessaires pour réaliser le travail, l'exigence de la perfection et la pression mise sur les travailleurs complexifient l'organisation du travail. Maruani et Nicole-Drancourt (1989) observent également les conflits entre les hommes-typographes et les femmes-clavistes et constatent que pour résoudre ces conflits, est installé un rideau qui sépare les deux groupes : *"Les correcteurs sont venus, au fil des années, travailler au même étage qu'elles. Ils se côtoient donc. Ils se côtoient, mais ne se voient pas. Un rideau les sépare. Il y aurait beaucoup à dire sur ce rideau, sur ce paravent*

165. Nous le verrons dans le chapitre 4 consacré aux conditions de travail, mais les erreurs ont des conséquences sur les primes de ces derniers.

qui masque d'indécentes différences." (p.43) Dans les exemples évoqués par Pierre, pas besoin de rideau. En effet, 9000 kilomètres séparent ces nouveaux "OS du tertiaire" des autres catégories d'employés. Il est alors plus aisé pour les donneurs d'ordres, de minimiser la valeur sociale de leur travail. Ces derniers ne sont mis au courant de ce que font les travailleurs qu'à travers les dirigeants d'entreprises de sous-traitance qui les emploient pour des salaires très faibles.

De nombreux travailleurs interrogés expliquent avoir réalisé des tâches de numérisation de documents administratifs, et ce, jusqu'à aujourd'hui. C'est dans un premier temps la mise en ligne des anciens actes d'état civil par la France qui va faire exploser ce secteur à Madagascar. Comme pour la BnF, les pouvoirs publics souhaitent procéder à une numérisation de masse pour mettre à disposition ces documents aux usagers des services publics. En parallèle de cette mise à disposition, des entreprises de généalogie commencent à se positionner sur le secteur de la numérisation des actes de décès et de naissances (Banat-Berger, 2010). Louis, dirigeant d'une entreprise présente à Madagascar et en Roumanie raconte cette époque : *"la saisie de données à l'époque, c'était de la saisie d'états civils pour les mairies françaises."* (Louis - CEO - Sous-traitant F). Ainsi, lorsqu'un citoyen fait une demande d'accès à un acte civil, les mairies font appel à un prestataire de numérisation qui scanne le document. Les travailleurs doivent ensuite saisir des données comme la date de naissance, ou les noms et prénoms des personnes. Cette phase a perduré jusqu'à aujourd'hui, pour la constitution des bases de données de généalogie ¹⁶⁶

Dans les années 2000, émerge également ce que Florence Maraninchi appelle la "numérasse", c'est-à-dire le mélange entre les documents papiers et le fait de passer par des applications dédiées pour certains actes administratifs. Un cas de figure bien connu des chercheurs est la nécessité de rentrer sur des applications les montants dépensés lors d'un voyage académique, ainsi que les factures correspondantes. De manière générale, le papier existe toujours, mais il est nécessaire de le numériser pour le faire rentrer dans des bases de données. Dans le cas précédemment cité, ce sera à au chercheur-usager de le faire.

Dans certains cas, cette tâche a également pu être externalisée. C'est justement ce que décrit Miky lorsqu'elle me raconte les tâches qui étaient réalisées chez le sous-traitant C aux alentours des années 2010 : *"il y a eu un moment en France, où on a dit faut plus faire de papier, faut digitaliser. On avait des dossiers scannés. Nos équipes devaient intégrer des informations dans des champs prédéterminés. Et ils saisissaient."* (Miky - Ancienne RH - Sous-traitant C) Dans la presse, des bons ou formulaires d'abonnement au format papier continuent d'exister : *"aujourd'hui on a pratiquement tous les acteurs de la presse (...) pour lesquels on saisit directement dans leur CRM (Customer Relationship Management) qui va générer l'installation de l'abonnement, l'envoi de l'abonnement, la facturation."* (Louis - Vice-président - Sous-traitant C) Si certains formulaires se numérisent, des documents officiels comme la carte d'identité et les RIB sont toujours parfois fournis

166. "on pouvait travailler pour (...) le concurrent de Filae. Filae a copié une boîte américaine qui s'appelait Family.org, qui recensait les fiches de naissance, de décès." (Louis - CEO - Sous-traitant F)

par les usagers et consommateurs dans un format qui nécessite une saisie. Christiane est une travailleuse du Cameroun qui travaille par le biais d'une plateforme "d'impact sourcing" (AIEthic) et qui y réalise des tâches variées, telles que à consulter des documents officiels, comme une carte d'identité, et à remplir les informations sur le site de son client, une agence d'intérim française¹⁶⁷.

Le cas des factures de supermarchés ou de restaurants illustre tout autant ce concept de "numérasse" et ce que cela nécessite en matière de travail externalisé. Le directeur du sous-traitant F explique cette intrication. L'entreprise fait partie des prestataires privilégiés de la plateforme Appen pour laquelle ils saisissent des informations présentes sur des factures (montant total, pourcentage de TVA...). Sur AIEthic, plusieurs travailleuses racontent également avoir réalisé ce type de tâches pour une grande entreprise française d'édition de logiciel à destination des entreprises. Cette dernière travaille principalement avec des cabinets d'experts-comptables, qui manipulent de nombreuses preuves de paiement, et propose un service de détection automatique des éléments présents sur ces documents. Comme l'indique leur site Internet : *"Les documents, pour lesquels nos machines ont détecté des incohérences, peuvent nécessiter plus de temps de traitement. Quoi qu'il en soit, ils seront tous traités dans un délai de 24 heures."* Cette durée de traitement peut étonner pour un service automatique. Là encore, des travailleuses de AIEthic continent africain sont chargées de vérifier la labellisation des éléments et de ressaisir ces montants dans une interface.

Les exemples de projets liés à la numérisation de documents textuels offrent d'autres perspectives sur le travail de l'IA. Même s'il s'agit de documents très bien structurés, le travail humain délocalisé reste nécessaire. Pour nous, cela va dans le sens de l'existence d'un continuum entre les tâches humaines et les tâches automatisées, où, finalement la véritable fonction du numérique au sein des organisations réside dans la capacité à coordonner ces deux faces de la production. Dans cette perspective, l'ensemble des projets d'automatisation algorithmique se situe entre deux pôles : un pôle de projets "très automatisés" et un pôle de "projets très humains". Cette perspective est abordée dans la littérature à travers le livre de [H. R. Ekbia et Nardi \(2017\)](#). Pour nous, l'IA renforce les tendances à l'externalisation permises par le numérique. Ainsi, l'application de l'automatisation algorithmique ne résiste pas à l'observation des tâches résiduelles qui sont également externalisées dans des pays en voie de développement, en plus d'être parfois dévolues à des consommateurs. De manière schématique, on fait alors face à deux modèles d'organisation du travail d'automatisation : d'un côté Google avec le travail des consommateurs (Captchas), et de l'autre la puissance publique française avec le recours aux travailleurs d'une de ses anciennes colonies.

167. "(...) on travaille sur les plateformes en back-office [...] On vous donne des documents d'identités, et ça reste strictement confidentiel (...) que vous devez les retranscrire sur (...) la plateforme du client, par exemple les noms, toutes les informations présentes sur le document (...) on retranscrit, on vous donne une image d'une carte d'identité et vous retranscrivez les informations en back-office sur la plateforme du client. C'est, comment ça s'appelle, gojob (...) Oui, c'est des pièces d'identité." (Christel - AIEthic)

6.1.2 Le travail algorithmique indirect des travailleurs malgaches au service des machines qui "ordonnent" le web

Plus récemment à l'échelle de l'histoire du Web, de nombreuses entreprises se sont mises à à faire de la "Search Engine Optimization" (SEO) pour s'assurer du bon référencement de leurs sites sur les moteurs de recherche. Ils font en général appel à des agences de marketing numérique, dont l'objectif est de mettre en place des "dispositifs de captation" des internautes (Mellet, 2011) comme l'écriture de blogs pour vanter les mérites d'un produit. À l'origine, l'algorithme "PageRank" derrière le moteur de recherche de Google devait trier les sites Web selon leur "autorité". Plus des sites web pointent vers un site, plus ce dernier sera classé tout en haut de la liste des résultats Google. Ainsi, se "*rendre visible, c'est obtenir des autres du « jus de lien »*". *Si celui-ci ne vient pas naturellement, il faut alors l'extorquer, l'acheter ou le produire artificiellement.*" (Cardon, 2013)

Pour faire venir des liens vers leurs sites web, les webmestres inscrivent leurs sites dans des annuaires, des blogs réputés et même sur Wikipédia. En réaction, Google a permis "*aux sites de démonétiser certains des liens qui partent de chez eux*". Depuis, l'industrie du SEO a tenté de créer artificiellement de l'autorité autrement à l'aide de "fermes à liens". Il s'agit par exemple de créer tout un écosystème de sites-écrans, autour d'un site marchand ("*money site*") : on pourra par exemple créer des blogs traitant du même sujet et composés de nombreux liens externes vers le site principal. Par la suite, cette industrie va se diriger vers la production de contenus destinés à renvoyer l'internaute vers un site marchand.

Dans la même veine, plusieurs sites de e-commerce ont recours à des plateformes ou à des entreprises de sous-traitance pour des tâches de nettoyage de leur catalogue et même désormais pour l'intégration de produits sur des catalogues et l'enrichissement des données des produits. C'est ce que raconte, le chef des opérations de l'entreprise E qui compte 2000 salariés à Madagascar. En plus de chercher les doublons sur le catalogue de Amazon, leurs équipes vont également réaliser des tâches de recherche d'informations sur de potentiels produits à intégrer sur le catalogue du site de e-commerce ¹⁶⁸.

Le développement de ces activités est permis par l'émergence du modèle des places de marchés. Sur le modèle de sites comme Amazon ou Ebay, de nombreux sites intègrent les boutiques de vendeurs tiers, en plus de proposer une boutique principale, sur laquelle ils vendent leurs propres produits. Que ce soit pour la visibilité des annonces sur Google ou sur les sites eux-mêmes, il est alors nécessaire de standardiser la présentation des produits et de faire coïncider la classification des tiers vendeurs avec celle utilisée par la

168. "*Rapidement, Amazon US nous recommande à Amazon Inde qui eux travaillent principalement sur l'entretien de la marketplace Amazon : les produits disponibles, les marques disponibles, le catalogue de produits, enlever les doublons, essayer d'affiner les recherches, tout ce qui tourne autour du catalogue de produits d'Amazon, (...) et ça évolue très rapidement vers de la recherche de contacts, d'abord du scraping de sites compétiteurs d'Amazon et (...) de la recherche d'informations, de l'enrichissement de données [pour compléter le catalogue]*" (Loik - COO - Sous-traitant E)

marketplace¹⁶⁹. La difficulté de cette tâche tient à la variété des produits à intégrer sur le site de e-commerce. Entre deux catalogues, les informations sur les produits diffèrent, mais aussi la manière de classer le produit. L'intégration des images des produits peut également s'avérer problématique si celles envoyées par les tiers vendeurs ne correspondent pas au format du site de e-commerce.

Encadré n°40 : Appréhender les catégorisations de données

”L'exemple horrible absolu (c'est BricoBoutik), le pire du pire. Le back-office est mal fait donc la majorité des vendeurs n'y arrivent pas. Le problème c'est que nous on prenne les catalogues de tous les tiers vendeurs et on va dire « cette référence-là ou cette catégorisation-là sur ce catalogue c'est cette catégorisation chez vous, etc. ». Vous imaginez sur des centaines de catalogues ? Avec des gens qui vous envoient des images qui ne sont pas au format et qui ne sont pas renommées, etc. ”

Louis - CEO - Sous-traitant F

Lors, d'un entretien, Ando, aujourd'hui annotateur chez le sous-traitant B raconte également une tâche de catégorisation de produits qu'il a réalisée lorsqu'il était chez le sous-traitant C. Le projet consiste à catégoriser les produits du catalogue d'une entreprise de e-commerce française¹⁷⁰. Ensuite, les travailleurs doivent écrire des descriptions de produits en se basant sur des présentations déjà existantes et ce, sans que le contenu ne soit détecté comme du "contenu dupliqué" sous peine d'être pénalisé sur le plan du référencement¹⁷¹.

Comme l'indiquait Louis, une partie de la difficulté réside dans la variété des produits vendus sur les *marketplaces*. Dans le cas du site évoqué par Ando, il raconte avoir eu beaucoup de problèmes pour réaliser l'annotation des pneus pour un site de e-commerce comportant de nombreuses références. Contrairement à d'autres projets, il n'existe pas de guide qui explique les différents produits et les catégories de produits. Le projet nécessite donc plusieurs niveaux de tâches : une part d'interprétation et de recherche d'informations (premier niveau), des échanges avec ses collègues pour résoudre le problème (deuxième niveau),

169. "sur tout ce qui est validation de base de données de produits. On travaille pour un concurrent d'Amazon. Sauf que la base de données en taxonomie elle est pas identique. Ils ont dit aux tiers vendeurs « Vous voulez tous être chez nous. On n'a même pas besoin de faire de pub alors c'est vous qui allez uploader votre catalogue. ». Donc ils le font, mais généralement mal. Et c'est là où on intervient parce que l'on connaît par cœur la taxonomie en question et donc on peut définir si c'est bon ou pas. On apprend un process et même s'il y a 50 catégories différentes, une fois que le savoir est fait, il est fait." (Louis - CEO - Sous-traitant F)

170. "A : vous voyez, quand vous cliquez sur un produit sur Cdiscount, et bien nous on s'occupe de les catégoriser d'abord, donc où ça va sur le site, et après on s'occupait des descriptifs. C'est pour ça que c'est un peu volé, on copie les descriptifs d'Amazon, on prend ... en fait c'était vraiment le process hein. On savait pas, y'avait une armoire, on gère une armoire du même genre, on prend les longueur/largeur/hauteur/profondeur, donc c'était ça ..." (Ando - Annotateur - Sous-traitant B)

171. En effet, même si Google s'en défend dans son centre d'aide, une croyance toujours tenace dans le monde du SEO indique qu'en cas de présence de contenus dupliqués entre deux sites, par exemple celui d'un site de e-commerce et celui de son fournisseur, le moteur de recherche est susceptible de pénaliser le site.

une connaissance du sujet par le manager (troisième niveau) et éventuellement l'implication du client (quatrième niveau) ¹⁷².

Ces projets requièrent parfois des spécialisations comme dans le cas de la saisie de produits pour une place de marché dédiée à des fournisseurs de matériels agricoles et industriels. Chez le sous-traitant B, plus d'une trentaine d'agents s'occupent d'intégrer des produits sur ce site, en tenant compte des spécifications des fournisseurs : *"Les tâches c'est plutôt d'essayer de reproduire la représentation sur le site du client sur l'interface d'exposition. On a le site de John Deer, on a plusieurs catégories de machines, on va (...) les reproduire sur le site d'exposition virtuelle avec le cahier des charges du site d'exposition, donc on va collecter les liens du produits, les specs, les brochures, et les transposer sur les sites d'exposition."* (Nahary - Chef de projet - Sous-traitant B).

Les difficultés du projet résident principalement dans le niveau de compréhension attendu dans des domaines très techniques ce qui passe par une bonne formation initiale ou bien par l'acquisition de qualifications spécifiques ¹⁷³. Nahary explique que les agents peuvent devenir "autonomes", mais que cela nécessite minimum trois ans de travail sur le projet : *"il faut compter au moins une année pour la montée en compétences, d'ailleurs quand on recrute on dit déjà à l'agent que si vous intégrez le projet c'est pour au moins trois ans, une année de montée en compétences et deux années où il va améliorer ses performances, parce que si mon projet y a des niveaux, y a le simple agent, le niveau intermédiaire et le niveau avancé tout en restant agent."* (Nahary - Chef de projet - Sous-traitant B)

Outre le travail lié aux catalogues de e-commerce, certains travailleurs racontent des projets relatifs à l'optimisation de la découvrabilité des sites web. Pour cela, il faut être détecté par le "robot.txt", l'outil de Google qui parcourt le Web pour indexer les sites web ¹⁷⁴. La rédaction de contenus pour l'optimisation du référencement fait également partie de l'éventail des services proposés par le sous-traitant C. Ainsi, en plus de la

172. *"A : vu que Cdiscount c'est très vaste, ça s'est passé tant de fois ... par exemple on était sur les pneus, je ne connaissais pas du tout. Il y avait plusieurs mesures, donc je sais pas, il y avait quelque chose par quelque chose. Et dans le fichier donné on m'avait demandé ... alors là je me souviens vraiment plus, il y avait « rayon », moi je me suis dit « mais qu'est-ce que c'est que ça ? » (...) comme d'habitude, il y avait mon manager à l'époque, donc je lui ai demandé. Et après, si lui il arrive pas à répondre, bah il demandait aux clients (...) sur cette histoire de pneus, je pense qu'il m'a juste montré le truc, ouais ... vu que j'ai déjà googlé tout et tout ... mais à l'époque non ça ne donnait rien, je n'avais pas compris encore (...) il n'y avait pas de consigne pour les produits, vu qu'il y a énormément de produits, même actuellement, sur Cdiscount"* (Ando - Annotateur - Sous-traitant B)

173. *"Parmi les domaines de compétences de virtual expo, l'industriel, l'agriculture, l'archi, le profil et la compétence des gens va être sur ces trois domaines là, on recrute au moins quelqu'un qui a eu un (Bac+2) dans ces trois domaines. Quand c'est un projet en architecture, donc on a pas mal d'ingénieur en travaux publics, ça a été décidé à notre niveau, mais auparavant, le client a juste fixé deux années d'études, mais nous lorsqu'on a traité les demandes des clients, sur le plan pratique c'est plus intéressant d'avoir des études au-delà de deux années universitaires, le raisonnement logique et l'intelligence, ça diffère entre ceux qui ont fait deux années et ceux qui ont des diplômes d'ingénieurs"* (Nahary - Chef de projet - Sous-traitant B)

174. Ainsi, comme le dit Cardon (2013), on *"peut retracer toute l'histoire de l'évolution de l'algorithme comme un jeu de chat et de souris entre les webmasters et la firme de Mountain View pour, d'une part, essayer d'agir stratégiquement sur le PageRank et, d'autre part, détecter et punir ces comportements en réformant l'algorithme avant que la détection d'une autre faille n'ouvre une nouvelle porte aux comportements stratégiques des internautes."*

saisie de données, certains employés réalisent des tâches rédactionnelles. Contrairement à la saisie pure, on y retrouve des travailleurs plus compétents, notamment en ce qui concerne la maîtrise du français. Ce type de projet a émergé autour des années 2010, avec le lancement de projets à destination de Ucatchit, une entreprise spécialisée dans la stratégie numérique. Se positionnant en tant qu'agence, cette entreprise a donc de nombreux clients ce qui a fait énormément varier les contenus réalisés dans cette entreprise. Nos entretiens permettent d'identifier les difficultés liés aux projets SEO : la partie rédactionnelle, la nécessité de maîtriser plusieurs thématiques et l'utilisation des mots-clés pertinents pour améliorer le référencement ¹⁷⁵.

Certains projets de rédaction sont encore plus complexes. M.N.A, interrogé au début de notre travail de terrain en 2020 nous explique une de ses expériences dans le secteur de l'externalisation. En tant qu'étudiant, il est employé pendant l'été pour rédiger des contenus journalistiques : *"J'ai fait de l'outsourcing pour la rédaction d'articles pour des contenus en ligne et pour des magazines comme Voici et Valeurs Actuelles, j'ai par exemple fait un article sur la Bourse, alors qu'il n'y a même pas de bourse à Madagascar. J'ai fait un article sur les municipales 2014."* (M.N.A - Ambassade de France à Madagascar). Comme l'explique Marion, consultante en marketing digital à qui je demande des explications sur ces projets : *"c'est un peu le géant Google ou Facebook qui réclame toujours plus à bouffer et tout le monde qui doit le nourrir, les entreprises sont obligées de faire ça à cause de ce fonctionnement"* (Marion - Freelance en marketing digital).

La numérisation est donc déjà le support de cette automatisation algorithmique, tout autant que les travailleurs de Madagascar, et vient en renforcer la portée. L'encadré qui suit illustre ce phénomène à travers l'exemple d'une entreprise de veille médiatique. L'objectif du projet est de permettre à des entreprises de suivre leurs retombées dans la presse. Le travail se traduit par de la saisie de mot-clés sur des interfaces dédiées. Ainsi dès qu'un travailleur repère une marque, il doit l'encadrer pour que l'information soit remontée à l'entreprise détentrice de cette dernière.

Encadré n°41 : L'annotation sur le projet VeilleMédia 1

"ZI : oui oui, VeilleMédia (...) c'était on va dire vers 2011. On a commencé petit avec la partie audio (...), comme je t'avais dit tout à l'heure,(...) par exemple (il y a) une émission de France Bleu Rhône-Alpes (...) et on saisit ... on ne fait pas de la saisie kilométrique, mais une synthèse de ce qui a été dit (...). Et il y a surtout les mots clés qui doivent être ressaisis. Quand c'est saisi dans une interface comme ça en fait, ça passe dans un autre outil qui prend les mots clés qui ont été saisis chez nous, par exemple Renault. (Là) où il y a indiqué « Renault », (cela) passe par des rédacteurs qui font la

175. "(...) on a commencé avec la plateforme (...) Ucatchit (...) ou quelque chose comme ça (...) l'équipe devait rédiger des rédactions de qui pourraient aller de 100 mots à 1000 mots, donc qui varient en termes de thématiques, de la vente High Tech, tout ce qui est Pierre et Vacances, tout ce qui est saut en parachute. Donc c'était vraiment très très diversifié, l'équipe devait insérer des mots clés pour vraiment booster les moteurs de recherche, varier toute la partie rédactionnelle en utilisant les mêmes mots-clés, donc c'était un challenge" (Miky - RH - Sous-traitant C)

synthèse de ce qui a été dit durant le journal « machin Carlos Ghosn qui dit que ... » (...) Il y a même la partie vidéo, donc on écoute et on regarde en même temps (...) Donc si il y a un logo Renault qui passe, il faut le noter dans le truc de saisie (...) Sur la partie presse (...) le monde, Le Figaro, la presse nationale, ils font des scans vraiment tôt le matin (...), et ça tombe directement dans une interface que l'on traite ici. (...) Ce sont les lecteurs presse (...) qui détectent le mot clé et qui valident (les mots) (...) et ce sont ces parties-là qui sont envoyées chez VeilleMédia (...).

Louis - Vice-président - Sous-traitant C

Le projet va connaître une évolution récente, expliquée par le vice-président du sous-traitant C, lors d'une visite de l'atelier de production en bordure de la capitale malgache. L'objectif n'est plus de saisir et d'encadrer des mots-clés pour faire remonter la présence médiatique, mais de le faire pour automatiser la détection de ces éléments dans des textes, des vidéos et des extraits sonores.

Encadré n°42 : L'annotation sur le projet VeilleMédia 2

Vers la fin des années 2018, ils ont investi beaucoup dans la recherche et le développement. Donc ils ont fait beaucoup de recherches, les audios-là, les retranscriptions automatiques dans l'interface ... il suffit que le rédacteur il découpe une saisie (...) ils cherchaient à voir (automatiquement) les pixels-là et les logos (...) de reconnaître le logo Renault, ou bien si il y a écrit Renault dans un bandeau là qui passe dans les journaux ... ce n'était pas encore au point mais il y a leur équipe de recherche et de développement qui voulait voir ça. Parce que la présence sur le média, ce n'est pas seulement ce qui a été dit mais le visuel également (...) (Ce sont) les surveillants média (...) qui regardent et qui saisissent en même temps oui. Ils regardent à quelle minute il y a un logo Renault qui passe (...) à telle minute, il y a un logo qui passe ... (...) vraiment c'est ça le business, c'est la présence des marques sur les médias. Pareil pour la presse, hein, si il y a des images, des visuels, c'est indiqué également, c'est indexé

Z : c'est à partir de ça qu'ils font leur reconnaissance ?

ZI : voilà (...) il y a l'image, et c'est surligné ... il y a des conditions aussi pour que ce soit validé en fait, donc si ça parle de faits divers par exemple, si le nom Renault a été mentionné dans un fait divers, c'est à rejeter, mais par contre si ça parle de bourse, ou bien de fusion avec une autre entité, là c'est à prendre. Il y a des consignes quand même spécifiques, il faut valider ou rejeter en fait la coupure ... il y a une coupure qui est directement découpée déjà par l'outil, et qui est validée soit rejetée par l'humain. Parce que la machine ne peut pas faire la différence je pense, d'interpréter. Et je pense que actuellement il y a toujours des opérateurs qui font le même truc (...)"

Louis - Vice-président - Sous-traitant C

Face à ces nouveaux modes de consommation de l'information, les entreprises françaises ont du s'adapter et mettre en place un processus d'externalisation permettant de produire à bas coûts des contenus sur le Web. Ces tâches de saisie, plus complexes que d'autres projets d'annotation, nécessitent des compétences culturelles et linguistiques plus importantes, la maîtrise du français étant essentielle pour produire ces articles.

* * *

À rebours des prédictions sur la disparition du travail, l'analyse d'activités de service numérique situées dans des pays comme Madagascar, montre l'émergence d'industries entières capables de répondre à l'émergence de ces nouvelles activités dans les pays occidentaux. Ces nouveaux besoins deviennent prépondérants à mesure que les algorithmes prennent de l'importance dans de nombreuses sphères économiques et notamment dans le tri de l'information. L'analyse de ces activités permet également de mieux comprendre la place des travailleurs humains dans la production de services automatisés qui s'articule autour de la continuité entre les anciens projets de numérisation et les nouveaux projets de *machine learning*. Ainsi, les projets d'automatisation algorithmique se situent entre deux pôles : le pôle des projets "fortement automatisés" et le pôle des projets "fortement humains". Ces tâches servent à l'*optimisation algorithmique*. D'une manière ou d'une autre, il faut pouvoir compléter ce qui est inachevé par les outils automatiques. De manière directe ou indirecte, les travailleurs de Madagascar mobilisent alors des qualifications linguistiques parfois très approfondies afin de réaliser cette optimisation, sans laquelle le fonctionnement des services numériques est susceptible d'être perturbé.

S'il serait nécessaire d'estimer l'impact de cette numérisation sur l'emploi en France et à Madagascar, nous pensons l'application des technologies d'automatisation reconfigure le rapport au travail dans l'espace (délocalisation) par l'utilisation de technologies de mise à distance du travail. Pour le dire autrement, il sera toujours plus aisé de voir dans l'automatisation le signal d'une disparition du travail et des qualifications si l'on place l'analyse dans un périmètre restreint : la sphère de production des entreprises occidentales.

6.2 Les tâches complémentaires à l'automatisation algorithmique : l'essor des travailleurs des relations clientes dans les pays du Sud Global

Au-delà de la transformation des documents papier en version numérique, la numérisation des relations de vente entre les entreprises et les consommateurs fait émerger tout un secteur externalisé. Comme l'expliquent Batt, Holman, et Holtgrewe (2009), le commerce international en matière de services s'est "intensifié à la fin des années 1990 en lien avec la dérégulation des industries de services, une libéralisation du marché permise par le General Agreement on Trade in Services, et les avancées en matière de digitalisation.". Le cas des centres d'appels devient alors emblématique de ces "laboratoires du travail du futur" (Buscatto, 2002b). Dans

plusieurs secteurs nécessitant une force commerciale, comme les assurances ou la banque, de nombreuses entreprises mettent en place une force de travail pour gérer les relations clients à différentes étapes.

Le début des années 2000 marque également l'émergence du e-commerce et l'essor de la livraison à domicile. Certaines entreprises déjà installées se convertissent à cette nouvelle modalité de vente, d'une part en numérisant leurs catalogues de produits et d'autre part en mettant en place des centres d'appels pour la gestion de la relation client. Le cas de l'entreprise de télécommunication Free est emblématique de cette évolution, puisqu'elle se lance sans point de vente physique, faisant le choix de la gestion entièrement à distance des relations avec ses clients : via un site Internet et un centre d'appels. Progressivement, ces *call centers* se développent, d'abord dans le monde anglophone, puis en France et dans les pays francophones du continent africain ¹⁷⁶.

Les pays comme le Maroc, la Tunisie et Madagascar se retrouvent alors en concurrence pour accueillir cette nouvelle activité. L'accent malgache est alors vu comme un avantage comparatif sur les pays du Maghreb : *"l'accent malgache se fait beaucoup moins sentir, et c'est un facteur de différenciation [pour attirer des clients].* (M.N.A - Ambassade de France à Madagascar). Déjà, dans une enquête réalisée pour le compte de la Banque Mondiale, Olivier Cadot et John Nasir (2001) identifiaient la maîtrise de la langue comme un facteur de développement, au même titre que le faible coût et la qualité du travail. Presque vingt ans plus tard, c'est toujours la "qualité" de l'accent malgache qui a conduit des entreprises à venir y installer des centres d'appels. Les évolutions de la relation client, notamment la part du contact par mail, ont ensuite conduit ces entreprises de sous-traitance à proposer d'autres services.

6.2.1 Le travail de la voix en miroir du travail gratuit des consommateurs

Au cours de notre enquête, de nombreuses personnes interrogées racontent avoir eu des expériences professionnelles comme *téléconseiller* ou *conseiller de vente*. À Madagascar, c'est au sein des mêmes entreprises que les travailleurs annotent des données et conseillent des clients. Comment donner du sens à cette unité de lieu ? Nous pensons que ces activités requièrent des qualifications similaires, car elles impliquent un travail relationnel et un rapport au *script*, aux prescriptions des donneurs d'ordres qui mettent en tension le travail de qualité avec la quantité de tâches effectuées. Enfin, le travail de la voix constitue un très bon exemple de ce qui est induit par l'automatisation. Que cette volonté de transformation se traduise par davantage de travail gratuit pour les consommateurs (par exemple : les bornes SNCF), ou par la mise en place

176. *"Les centres de gestion de la relation clientèle ou centres de contacts, hotline, ou plus communément appelés centres d'appels sont nés dans les années 1960 aux États-Unis d'Amérique avant de s'étendre à l'Australie et à l'ensemble des pays anglo-saxons. C'est ainsi qu'ils arrivent en Europe via le Royaume-Uni dans la seconde moitié des années 1970, avant de conquérir par la suite l'Irlande. Ils y connaissent un développement sans cesse grandissant. Ils apparaissent d'abord dans les secteurs de la finance et des assurances, répondant au développement de prestations de services de plus en plus dématérialisées. Ce n'est que bien plus tard, à la fin des années 1980, qu'ils font leur apparition en France. En Afrique, ils ne connaissent leurs débuts que vers la fin des années 1990, d'abord en Afrique du Sud, puis au Maghreb (Maroc, Tunisie, Algérie, Égypte) avant de s'implanter au Sénégal."* (Tiné, 2021)

d'algorithmes, l'incomplétude de ces nouveaux processus nécessite un travail humain de conseil qui contribue à stabiliser le processus de production du service (SHESTAKOVSKY, 2015). Pour nous, ces activités sont donc profondément indissociables de l'économie numérique, dont elles sont l'un des fondements.

Le rapport au "script" et la standardisation de cette activité comme témoin ou non d'une éventuelle "déqualification" a fait l'objet de plusieurs travaux. Ces derniers questionnent l'organisation et les conditions du travail au sein d'entreprises qui vendent des services liés aux relations clients. Comment les travailleurs vivent-ils leur travail ? Dans quelles mesures s'impliquent-ils dans leur travail ? Comment les organisations les mettent-elles au travail ? Ces questions dessinent une ligne de fracture parfaitement résumée par Calderón (2006) qui décrit le processus de rationalisation des procédés de ventes par téléphone dans une entreprise publique. Selon lui, la différence entre les centres de services et l'industrie repose sur le fait que le produit est le travailleur lui-même : *"En somme, la subjectivité du travailleur, son attitude, ses émotions, sont indissociables du produit."* Dès lors, l'organisation du travail repose sur une tension entre le contrôle exercé par la direction sur la subjectivité des travailleurs et l'autonomie de ces derniers face aux demandes des clients¹⁷⁷. Cette *modulation* de l'implication se traduit par les scripts, par les formations et par des dispositifs de contrôle (écoute des travailleurs...). La nécessité de satisfaire le client de manière qualitative va toutefois limiter cette rationalisation : *"En situation de travail, les téléconseillers ne doivent pas se contenter de répéter un même script. Ils doivent donner au client l'impression qu'ils l'écoutent"* (Calderón, 2006)

Or, comme on l'a également vu dans le cas des projets de numérisation, l'implication du travailleur oscille en permanence entre le niveau de qualité attendu, c'est-à-dire la dimension qualitative de son travail, et des objectifs de performance (nombre de documents traités). Selon (Calderón, 2006), on y retrouve donc un continuum entre d'un côté le strict respect des consignes (le script) et la prise d'initiative du travailleur. Patrick, responsable dans un grand centre d'appels de Madagascar raconte avoir auparavant travaillé pour l'entreprise Eufonie au sein de laquelle il faisait partie d'une équipe de gestion du service client pour le compte de Somfy, entreprise française de domotique. Au sein de cette entreprise, il conseille des clients par téléphone, avant la vente. Après la vente, une autre équipe conseille également le client sur l'utilisation des produits, par téléphone et par mail. Pour répondre aux demandes d'informations sur les produits, les conseillers ont *"un petit script"* qui selon lui *ne sert pas à grand-chose* (Patrick - Responsable centre d'appel), car il est plutôt nécessaire de bien connaître et comprendre les besoins du client. Au-delà du script, les conseillers reçoivent *"une formation avec une représentante de Somfy et donc avec une formation sur le produit et avec des exemplaires de chaque produit"*. En indiquant que *"le script ne sert à rien"*, Patrick illustre une

177. *"Les centres d'appels représentent ainsi l'évolution peut-être paradigmatique du taylorisme vers une nouvelle organisation de type néo-tayloriste qui requiert, à la différence des organisations strictement tayloriennes, une certaine implication de la part des travailleurs ; cette implication étant modulée, nous l'avons montré, par les dispositifs de formation mis en place. Or, dès lors que le temps subjectif de production devient une préoccupation de la direction, cela crée une tension puisque la finalité même du travail devient un enjeu : pourquoi travaille-t-on ? Pourquoi s'implique-t-on dans la recherche de nouvelles manières de procéder ?"* (Calderón, 2006)

dimension perceptible dans l'univers de l'annotation : ce travail repose sur une tension entre le façonnage des subjectivités et la capacité d'initiative des travailleurs (agentivité). Avec une différence fondamentale, les données produites par les travailleurs pour l'IA sont susceptibles de perturber le fonctionnement du modèle si elles ne sont pas au maximum standardisées.

Toujours d'après (Calderón, 2006), cette tension explique également comment cette dimension est *qualifiée* (valorisée) ou non par la direction comme par les travailleurs. En effet, *"le principal problème est que l'enjeu de la finalité du travail, la compétition pour le contrôle du sens accordé à l'activité entre les salariés et l'encadrement, semble pour l'instant circonscrit à l'individu, et n'est pas reconnu par la direction dans la plupart des cas."* Cette non-reconnaissance des compétences qui dépassent du travail prescrit constitue une tension déterminante dans l'évolution de la profession : *"dévalorisante, déshumanisante et répétitive pour les salariés ou, au contraire, une profession qui intègre la dimension qualitative de la relation"*. (Calderón, 2006). C'est à l'aune de cette tension qu'on peut analyser la littérature sur la qualification des travailleurs des centres de services. L'enquête de Buscatto (2002b) sur un centre d'appels décrit également le processus de formalisation et de taylorisation de l'activité de vente par téléphone et les résistances et négociations des travailleurs. Par exemple, le fait que, pour les managers du centre, *"l'adéquation avec le script n'apparaît pas comme un enjeu en soi"*. Le dilemme quantité / qualité, et donc la nécessité de bien servir des clients variés à moindre coût (en réduisant le temps des appels) et les stratégies de résistance des salariés, créent ainsi une limite à la rationalisation du processus de travail¹⁷⁸. Dès lors, le travail *« exige en permanence une interprétation sélective, parfois arbitraire, faisant jouer un rôle décisif à l'expérience de situations antérieures, à l'intuition et à des connaissances idiosyncrasiques. Les méthodes et les savoirs formalisés sont au mieux des lignes de conduites (guidelines) indispensables »* (Gadrey, 1994 cité dans (Buscatto, 2002b)). C'est cette dimension, l'interstice entre le travail prescrit et le travail réel, qui fait de l'activité de vente par téléphone une activité qui n'est pas dénuée de qualification.

Dans une enquête comparative sur trois call centers, Béraud, Colin, et Grasser (2008) mobilisent un cadre théorique dans lequel la qualification *"s'envisage comme un rapport social"* et croise trois variables : le "travail effectué", les qualités des travailleurs et la valorisation de l'emploi (salaire). Les auteurs soulignent la variété des organisations et des activités des téléopérateurs. C'est encore plus vrai pour les "métiers de l'outsourcing" observés à Madagascar. Ils identifient toutefois deux aspects du travail qui limitent la professionnalisation des travailleurs : le "caractère répétitif" et les "organigrammes plats" (*dead-end jobs*). Par ailleurs, ils constatent des variations entre les entreprises sur leur capacité à créer un cadre qui *"favorise ou non la professionnalisation"*.

L'enquête de Cihuelo (2010) va contredire l'hypothèse d'un travail "déqualifié" considérant que l'autonomie

178. Jamie Woodcock considère que c'est la mesure de l'activité qui permet de résoudre le dilemme entre les objectifs quantitatifs et qualitatifs

nécessaire au service des clients est investie par les travailleurs pour développer leur savoir-faire centré sur la satisfaction des clients. Cette "réappropriation de l'activité" participant alors à constituer "une identité collective" faisant "émerger dans les centres d'appels étudiés une figure du professionnel de la relation de service, proche de celle du « professionnel de service public » (Cihuelo, 2010). L'espace d'autonomie laissé vacant par le dilemme entre la qualité du service rendu (et la satisfaction des clients) et la quantité rendue nécessaire par la contraction du coût du travail pratiquée dans les centres de services est ainsi rempli par les capacités d'adaptation des travailleurs. Ce *travail émotionnel* est une des dimensions les plus structurantes du métier d'autant plus lorsque celui-ci est externalisé puisqu'il se double d'une dimension culturelle dans les interactions sociales avec le client : *"l'aspect subjectif de la main-d'œuvre, sa composition qualitative, intervient lors des processus de sélection et de recrutement. À cet égard, signalons l'importance accordée aux « compétences sociales » (...). La solidité psychologique et, bien sûr, la motivation, mais aussi d'autres aspects comme la personnalité du candidat, son énergie, l'enthousiasme supposé, le « feeling », les capacités relationnelles (le ton, le sens de la discussion)." (Calderón, 2006).*

La connaissance du produit et le travail émotionnel rendent selon Patrick, ce travail "assez complexe" et donc "bien rémunéré" pour le pays. Il est également important de noter que ce dernier est titulaire d'un master en droit et en gestion et qu'il parle parfaitement français, compétence pour laquelle il a été testé à son entrée dans le centre d'appels. Cet aspect de la division internationale du travail de services a également un impact sur la notion de qualification du métier, puisqu'il repose alors sur la maîtrise de la langue et du contexte culturel du pays donneur d'ordres, créant ainsi une sélectivité importante. C'est ce que me raconte M.N.A lorsqu'il me parle de son recrutement chez le sous-traitant C : *"Ils recrutait beaucoup de gens qui avaient des liens avec la France, par exemple si tu avais un bac français ou un lien avec la France. Tu avais des questions sur l'histoire de la France, comme les questions qu'on pose pour se faire naturaliser, c'est vraiment pas évident pour un malgache d'être en prise avec tout ça. Sur 50 personnes, on était trois à être pris."* (M.N.A - Ambassade de France à Madagascar). Une autre perspective sur ces centres de services consiste à étudier les compétences valorisées par les donneurs d'ordres et fait le lien avec la littérature post-coloniale. Les travaux d'Aneesh Aneesh (Aneesh, 2015) sur le travail communicatif et affectif des employés des centres de services montrent que c'est grâce à la neutralisation de leur accent et à la *mimesis* (imitation) que les travailleurs rendent possible la communication avec les consommateurs américains. Ils montrent également l'impact sur les travailleurs de l'adaptation de ces centres aux contraintes du marché américain, notamment en termes d'horaires de travail. Dans le cas des centres de services français, Hechiche Salah, Ben Radhia, et Ben Ammar-Mamlouk (2009) et Tine (2021) montrent également ce processus de minimisation de l'identité culturelle des travailleurs. Ce qui se traduit notamment par l'obligation de parler uniquement français ou de s'exprimer sous un autre nom. Toujours à propos des centres d'appels, cette fois aux Philippines, Jan Padios parle d'un processus de "rappel colonial" ("colonial recall"), montrant que les relations post-coloniales sont redéfinies comme une valeur ajoutée pour les entreprises américaines. ROBERTS (2019) évoque quant à elle

une "déconnexion" entre l'environnement culturel des entreprises de centre d'appels, marqués par la culture et la langue des pays occidentaux et le contexte local des travailleurs, ce qui peut aboutir "à la traduction d'un travail bien fait en fonction de l'aptitude d'un employé à passer d'une culture à l'autre et d'une langue à l'autre." (p.46)

L'émergence de ces services sur la Grande Île est corollaire à l'apparition des sites de e-commerce et de la réglementation en matière de coût d'accès aux services. En effet, à l'époque, les numéros sont surtaxés. La loi Chatel du 3 janvier 2008 pour le développement de la concurrence au service des consommateurs puis la directive européenne 2011/83/UE relative aux droits des consommateurs vont complètement changer la donne. La première dispose que les services d'assistance doivent être accessibles à un tarif "fixe et non surtaxé" insistant sur le fait de ne pas faire payer le temps d'attente des utilisateurs avant la prise en charge. La loi concerne tous les secteurs, pour peu qu'ils offrent un service client. La directive européenne rend obligatoire l'affichage d'un numéro de téléphone pour contacter un e-commerçant. Déjà à l'époque de la loi Chatel, les entreprises de services clients comme les e-commerçants pointent les conséquences en matière d'emplois : "Dans un courrier au secrétaire d'État à la Consommation, l'Association française des relations clients (AFRC) demande d'urgence un moratoire, faute de quoi, des conséquences sur plusieurs milliers d'emplois seraient à craindre dans ce secteur où trente-cinq mille emplois sont déjà délocalisés." (Le Parisien). Le directeur du sous-traitant C explique : "on a vu arriver très très rapidement aussi sur la relation client tout ce qui est e-commerçants (...) et la presse [qui] avaient besoin aussi que la relation client soit exportée, parce que ça coûte quand même assez cher [...] Les e-commerçants ont pas eu le choix pour deux raisons : parce que avant il n'y avait pas l'obligation qu'ils aient un service client, (et) ils vendent en masse. (...) Dès qu'ils ont ouvert un service client, ils sont arrivés chez nous, comme Rue du commerce parmi les premiers"

La baisse du tarif d'appel aux services clients pousse finalement de nombreuses entreprises à trouver des solutions moins onéreuses. Pour soutenir les besoins toujours plus importants des clients français, ces dernières vont se tourner vers des entreprises comme le sous-traitant C (les entreprises de sous-traitance). Comme l'indiquent Béraud et al. (2008), le positionnement de ces entreprises est très différent des centres de téléservices internes aux entreprises : "La relation client à distance n'est plus la marge d'une activité principale, mais leur véritable métier." Dès lors, elles doivent jouer sur les coûts tout en maintenant une qualité suffisante pour être rentable. Certaines se sont donc concentrées sur les prestations dites à "faible valeur ajoutée", ce que le Vice-président du sous-traitant C désigne par le "niveau un" : "Si vous êtes sur SFR (...), vous avez de la chance d'arriver ici à votre niveau un alors qu'avant c'était pas du tout dans leur choix. (...) On est un marché de consommateurs, donc forcément, on veut que ce soit moins cher (...)" (Louis - Vice-président - Sous-traitant C) Dans cette organisation, les travailleurs malgaches se chargent des demandes clients considérées comme les plus simples. Reste à des employés français de répondre aux sollicitations les plus techniques ou les plus problématiques, par exemple les clients mécontents. Plus récemment, il observe éga-

lement un phénomène de ”respécialisation” de certaines boutiques en ligne face aux généralistes, que sont Amazon ou Ebay. Dès lors, il est nécessaire de disposer de conseillers susceptibles d’apporter une réelle valeur ajoutée au client. Cette évolution maintient toutefois le besoin de travailleurs à moindre coût, qui vont permettre aux entreprises de garder suffisamment de marges pour financer des conseillers spécialisés.

Encadré n°43 : La transformation du e-commerce et son impact sur le travail à Madagascar

”Il y a (...) une nouvelle tendance, parce qu’avant tout le monde faisait tout et n’importe quoi. Ça a commencé par certainement un peu eBay, mais aujourd’hui, c’est comme sur Amazon, rue du Commerce, ... rue du Commerce c’est un exemple qui est assez flagrant. Quand je les ai connus, ils étaient spécialisés sur le high tech au départ et maintenant on y trouve même des sex toys et des porte-clés (...). Et problème (...) il y a Amazon (...) c’est une machine de guerre. [Face à ça] on a de plus en plus de e-commerçants spécialisés qui se respécialisent. (...) Alors j’ai un exemple (...) c’est un des premiers acteurs européens sur le marché d’accessoires de motos. Et avec eux (...) la partie relation client, on la traite qu’à l’écrit. Et souvent pour une partie donc le mail [même si] les chats aussi se déportent beaucoup. Et à Mada, c’est plus facile de trouver des gens qui savent bien communiquer à l’écrit aussi plus qu’à la voix, puisque la voix, c’est toujours un peu compliqué (...). Et eux se déportent, on a ça pour pas mal. On a un client qui est (...) sur la vente de fleurs dans le nord de la France. Mais ils veulent garder ce côté spécialiste (...). Et ça, on le déporte pas.”

Louis - Vice-président - Sous-traitant C

Finalement, les évolutions réglementaires et la nécessité de maintenir des coûts faibles pour les services clients poussent les entreprises à réorganiser le travail. Les travailleurs des pays du Sud Global s’occupent en théorie des échanges les plus simples. Qu’elles sont les implications de ces évolutions pour les travailleurs de Madagascar? Notre enquête nous a permis d’identifier plusieurs cas dans lesquels ce travail de la voix vient en complément de services automatiques. Le cas le plus emblématique est celui du service client de l’entreprise de livraison de repas Deliveroo. Une littérature conséquente a mis en avant le caractère ”algorithmique” du management, notamment au travers de la notation des chauffeurs. Cette dimension est limitée par les stratégies de résistance ou de bricolage de certains travailleurs des plateformes qui apprennent à contourner l’algorithme (Rosenblat & Stark, 2016). Des recherches soulignent encore la participation des chauffeurs à produire les données qui façonnent les services de travail platformisé (A. A. Casilli, 2019).

Dans le cas de Deliveroo, le service client a été délocalisé à Madagascar. Ainsi, comme le raconte un chauffeur ”si on avait un pneu crevé ou si le client ne répondait pas, on devait appeler Madagascar.” (SudInfo.be). C’est également ces travailleurs qui doivent trancher les différends entre les clients et les livreurs en cas de

problème concernant la livraison : *”En cas de litige avec un client, il faut passer par une centrale d’appel, basée à Madagascar. Dans l’autre hémisphère, un conseiller tranche le différend entre deux personnes qui se trouvent sur le même pas-de-porte. C’est la garantie d’une demi-heure de perdue et c’est souvent au coursier qu’on donne tort.”* (Le Monde) C’est ce type d’expérience que me raconte Stéphane, aujourd’hui chef de projet dans une entreprise d’annotation de données pour l’IA. En effet, il commence sa carrière comme conseiller client et s’occupe ainsi des problèmes liés à la livraison des commandes. Ce faisant, ces travailleurs prennent en charge ce que le service ”automatique” ne peut pas faire. Encore une fois, malgré la tentative de standardisation à l’extrême du service de livraison qui se traduit par ”l’algorithme”, des travailleurs humains doivent trancher en cas de problème¹⁷⁹. Ce faisant, les travailleurs de Madagascar contribuent même à l’exercice d’un ”pouvoir de direction” vis-à-vis des livreurs et constituent une caractéristique du service qui pourrait tendre à considérer les livreurs comme des salariés et non des prestataires de service.

Les services clients et cette ”standardisation” des interactions constituent également un prélude à l’émergence de relations clients automatisées. C’est ce que raconte Patricia, responsable d’une start-up qui produit des ”call bots”, c’est-à-dire des robots vocaux censés répondre aux clients par téléphone. Ce type de produit a par exemple été utilisé par le ministère de la Santé au moment de la crise du Covid *”en complément du service humain”* qui est délocalisé. L’objectif est de ”trier” les appels en proposant des réponses automatiques pour les demandes les plus simples et ainsi de mieux gérer des afflux soudains d’appels, comme c’est le cas lors d’une crise. Pour notre répondante, les services clients sont nécessaires, car les entreprises ont créé de plus en plus d’espaces en ligne pour la vente et pour gérer les interactions avec les clients. Toutefois, cette ”automatisation” de la relation client a des limites dans la capacité des clients à utiliser ces services, ce qui a comme conséquence la nécessité de mettre en place des plateformes d’appels.

Encadré n°44 : La transformation du e-commerce et son impact sur le travail à Madagascar

”La digitalisation s’est faite à grand renfort de mise en place d’espaces en ligne.[Mais], beaucoup de grandes marques sont très déçues par rapport au poids qu’il a fallu mettre sur (...) ces espaces en ligne parce que finalement la fréquentation n’est pas celle qui est attendue. À chaque fois qu’un client appelle un conseiller, le conseiller lui rappelle qu’il peut aller sur son espace en ligne, sauf que le frein majeur c’est la faible récurrence, c’est-à-dire que lorsque vous allez sur votre espace en ligne une fois par an parce que vous avez besoin d’une attestation ou autre, vous ne savez plus où c’est, vous avez perdu votre mot de passe, et finalement pour tout un chacun le téléphone est un accès plus direct, plus simple et instantané. Donc la digitalisation c’est très bien, mais ça a des limites.”

179. *”J’ai commencé au bas de l’échelle, j’étais conseiller client en appel entrant - vous connaissez très bien j’en suis sûr, pour Deliveroo. Donc là, on était responsable des commandes, on s’assurait que les livreurs livrent bien les choses.”* (Simon - Chef de projet - Sous-traitant G)

Patricia souligne par ailleurs la complémentarité entre la standardisation des appels (les "scripts") et la mise en place de robots vocaux : *"Quand on propose de digitaliser une partie des appels téléphoniques (...) on va utiliser (...) tout ce qui a été déployé pour connecter l'espace client (...) on va le réutiliser, donc on va optimiser la digitalisation qui a été démarrée avec l'espace web."* Pour ce faire, la start-up travaille avec son client à partir des scripts des téléconseillers afin de définir *"la typologie des cas d'usage à adopter"* afin de *"construire la conversation"*, c'est-à-dire l'arborescence des questions potentielles et des réponses à apporter. En cas de requêtes non standards, l'appel va être automatiquement transféré à un humain : *"Dès que la question est plus pointue, on a choisi de transférer ces appels à un humain."* Dans le cas du Covid-19, notre répondant prend la différence entre une question portant sur les documents nécessaires si le citoyen souhaite sortir après le couvre-feu et une question portant sur : *"qu'est-ce que je dois faire comme attestation pour aller nourrir mes chevaux au pré ?"* Les interactions avec le call-bot peuvent également être analysées pour proposer d'autres scénarios de réponses, notamment si les conditions du service évoluent.

Ces éléments dessinent une ligne de fracture parfaitement résumée par [Calderón \(2006\)](#) qui décrit le processus de rationalisation des procédés de ventes par téléphone dans une entreprise publique. Selon lui, la différence entre les centres de services et l'industrie repose sur le fait que le produit est le travailleur lui-même : *"En somme, la subjectivité du travailleur, son attitude, ses émotions, sont indissociables du produit."* Dès lors, l'organisation du travail repose sur une tension entre le contrôle exercé par la direction sur la subjectivité des travailleurs et l'autonomie de ces derniers face aux demandes des clients. Cette *modulation* de l'implication se traduit par les scripts, par les formations et par des dispositifs de contrôle (écoute des travailleurs...). La nécessité de satisfaire le client de manière qualitative va toutefois limiter cette rationalisation : *"En situation de travail, les téléconseillers ne doivent pas se contenter de répéter un même script. Ils doivent donner au client l'impression qu'ils l'écoutent"* ([Calderón, 2006](#))

Comme dans le cas du travail de l'automatisation algorithmique, on constate que les travailleurs sont soumis à l'utilisation de l'IA dans leur travail. Nous considérons en revanche que ce travail est "complémentaire". En effet, s'il ne s'applique pas *de manière "substantielle" au fonctionnement des algorithmes*, mais permet au contraire de garantir le fonctionnement du service automatisé en offrant aux consommateurs de l'aide sur leur utilisation. En comparaison d'autres tâches liées à l'automatisation, ce travail est particulièrement *relationnel*, puisque les travailleurs sont directement en contact avec les consommateurs. Au-delà de cette dimension relationnelle qui demeure difficilement automatisable, le rapport des travailleurs au script d'appel et à la qualité de services illustre une tension propre au secteur de services. Il faut que cela coûte le moins cher possible, mais que la prestation soit de qualité au risque de crispier le consommateur.

Or, le travailleur est au centre de la prestation : ce sont ses qualités et ses qualifications qui sont finalement

vendues aux donneurs d'ordres : "dès lors que le temps subjectif de production devient une préoccupation de la direction, cela crée une tension puisque la finalité même du travail devient un enjeu : pourquoi travaille-t-on ? Pourquoi s'implique-t-on dans la recherche de nouvelles manières de procéder ?" (Calderón, 2006). Cela nécessite alors une capacité d'initiative pour sortir du script, en fonction de cas d'usages non prévus initialement, en plus de valoriser des qualifications coloniales (linguistiques notamment). Les entreprises elles-mêmes sont d'ailleurs prises dans cette tension dans leur capacité à associer des indicateurs de qualité et de quantité.

6.2.2 Le travail du clavier : la relation client à l'ère du Web

Avec la poursuite de la numérisation du commerce et l'essor des espaces de discussion en ligne, de nouvelles activités se numérisent, ce qui rend possible leur externalisation dans les pays du Sud Global. Les relations avec les clients se font par mail, par des chats en direct ou par les réseaux sociaux, ce qui rend du même coup nécessaire la modération de ces espaces de discussion. Comme l'explique Louis du sous-traitant C : "la presse s'est démocratisée puisqu'ils ont continué eux à aller plus vite (...) là, on arrive dans les métiers un peu plus avec la fibre, donc y'a à peu près onze ans ... (...) où on a commencé à faire de la modération (...) on modérait des SMS ! et on faisait du traitement de communauté. (...) Et puis forcément au bout de ... je ne sais pas, vers 2016 on va dire 2016, ça commence à complètement décroître les SMS (...) Et là, on est rentré plus dans de la modération".

Dans de nombreux cas, l'intégration des productions écrites des consommateurs aux modèles d'affaire des entreprises va pousser ces dernières à se saisir de ces pratiques. L'apparition des commentaires sur les sites de presse (Smyrnaiois & Marty, 2017), les pratiques de notation des services et produits vendus sur Internet (Beauvisage, Beuscart, Cardon, Mellet, & Trespeuch, 2013) ou la dimension relationnelle des réseaux sociaux sont autant d'expression de l'émergence de nouveaux modèles d'affaires. Ils reposent sur les interactions entre utilisateurs autour des contenus publiés sur la plateforme par des individus ou des marques. Cependant, la prolifération des discours haineux, et les manipulations de l'opinion publique amorcent un "désenchantement de l'Internet" qui conduit les entreprises à faire évoluer leurs pratiques en matière de modération (Badouard (2017)). Dans le cas de la presse, Smyrnaiois et Marty (2017) expliquent que ce mouvement s'est traduit soit par une suppression ou au moins une restriction des espaces communautaires, soit par un renforcement des équipes de modération¹⁸⁰. Selon lui, un "marqueur tangible du statut [peu valorisant] conféré aux commentaires par les rédactions est le fait que leur traitement soit le plus souvent externalisé et industrialisé."

180. "Dès lors, différentes stratégies ont été mises en œuvre par les éditeurs concernant le filtrage de commentaires : la suppression pure et simple des espaces qui leur sont dédiés pour une partie ou pour la totalité des articles ; l'obligation de s'enregistrer ou même de disposer d'un profil éponyme pour pouvoir commenter ; enfin le renforcement de la modération (Goodman et al., 2013) (...). Sans y consacrer des moyens conséquents en interne (...)" (Smyrnaiois & Marty, 2017)

L'affaire Cambridge Analytica et l'ingérence de la Russie dans les campagnes présidentielles américaines et françaises ont conduit à une prise de conscience par les pouvoirs publics des enjeux de régulation de la modération, notamment à travers le "Digital Services Act" (DSA) dans l'Union Européenne ou la loi Avia en France. Selon [Badouard \(2021\)](#), la conséquence principale est l'intensification de la modération par les plateformes nécessitant une évolution des pratiques de modération. Deux stratégies sont alors mises en place : d'une part, diminuer la visibilité des informations jugées peu viables ou haineuses minimisant ainsi le risque de "sur-censure"; d'autre part, parier sur l'automatisation de la modération. Ainsi, le *"paradoxe qui caractérise le travail de modération (...) est que son « invisibilité » fait partie intégrante d'un processus (...) dont le but ultime est de désigner (...) les contributions (...) qui remplissent les conditions de la « visibilité »."* ([Smyrnaiois & Marty, 2017](#)) Tout ceci conduit à renforcer l'invisibilisation du travail de modération déjà identifié par [ROBERTS \(2019\)](#); [Smyrnaiois et Marty \(2017\)](#). D'ailleurs, malgré *"leur rôle décisif dans cette élaboration collective de ce qui mérite d'être publié, les acteurs de la modération n'obtiennent pas « la reconnaissance de leurs savoirs et savoir-faire en termes de compétences, de statuts, de rémunération ou de promotion, par leur employeur, leurs confrères, ou les personnes-ressources"* ([Smyrnaiois & Marty, 2017](#)).

Au cours de l'enquête, de nombreux travailleurs ont évoqué avoir exercé des fonctions de "maileurs", de "téléconseiller par mail" ou de "modérateur". Derrière ces différentes appellations, il s'agit par exemple de prendre en charge des réclamations : *"Si par exemple tu as une réclamation auprès de Air France, il y a quelqu'un à Mada[agascar] qui répond."* (M.N.A - Ambassade de France à Madagascar). Ce type d'activité est également très présent sur les sites de vente en ligne comme Amazon. L'un de nos enquêtés a ainsi pris en charge les réclamations en matière de colis pour un site de e-commerce. Cette tâche est parfois perçue comme moins difficile que les appels, car les travailleurs ont une plus grande distance émotionnelle vis-à-vis des clients, comme l'indique Radera dans un entretien : *"Si c'est des appels entrants c'est plus difficile. Il y a beaucoup de stress. Les gens qui appellent ils sont toujours méchants. Non, ils sont pas méchants, mais ils sont toujours en colère. C'est plus facile que le call, donc j'ai préféré opter pour le chat."*¹⁸¹ Toutefois, cela nécessite un niveau important en français ce qui est le cas de Radera qui a étudié le français à l'école et au contact d'amies françaises.

Comme pour les activités d'appels, il existe des scripts pour gérer les interactions avec les clients. C'est ce que raconte Andotiana, dont le travail consiste à *"voir le panier des clients, voir le produit ou l'objet qui est contenu dedans et de valoriser l'objet en question, afin que le client finalise son achat"*, et ce, en passant par une *"icône de discussion qui est comme un conseil client automatisé"*¹⁸². Pour ce faire, ils reçoivent une "une

181. Cette dimension du travail relationnel effectué par les travailleurs de Madagascar est cohérente avec des recherches réalisées sur les "scripts" des automates conversationnels. Ainsi [Velkovska et Zouinar \(2020\)](#), montrent que les promoteurs de ces modes de communications ont tendance à *"à invisibiliser la multiplicité, la complexité et l'ambivalence des usages réels qui débordent les scripts inscrits dans les dispositifs."* en mettant en lumière *"les processus d'appropriation qui les installent graduellement dans les activités quotidiennes"*

182. *"Parce que une fois que on est sur Cdiscount on peut parcourir plusieurs rayons, ou plusieurs secteurs, que ce soit des biens*

formation spéciale côté trame client, c'est à dire une trame spécifique qu'on doit présenter au client, comment s'engager au sein du client, comment l'inciter à valider son achat". Comme pour le travail d'appel, la pression est importante, ce qui pousse Andotiana a quitté ce travail : "je ne me suis pas senti très à l'aise, vu la pression du travail au quotidien, aussi les charges horaires qui ne sont pas flexibles, mais imposés (...) j'ai décidé d'arrêter." C'est aussi ce que raconte Saroy, aujourd'hui annotateur de données et qui a commencé sa carrière en tant que "téléconseiller" chez ADM Value. Son travail consiste alors à répondre aux mails des clients de chez TotalEnergies et à résoudre leurs problèmes, par exemple sur la facturation. La pression se traduit cette fois par les changements de consignes et par la quantité d'informations à retenir pour réaliser la tâche : "Tout peut changer à n'importe quelle heure. Et le process de TotalEnergies, on doit le connaître par cœur. Il y avait beaucoup beaucoup de choses à retenir. Et même les anciens disent qu'on ne peut pas maîtriser ça. C'est ça la pression chez ADM. C'est dans le travail. Par exemple, on ne traite pas les factures des clients résiliées un jour, et le lendemain il faut les traiter, et puis quelques minutes après, ils vont dire « on ne traite plus ». Ça change tout le temps et il y a beaucoup d'informations que tu dois retenir. C'était ça la pression." (Saroy - Annotateur - Sous-traitant G)

Une autre tâche évoquée par les personnes interrogées consiste à imiter le comportement d'un programme automatique. Par exemple, Patrick a travaillé pour Travelcar, une société française d'autopartage. Dans ce cadre, il effectue un travail à mi-chemin entre le service client et la fausse intelligence artificielle (*AI impersonation*) qui consiste à opérer l'appariement entre la demande de location l'offre de véhicule. Par ailleurs, il a également travaillé pour le compte de SocialShaker, une agence de communication sur les réseaux sociaux qui s'est spécialisée dans la création de jeux sur Facebook ou Twitter. Dans ce cadre, des travailleurs de Eufonie répondaient aux clients via le système de chatbot du site de l'agence ¹⁸³.

Foanoa explique un autre projet de relation client par chat qui porte sur un site de vente de vêtements de seconde main. Sur ce dernier, un encart permet aux clients de faire des requêtes. Ces dernières sont d'abord prises en charge par une solution censée être automatique qui s'occupe de l'introduction de la conversation ("bonjour"), des demandes de base sur le compte du client ("redonnez-moi votre adresse mail") et qui enfin propose différentes catégories qui permettent de trier la demande de l'utilisateur (par exemple : "Demande pré-achat"). Le cas échéant, le chatbot dirige le client vers la Foire Aux Questions (FAQ). Toutefois, il propose systématiquement si le client a besoin d'informations supplémentaires ce qui permet aux clients de prendre

mobiliers, des produits high-techs, c'est tellement vaste ... et puis après on peut sélectionner au hasard « cet objet-là me plaît ». Et le client le met directement dans son panier. Mais il ne valide pas directement son achat, mais il reste encore pas décidé ... et notre rôle à nous c'est de le faire décider d'acheter le produit et de valoriser le produit."(Andotiana - Annotatrice - Sous-traitant B)

183. *"le voyageur laisse son véhicule à l'aéroport et d'autres voyageurs peuvent le louer et eux faisaient de la mise en relation entre les deux. Un client qui appelle en mode « je viens d'arriver à Orly », je souhaite louer un véhicule pour trois jours, avec un siège bébé », on les interroge et ça rentre dans la base de données jusqu'à tomber sur le véhicule qui correspond et ensuite on contact le propriétaire"(Patrick - Responsable centre d'appel)*

contact avec un conseiller client à Madagascar (niveau deux)¹⁸⁴.

Dans ce cas, les travailleurs se retrouvent à gérer les cas qui ne sont pas prévus par le "script" du chatbot ou par la *Foire Aux Questions* (FAQ). Le chatbot constitue donc un outil de triage des utilisateurs, avec un fonctionnement très proche des serveurs vocaux interactifs des centres d'appels. Les clients y choisissent une catégorie et c'est seulement après ce choix qu'ils peuvent interagir avec un conseiller humain. L'enjeu de l'interaction n'est pas de proposer de l'aide, mais plutôt de rendre visibles les ressources existantes consultables directement par l'utilisateur. Reste aux travailleurs humains de régler les problèmes qui nécessitent des échanges plus longs et plus de technicité.

À la différence de la relation client par chat, la modération consiste prioritairement à identifier et supprimer des messages ne respectant pas un espace communautaire (forum, réseau social). Dans certains cas, il s'agit également de répondre aux utilisateurs, comme c'est le cas pour Patrick. Au début de sa carrière professionnelle, il est responsable d'une équipe de modération des photos, des messages et des profils du site de rencontre Meetic. Une première équipe pouvait supprimer le message en disant à l'internaute : « *vous n'avez pas la possibilité de dire ça* » tandis que l'autre s'occupe de vérifier les photos postées par les utilisateurs afin de détecter la présence de faux profils. Après avoir été formés pendant une semaine sur les règles du site et celles qui s'appliquent à la suppression des photos et messages indésirables, les modérateurs devaient se relayer en permanence pour modérer le site. Jo, qui a également travaillé sur le projet Meetic, raconte qu'il ne souhaite plus travailler pour un site de rencontre en raison des contenus publiés par les internautes¹⁸⁵. On trouve également la modération des annonces publiées sur des sites de revente de produits Au-delà de la modération des textes et des images, le travail consiste à corriger les annonces et à parfois les traduire¹⁸⁶.

Un autre projet lié à la modération, apparu lors de nos entretiens concerne un réseau social de publication de vidéos. Confronté à l'explosion des vidéos publiées et face aux nouvelles obligations en matière de retrait de contenus¹⁸⁷, l'entreprise qui gère le réseau social a mis en place un projet d'automatisation de la suppression

184. "M : c'était des chat humains, y'a des requêtes qui tombent ... y'a des personnes par exemple qui veulent vendre ces baskets, qui veulent savoir à quel prix, comment ... Q : mais le client, moi, si je vais sur vestiaire collective, c'était des chatbot ou ... ? M : non c'est des personnes (...) ce qu'on appelle le niveau 2. Ce chatbot là traite en premier lieu « *bonjour, qu'est-ce qu'on peut* » ... y'a des choix 1/2/3/4, etc. ... et dès qu'il ne règle pas (le problème), il monte au niveau deux et là c'est chez nous." (Foanoa - Pôle commercial ComData)

185. "Enquête : Je cherche dans le domaine des modérateurs mais pas dans un site de rencontre. Si je vois un modérateur de vente, ça me suffit, parce que pour le modérateur de vente je pense ce sont des articles qu'on met en vente donc je pense qu'il y a pas des gens qui vont faire des insultes bizarres, ou des pervers. Je me vois pas travailler pour un site de rencontre. Q : Pourquoi ? C'est à cause des photos ? Enquête : Oui des photos vulgaires qu'on voit tous les jours (rire) . Il y a des fois j'ai envie d'arrêter mais quand on trouve pas encore, on reste. C'est ça le problème." (Jo - ADM Value)

186. "Il fallait aussi corriger les fautes (...) Déjà il fallait enlever les fautes, la ponctuation, il fallait catégoriser les produits. Il fallait catégoriser si c'était des vêtements, des assiettes, si c'était des voitures, des terrains. Il fallait corriger l'annonce, parfois quelqu'un pouvait écrire en Français, en Anglais, dans sa langue vernaculaire [et] il fallait traduire (...) si quelqu'un avait mis quelque chose en Anglais, il fallait mettre en français." (Mafo - AIEthic)

187. Et notamment le "Digital Services Act"

des contenus interdits par sa charte. Pour ce faire, elle est passée par une grande entreprise américaine qui propose, en plus d'un service automatisé de retrait de contenus, de nombreux services sur tout le cycle de développement de l'IA : intégration de services "cloud", services d'entraînement de modèles et accès à une force de travail pour l'annotation. Une fois que le modèle de modération est entraîné à la fois sur des cas généraux et sur des cas spécifiques au réseau social, les contenus sont analysés pour être acceptés ou rejetés. Si la marge d'erreur du modèle est trop importante, les vidéos sont traitées par des travailleurs humains. Au sein du sous-traitant G, ils sont presque une dizaine à travailler sur ce projet depuis Madagascar. L'objectif est d'annoter les vidéos de ce réseau social selon des classes prédéterminées comme le racisme, le terrorisme ou la présence de drogues. Concrètement, les annotateurs reçoivent la vidéo via une interface, ils doivent la visionner et l'interpréter puis ils doivent lui accoler la classe correspondante afin de la classer.

Encadré n°45 : L'annotation d'un projet de modération de contenus vidéos 1

S : c'est pour filtrer les vidéos SocialMedia moi je ne (l'utilise) pas. Mais pour SocialMedia, là ils vont juste bloquer les vidéos qui sont assez explicites, et aussi qui disent des choses assez hard comme du racisme, ou aussi le caractère terroriste, des choses assez hard comme ça. Il y a plusieurs catégories que l'on traite, le terrorisme, les spams, les hackers aussi et les images explicites, sur tiktok on en a vu beaucoup ! Et aussi les jokes, le vaping (...) pour créer un algorithme qui filtre directement les vidéos (qui) les bloque directement et que ça passe pas

Z : mais du coup vous ne faites pas de séquençage, rien, c'est juste de la catégorisation de vidéo ?

S : oui oui c'est bien ça, ce n'est pas de l'annotation d'image mais essentiellement de l'annotation de vidéo. Donc on classifie, on catégorise les vidéos, un peu comme je vous ai expliqué, si c'est des images reviews on clique sur review, si c'est asmr on clique sur asmr ... mais là c'est un peu plus complexe parce que là on filtre vraiment des mots clefs, on regarde sur les métadonnées, sur le son, sur la vidéo, sur la gestuelle de la personne sur la vidéo aussi ... (...) s'ils font des insultes avec la main, majeur, ou bien aussi le son, quand par exemple des audio clips avec « suck my dick », (...) des choses comme ça, donc là rapidement ils les filtrent. Et c'est pas donné aussi pour les annotateurs de le faire, parce que là y'a des personnes qui sont vraiment pas à l'aise"

Simon - Chef de projet - Sous-traitant G

Les lots de vidéos arrivent de manière variable et peuvent représenter jusqu'à 1500 vidéos par semaine à annoter. Ces dernières durent entre quelques secondes et jusqu'à plusieurs minutes et sont consultées par trois travailleurs. Deux travailleurs examinent chaque vidéo. Comme l'explique Stéphane, chef du projet : *"il y aura un double traitement sur la même vidéo, un annotateur traite une vidéo et un autre traite la même vidéo. Cette annotation au consensus permet de définir "un score de similarité entre les annotations des 2 annotateurs"* et de faciliter le travail de vérification des annotations qui est réalisé par un troisième travailleur.

Pour réaliser cette tâche, ces derniers ont jusqu'à deux fois le temps de la vidéo, par exemple deux minutes si la vidéo dure une minute. Les travailleurs doivent réaliser trois types de classification. Une première, considérée comme plus simple par les travailleurs, consiste à dire si une vidéo est une "opinion" ou s'il s'agit d'une vidéo humoristique ou informative. Cette première catégorisation sert à identifier les contenus qui pourraient être choquants, mais qui sont informatifs ou qui sont le reflet d'une expression artistique sur des sujets d'actualité ou "d'importance sociale"¹⁸⁸ et qui sont acceptés par la plateforme. La deuxième classification consiste à identifier le thème de la vidéo : est-ce qu'il s'agit de racisme, d'homophobie ou d'une vidéo à caractère sexuel ? Enfin, les travailleurs doivent indiquer si la vidéo est "haineuse" (*hate*) ou "non haineuse" (*not hate*).

Encadré n°46 : L'annotation d'un projet de modération de contenus vidéos 2

Q : « Vous avez combien de labels/classes ? »

R : « Plus d'une dizaine au total. Il y a deux groupes. Le deuxième groupe a le plus de labels. Le premier groupe, il y a des labels comme : opinion, choc, sarcasme, description, not target. Dans le deuxième groupe, il y a : race et genre, à caractère sexuel si ça parlait de haine envers les homos ou autres, et aussi d'autres labels. »

Saroy - Annotateur - Sous-traitant G

Le score de consensus est calculé selon la similarité entre deux annotateurs sur les trois catégories d'annotation. Si, le score est de 100%, les annotations n'apparaissent pas dans la plateforme de vérification. Lorsque le score est inférieur à 100%, cela veut dire que potentiellement "un annotateur a faussé son annotation" ou s'est trompé, les vérificateurs vont les traiter prioritairement. D'ailleurs, Stéphane raconte également que cette vérification sert d'outil de contrôle du travail des annotateurs afin de voir "si un annotateur a besoin de plus de formation s'il refait tout le temps les mêmes erreurs." Selon les périodes, les travailleurs doivent traiter différentes thématiques, qui reprennent les catégories du réseau social. Ainsi, dans la catégorie *stupéfiants* ou *produits interdits*, on retrouve la "drogue, alcool, vaping, smoke" et on retrouve des annotations liées aux "activités sexuelles, c'est-à-dire "tout ce qui est en rapport avec la sexualité, qui peut conduire ou inciter les gens, ou montrer des actes qui ne sont pas appropriés." (Onja - Annotateur - Sous-traitant G) Pour réaliser ces annotations, les travailleurs sont formés au début du projet sur la base du guide d'annotation envoyé par l'entreprise d'IA. l'objectif est d'explicitier les catégories aux travailleurs. Toutefois, les annotateurs ont régulièrement des doutes, notamment sur des sujets d'actualités ou à visée informative. C'est le cas par exemple pour Saroy qui explique avoir eu des difficultés à classer une vidéo représentant une manifestation sur le conflit israélo-palestinien¹⁸⁹.

188. Je reprends ici les termes des règles communautaires de la plateforme.

189. R : « Il y en a(...) pas mal [de vidéos difficiles à classifier]. Par exemple, il y a des gens qui manifestent, qui montrent un drapeau

Pour résoudre ces doutes, les annotateurs font parfois appel au donneur d'ordres à travers un document partagé leur permettant de poser leurs questions. Dans d'autres cas, ils prennent une décision au consensus, soit entre le travailleur et son chef de projet, soit avec l'ensemble des travailleurs, selon le degré de difficulté de la vidéo. Cet élément du travail renforce le dilemme entre la quantité de vidéos réalisée et la qualité de l'annotation. En effet, les travailleurs reçoivent des lots de vidéos qu'ils doivent traiter selon une échéance pré-déterminée par le client. Par conséquent, plus le traitement d'une vidéo est difficile, plus les travailleurs doivent faire remonter les problèmes à leur hiérarchie pour obtenir une réponse, ce qui implique aussi un ralentissement du travail. C'est aussi ce qui explique la mise en place de procédés de résolution par "jugement majoritaire" entre les travailleurs. Ces derniers étant dans le même bureau, il est plus rapide de se mettre d'accord, que d'attendre la réponse du client ¹⁹⁰.

Lorsque des cas difficiles reviennent, les annotateurs constituent leur propre document individuel. De manière régulière, le chef de projet organise des réunions avec les annotateurs comme lorsque les consignes évoluent ou pour standardiser les réponses aux cas complexes les plus fréquents : *"si c'est par rapport à des vidéos spécifiques qu'on a l'habitude de voir, et qu'on veut que les annotations soient homogènes (...) On fait un point pour orienter tout le monde sur les tags qu'il faut mettre."* (Simon - Chef de projet - Sous-traitant G). Ce type de point n'est pas demandé par l'entreprise cliente, mais sert à améliorer le temps et la qualité de traitement des agents en homogénéisant les réponses. Les difficultés de traitement peuvent également surgir de différences culturelles en fonction du pays d'origine des vidéos, en l'occurrence, les pays francophones et anglophones. Ces différences culturelles peuvent notamment se traduire par un rapport différent à la sexualité : *"(Pour la sexualité) Ils ne tolèrent pas ce genre de choses (chez nous), donc on est déjà habitué de ne pas voir ce genre de vidéos concernant la sexualité. Ainsi, on sait naturellement que ça ne doit pas apparaître dans la rue ou autre."* (Onja - Annotateur - Sous-traitant G). Le degré variable de maîtrise de la langue française et la fréquence d'expressions métaphorique renforcent donc la complexité de ces tâches ¹⁹¹.

Dans son livre, **ROBERTS** (2019) montre que les outils automatiques ont permis de faciliter la distinction entre les modérations simples et celles qui sont complexes. Les tâches complexes sont par exemple celles liées à la détection d'un discours de haine. **ROBERTS** (2019) observe aussi que cette distinction permet de

comme celui de la Palestine et des policiers se jettent sur la personne qui montre le drapeau. Ce genre de vidéos est difficile parce que des fois, on dit que ce n'est pas de la haine (...) Je consulte mon supérieur. Je demande au chef de projet. C'est lui qui décide, avec moi, et je classifie selon notre consensus (...) Elle a dit que ce n'était pas de la haine. Il montrait un drapeau, et ce n'était pas de la haine. Donc on met « not hate.» (Saroy - Annotateur - Sous-traitant G)

190. *"Oui, ça nous arrive régulièrement de ne pas être d'accord sur plusieurs vidéos, parce que chacun à son avis. Certains disent que ça ne reflète pas du tout les demandes du client et certains disent que si. Ça se joue dans les petits détails, comme une façon de bouger, de marcher, ou une intonation dans la voix. C'est pour ça qu'il y a des débats sur la décision finale. Si c'est vraiment compliqué, on décide de les mettre sur le fichier et c'est au client de donner son avis. (...) Ça dépend de la vidéo. Si on arrive à débattre et avoir une conclusion positive de la part de tout le personnel ou de la majorité, on le fait directement."* (Onja - Annotateur - Sous-traitant G)

191. *"Il faut vraiment être quelqu'un qui visionne beaucoup de films ou ce genre de choses pour reconnaître directement que cette image est peut-être caractérisée dans des critères que le client a demandé. Et chez nous, dans le pays, la culture est assez limitée pour ce genre de choses."* (Onja - Annotateur - Sous-traitant G)

délocaliser les tâches de modération les plus simples dans des pays à faibles coûts. Cependant, à travers des différents exemples rencontrés au cours des entretiens, on a pu constater la variété des tâches de modération. Les travailleurs malgaches se trouvent en effet bien souvent mis face à des cas particulièrement complexes à traiter.

Encadré n°47 : L'annotation d'un projet de modération de contenus vidéos 3

”il y avait une dame qui versait de l'eau et sur les verres il y avait tout un tas de gens et on savait pas que c'était raciste, elle versait de l'eau elle écrivait sur les verres il y avait les gays, les noirs, les blancs et les moches et les jolis et elle versait de l'eau, elle versait de l'eau dans un verre. En premier lieu on l'a dit aux supérieurs, il nous a dit de ne pas annoter la vidéo, on avait posé nos questions sur le fichier questions et même le client avait des doutes sur la vidéo, et après on a fait un point sur la vidéo et tout et on avait pris la décision de prendre la décision de tout le monde et on annoté la vidéo par démocratie entre collègue. Je pense qu'on a fait raciste, mais surtout sur le fait de la personnalité, raciste sur la personnalité des gens. Nous on pensait beaucoup que la femme faisait des blagues et il y a d'autres qui disaient que c'était une blague, et les autres personnes disent que c'est pas une blague et qu'elle est vraiment raciste.”

Soly - Annotateur - Sous-traitant G

Comme le montre **ROBERTS (2019)** les tâches de modération sont en réalité très complexes et nécessitent une capacité d'interprétation et de contextualisation de la production du message (pays, mode d'expression) par les travailleurs. Ces capacités à comprendre une vidéo publiée sur un réseau social sont aussi renforcées par la présence de toute une organisation du travail qui facilite les échanges d'informations sur les cas les plus difficiles à traiter. Si les acteurs du secteur soulignent la part toujours plus grande des modérations automatiques, des cas comme celui-là nuancent ce constat.

Un autre cas cité par une modératrice de l'entreprise WebHelp illustre ce réagencement entre travail humain et automatisation, celui d'un grand quotidien national français. Deux catégories de travailleurs s'occupent de ce projet. Certains sont salariés et font une relecture a priori de tous les commentaires publiés sur le site du journal. D'autres sont des travailleurs indépendants qui chargent *a posteriori* des commentaires publiés sur les réseaux sociaux du journal. Sur ce second espace, la masse de commentaires a poussé l'entreprise à s'appuyer sur un programme automatique qui identifie les messages qui ne requièrent pas d'intervention humaine : la simple mention d'un ami, les commentaires contenant certaines insultes. L'outil sert également à établir des priorités pour les messages en fonction de la présence de certains mots : "islam", "migrants" ou "gay". Plus ces mots sont présents, plus le message est traité en priorité par les modérateurs de Madagascar. On retrouve donc une séparation des tâches dans laquelle les travailleurs de cette entreprise s'occupent

des tâches les plus complexes. Or comme l'explique, Gillespie (2020) : "du moins pour l'instant, l'écrasante majorité des contenus identifiés automatiquement sont des copies de contenus qui ont déjà été examinés par un modérateur humain. De telles statistiques [sur la suppression de contenus] sont délibérément trompeuses, car elles laissent entendre que les techniques d'apprentissage automatique repèrent avec précision de nouveaux cas de contenus odieux, et pas seulement des variantes d'anciens contenus". L'automatisation algorithmique de ce processus illustre qu'y compris lors des cas de tentatives d'automatisation plus complexes, des travailleurs humains sont requis pour faire fonctionner les outils de modération.

Comme dans le cas du travail des *call centers*, ce travail du clavier montre cette fois l'intrication entre le travail humain et les modèles d'IA. Que ce soit pour venir en aide à des algorithmes de modération de contenus peu performants ou répondre à des demandes de clients sortant du script, l'intervention humaine devient difficile à distinguer des tâches exécutées de manière purement automatique - et invisibles au regard des consommateurs - les travailleurs réalisent pourtant tout un ensemble de tâches qui impliquent des compétences. Pour ces derniers, cela nécessite pourtant un travail de *contextualisation* important. Enjeu commun à l'ensemble des tâches liées à l'automatisation, c'est la compréhension d'une culture inconnue qui devient alors une qualification particulièrement importante, ce qui évoque également le travail de Padios (2018) à propos du processus de "rappel colonial" ("*colonial recall*"). Elle montre en effet que les relations post-coloniales sont redéfinies comme une valeur ajoutée pour les entreprises américaines d'externalisation de services. Les anciens pays colonisés disposent de compétences linguistiques et d'une proximité culturelle avec les pays donneurs d'ordres dont bénéficient les entreprises de services. Ce faisant, le continuum entre numérisation et automatisation participe à la réactualisation des dynamiques post-coloniales en permettant aux entreprises numérisées des pays du Nord de bénéficier des services indispensables au développement de l'IA à moindre coût : sans modération effectuée par des travailleurs des pays pauvres, les réseaux sociaux occidentaux seraient sans doute bien incapables d'assumer la modération des torrents de haine qui s'y déversent.

* * *

Dans cette deuxième section, nous avons analysé plusieurs activités externalisées à Madagascar. Sur la base de l'article de Shestakofsky (2017), nous montrons que ce résultat est généralisable à l'ensemble des activités complémentaires à la numérisation. Nos entretiens soutiennent ainsi l'hypothèse de la continuité d'un travail humain lors des phases d'automatisations algorithmiques. Notre travail permet alors d'observer ce "*travail algorithmique supplétif*" à l'échelle de toute une chaîne de valeur entre la France et Madagascar. Pour nous, ce travail humain se traduit de deux manières. Un travail supplétif qui s'inscrit dans des processus automatisés et qui sert à pallier les manques des modèles, comme dans le cas de la numérisation de textes. D'autres formes de travail complémentaires sont liées aux relations avec les consommateurs, qui servent à apporter de l'aide dans l'utilisation de certains services notamment numériques. In fine, cela

illustre l'intransférabilité complète des connaissances humaines dans les machines.

La mise en place de ces chaînes de production se traduit par une tension entre qualité de leurs réponses et quantité de messages ou d'appels traités, phénomène déjà discuté dans la littérature sur les centres d'appels et qui demeure controversé. Certains considèrent que les travailleurs peuvent se réappropriier l'activité en sortant du cadre formulé par les clients finaux grâce à leurs compétences relationnelles (Cihuelo, 2010), tandis que d'autres enquêtes insistent sur le contrôle exercé sur les travailleurs à travers des mécanismes de mesure de la qualité et de la quantité. Notre travail montre que ces activités humaines, essentielles aux services numériques nécessitent, comme dans le cas de l'annotation, des compétences plus étendues que ce que les discours des donneurs d'ordres laissent parfois entendre. Derrière ce mythe du "travail facile" (Maruani & Nicole-Drancourt, 1989), il y a une minimisation du travail réalisé par les travailleurs déjà partiellement identifié dans le contexte de l'annotation par Irani (2013). Cette minimisation se traduit, comme on le verra dans notre prochaine partie, par de faibles salaires et dans le cas des centres d'appels par une *tachification* partielle du travail à travers des objectifs d'appels qui donnent lieu à des primes. Pourtant, on constate qu'à Madagascar, ces emplois se traduisent par un *travail relationnel* qui se double d'un *travail culturel*, déjà présent lors du recrutement avec la prédominance de locuteur français chez les travailleurs. De la même manière que pour l'annotation, on retrouve une forme *d'agentivité* dans le travail quotidien des employés de centre de services.

Nous montrons par ailleurs la continuité entre numérisation et automatisation à travers des projets qui illustrent les apports du *machine learning* à des projets pré-existants. Cela permet aussi d'illustrer le continuum entre des activités très automatisées et des activités qui le sont beaucoup moins. On retrouve ici l'intrication entre travail humain et travail de l'algorithme, déjà observé par H. R. Ekbia et Nardi (2017) pour désigner les nouveaux mécanismes d'extraction de valeurs qui sont mis en place par les entreprises du numérique. Cette dimension a toute son importance si l'on se rappelle la définition de l'automatisation proposée par Grosin (1960). D'un point de vue sociologique, il considère qu'on ne peut parler d'automatisation que lorsque les produits ne sont pas touchés par la main de l'homme. Or, les exemples cités dans cette dernière partie montrent que la délivrance de services numériques commerciaux implique systématiquement du travail humain. Dans certains cas, les travailleurs sont mêmes impliqués dans des services dont la promesse est l'automatisation complète du processus. Nos résultats illustrent la centralité du travail humain, relégué aux marges de l'emploi, mais pas des chaînes de production.

Chapitre 7 - Les pratiques d'annotations : un travail de classification de données et de problématisation, négocié entre acteurs de la chaîne de valeur

La littérature sur les effets de l'intelligence artificielle a dans sa majeure partie postulée non seulement la déqualification de nombreux emplois, mais également la fin du travail humain (Brynjolfsson et al., 2017; Frey & Osborne, 2017). Trois éléments sont au coeur de cette vision des effets de la technologie sur le travail. Tout d'abord, l'analyse de la "faisabilité de l'automation", sous l'angle de la répétitivité des tâches du point de vue individuel : "Comme les tâches routinières peuvent être cartographiées par des algorithmes, elles peuvent être prises en charge par (...) des machines.". Ensuite, l'idée d'une "destinée manifeste" de l'automatisation qui est présentée comme allant de soi, ces auteurs minimisant les processus sociaux susceptibles de réduire l'impact de la technologie. Enfin, la conséquence de l'automatisation est déterminée par le fait que les travailleurs doivent lutter contre les machines pour les emplois routiniers. En effet, la configuration institutionnelle spécifique à certains pays (système de formation, présence des syndicats) peut limiter l'effet de l'automatisation. Krzywdzinski (2022) a remis en question ce qu'il nomme le "Routine-biased technological change" en développant plusieurs arguments. Le premier concerne le manque de définition des emplois routiniers, et ce, alors que certains métiers dits de routine nécessitent également des compétences de résolution de problèmes et de l'improvisation. Il note également la variation des effets de l'automatisation sur l'emploi selon les pays en fonction de la représentation syndicale des travailleurs, elle-même associée à l'"enrichissement du travail".

Au regard de ces critiques, notre perspective de recherche consiste donc à analyser la qualification requise par ces activités qui complètent la numérisation, au plus proche de ce que les travailleurs font réellement pour alimenter l'IA. Ce faisant, nous contribuons à la littérature sur les processus d'apprentissage des compétences mobilisées dans le travail des données. Pour nous, cela remet en question une tendance de la littérature à analyser ce travail comme "micro".

D'après Stroobants et Singly (2016), "la qualification représente (...) une relation stabilisée entre les qualités supposées acquises par des catégories d'individus et des qualités supposées requises par des catégories d'emplois" (p.63). La qualification peut faire l'objet d'une codification au sein des nomenclatures professionnelles. Derrière les standards de qualification qui s'appliquent à un poste, existent également des qualifications individuelles parfois non prises en compte. La difficulté des travaux portant sur la qualification est alors de bien situer l'échelle d'observation et de mesure de la qualification. Doit-on étudier uniquement la qualification en tant que convention? Doit-on prendre en compte les "qualités individuelles" et comment? Toujours selon Stroobants et Singly (2016) la "tentation est grande, pour les acteurs, comme pour les chercheurs (...) de se

référer au travail effectivement accompli, comme s'il était plus réel que la qualification conventionnelle". (p.69). Or, d'après eux, "la qualification est par définition conventionnelle" et "ne vise pas à reconnaître la totalité des qualités". Ils considèrent ainsi que c'est à travers un contexte localisé, par exemple, au sein d'une entreprise, que les travailleurs font valoir leurs qualités individuelles et que ces dernières sont reconnues comme utiles ou non ¹⁹².

La mesure de la qualification d'un poste de travail est donc marquée par une tension forte entre les qualités qui permettent d'accomplir réellement le travail et les qualités supposées, définies et reconnues par l'employeur, et ce d'autant que le processus de standardisation des qualifications est affecté par les hiérarchies sociales du milieu professionnel des travailleurs ¹⁹³.

L'analyse des conséquences de l'automatisation sur la qualification du travail a ainsi fait l'objet de plusieurs méthodes de mesure, certaines s'appuyant notamment sur la fréquence de tâches routinières. Les postes de travail sont évalués à l'aune des qualifications requises telles que définies par les classifications d'emplois qui servent à la statistique publique. L'objectif de cette partie est d'aborder les qualifications sous l'angle des activités réalisées quotidiennement par les annotateurs de données en suivant une vision substantialiste des qualifications (voir l'introduction du chapitre 6).

Une piste d'analyse des activités se trouve dans les études publiées durant la période de l'automatisation basée sur l'informatique (autour des années 1980) (Du Tertre & Santilli, 1992). L'arrivée de la bureautique ravive en effet les questionnements sur la mesure de la productivité induite par l'intégration de ces technologies dans les postes de travail. S'il est difficile de mesurer cette productivité, certains y voient tout de même une amélioration de l'efficacité des organisations (Ktitareff et Le Tellier, 1987). Cette hypothèse est cependant contestée par Norbert Alter (1985). Pour ce dernier, les conséquences positives de la bureautique sont peu évidentes à démontrer, car une les « gains en efficacité du travail, variable qualitative non mesurable, sont supérieurs aux gains de productivité. » (Alter, 1985). Ce sont finalement les transformations de l'organisation du travail qui changent la nature des tâches, parfois vers un travail plus qualitatif, qui absorbent *in fine* les gains de productivité ¹⁹⁴.

Outre la bureautique, les système-experts émergent également à cette époque comme des programmes informatiques qui disposent de connaissances suffisantes pour fournir une expertise similaire à un travailleur

192. Ainsi, selon eux "utiliser un téléphone, reconnaître un visage sont des actes qui font appel à des savoir-faire élaborés, difficiles à automatiser. Il s'agit pourtant de capacités si répandues et si ordinaires que nul ne songerait, actuellement à les faire valoir pour briguer un emploi" (p.69).

193. C'est tout le sens du travail de Irani (2013) qui montre bien que les entreprises utilisant Amazon Mechanical Turk mettent en place une stratégie de minimisation de la contribution des travailleurs des données pour valoriser leur position d'innovateurs.

194. Une présentation scientifique pouvait ainsi parfois s'accompagner de "transparents" qu'il fallait préparer. L'arrivée de l'informatique permet plus facilement de réaliser ces présentations avec des logiciels, mais, d'un autre côté, les possibilités offertes sont plus nombreuses et les attentes plus importantes ce qui demande également du temps.

humain. Selon [Du Tertre et Santilli \(1992\)](#), pour analyser cette nouvelle forme d'automatisation de manière critique, un axe de travail part de la dichotomie entre le travail explicite, qui serait donc plus facilement codifiable, et une partie du travail dit informel, composé de *tours de main* acquis par les travailleurs de manière implicite et donc plus difficilement incorporables à la machine.

Or, pour produire les systèmes-experts (SE), il est nécessaire de procéder à un travail de codification des tâches. La question est alors de savoir comment formaliser des savoirs tacites, parfois très personnels et donc propres à chaque travailleur. En simplifiant le processus de travail pour le rendre codifiable, ces systèmes deviennent réducteurs et donc par la même biaisés par ces choix¹⁹⁵ et ne restituent donc pas complètement le travail à réaliser. Un article de [De Terssac, Soubie, et Neveu \(1988\)](#) restitue ces enjeux en montrant que les SE procèdent à une simplification du problème à automatiser et donc une simplification des connaissances qui s'appliquent à la codification. Ce faisant, ils ne tiennent pas compte des contextes sociaux et culturels qui donnent pourtant du sens à la tâche à effectuer par un travailleur. Les auteurs nuancent notamment l'idée selon laquelle l'incorporation des savoirs-experts dans les machines permet de se passer de ces derniers. Ils montrent en particulier que les SE s'appliquent souvent à des fonctions de diagnostics pour lesquelles les entreprises ne disposaient initialement pas d'éléments de contrôle¹⁹⁶ : *"En schématisant un peu, on aurait pu dire que l'intervention humaine commence où finit l'automatisation... Cette frontière entre le fonctionnement des automatismes et celui des hommes semble à nouveau mise en question par la génération actuelle de technologies qui s'implantent dans le domaine de la reconnaissance de forme, dans la compréhension du langage ou dans le domaine du traitement des raisonnements"* [De Terssac et al. \(1988\)](#)

Ces nouvelles technologies permettent d'après [Du Tertre et Santilli \(1992\)](#), une *"flexibilité technologique"*. Les entreprises peuvent désormais fabriquer simultanément plusieurs produits, adapter le processus de production plus efficacement, ou devenir capables de répondre plus promptement aux évolutions du marché en adaptant les volumes. Une première conséquence sur le travail porte sur son intensité. Sur ce point, [Du Tertre et Santilli \(1992\)](#) différencient l'intensité *directe* du travail de l'époque tayloriste de l'intensité *connexe*. La flexibilité des machines basées sur l'informatique nécessite un temps d'adaptation, ce qui pose alors la question de l'*amplitude* et du *degré* de l'automatisation. Plus ces deux éléments sont forts, plus l'adaptation demandera du temps aux ouvriers, limitant alors la productivité. Par conséquent, ce qui permet aux entreprises de réellement gagner en productivité, c'est la capacité des travailleurs à absorber ces changements et à les mettre en place dans les meilleures conditions possibles.

Les conséquences sur la productivité mises à part, quels sont les effets de l'automatisation algorithmique sur le travail ? [Braverman \(1998\)](#) développe à cette époque la thèse de la *"polarisation des compétences"* qui

195. Ce qui rejoint une critique désormais classique en sociologie de l'IA

196. [Du Tertre et Santilli \(1992\)](#) parlent également de "boîtes noires" pour désigner ces SE, en indiquant qu'il sera difficile de maîtriser ce qu'il y a dans la machine, ce qui nous rappelle des débats très actuels en matière d'intelligence artificielle

théorise le mouvement par lequel une faible minorité de travailleurs est capable de tirer son épingle du jeu grâce à ses qualifications, tandis que le plus grand nombre se voit déqualifié. Une deuxième thèse part des processus de qualification collective pour en conclure à une requalification des ouvriers, qui émerge de l'automatisation. D'autres sociologues considèrent qu'avec les machines, les travailleurs voient leur espace de liberté et d'autonomie se restreindre à leur détriment et au profit des tâches accomplies par les machines. Par conséquent, on assisterait à l'aboutissement d'un processus de déqualification entamée avec l'industrialisation. Dans ce processus de substitution du capital au travail, les ouvriers sont dépossédés de leurs compétences par leurs hiérarchies et considérés à des machines. Cette explication est reprise par Michel Freyssenet qui parle de *"déqualification du plus grand nombre des travailleurs et de surqualification d'un petit nombre"* (M. Freyssenet, 1974, p. 8, p. 10.)¹⁹⁷.

Zarifian (1983) propose une autre interprétation du phénomène en analysant l'automatisation sous l'angle des *"qualifications collectives"*. Selon lui, il paraît difficile d'analyser les effets des machines automatiques à l'aune de la qualification individuelle et considère qu'il *"est possible de situer l'existence d'un réseau dense et permanent d'échanges d'informations, indispensables à la réalisation de la production et qui constitue de fait (– quand bien même il resterait informel –) un savoir collectif."* Ce faisant, il sort de l'opposition entre savoir concret et abstrait, savoirs codifiés et savoirs informels. En revanche, il observe la poursuite de la séparation du travail et des travailleurs lors de l'automatisation, notamment entre les ingénieurs et les ouvriers. D'après lui, chaque époque de l'automatisation produit de nouvelles formes de travail en réseau qui reconfigurent les rapports au sein de l'entreprise. Les rôles des ouvriers changent, mais cela n'empêche pas le développement de compétences collectives notamment liées à la circulation des informations. Il prend pour exemple le développement des interphones qui permettent aux ouvriers de communiquer et de se coordonner entre eux sur les caractéristiques d'un produit ou les problèmes éventuels de la chaîne de production. Ainsi, au *"degré le plus élevé de modernisation, nous emploierons la notion de "spécialisation des tâches" plutôt que de parcellisation, pour éviter l'image d'une division du travail qui atomiserait les tâches alors que dominant, dans ce système très intégré de production, les exigences de coordination et de communication."* (Zarifian, 1983) Il observe également que l'impératif de flexibilité peut se traduire par la création d'unités de production autonomes et le développement de la polyvalence chez les ouvriers. Il parle alors d'une *remobilisation* des compétences ouvrières. L'intérêt de son analyse est de questionner le caractère dynamique de la production et les compétences qui découlent de ce mode de fonctionnement en matière de communication. Enfin, la prise en compte du caractère collectif des savoirs ouvriers donne à voir une vision plus nuancée de l'hypothèse de la déqualification, en soulignant les mécanismes d'apprentissage collectif qui persistent dans les pratiques ouvrières. Cette perspective théorique ouvre la voie à l'analyse des logiques de *"reprofessionnali-*

197. Du Tertre et Santilli (1992) analysent cette dynamique en indiquant que : *"les machines « ne sont pas neutres », leur conception et leur production correspondent à un certain antagonisme entre le capital et le travail."* On y retrouve une idée, également portée dans la sociologie de l'innovation, selon laquelle les innovations techniques sont *"soutenues et justifiées par le fait qu'elles permettent de réduire les coûts de main-d'œuvre"* (D. MacKenzie & Wajcman, 1999)

sation” des ouvriers¹⁹⁸.

Le développement de connaissances tacites est, d’après Jones et Wood (1984), une limite de l’automatisation. Le propre du travail serait en effet de s’appuyer sur une *”connaissance tacite (qui) s’apprend à travers l’expérience individuelle, (qui) est difficile – souvent impossible – à exprimer dans un langage explicite et formalisé, (et qui) est liée généralement à une situation spécifique”*. C’est également dans cette perspective que Cavestro (1984) analyse la requalification du travail sous l’angle de la maîtrise des technologies¹⁹⁹. Ainsi, l’expérience concrète du travail est plus complexe que, les managers qui établissent les situations d’emplois, les informaticiens qui développent des systèmes experts ou que les catégories utilisées par les sociologues. Ces qualifications tacites regroupent des apprentissages qui découlent de la pratique de tâches *rutinières* et pour lesquelles les travailleurs développent parfois des *”tours de main”*, des *”trucs”*, mais aussi des qualifications de *”coopérations”* reposant sur l’expérience du travail collectif.

Par ailleurs, De Terssac et Coriat (1984) s’attaquent à la *thèse de la marginalisation* selon laquelle le savoir ouvrier risque de disparaître en étant incorporé dans les machines. Cela nie, selon eux, l’activité réelle de l’ouvrier ainsi que les savoirs tacites développés par les travailleurs. Ce serait également réduire l’activité uniquement à ce qui est prévu par les machines. Or, le caractère imprévu et dynamique de la production conduit les opérateurs à réaliser des tâches non prévues. Ce constat conduit De Terssac et Coriat (1984) à souligner les limites de la thèse de la disparition du savoir-faire ouvrier. Selon eux, *”il existe en fait « un autre type d’approche des fonctions ouvrières, fondé sur l’analyse de l’activité réelle, c’est-à-dire des procédures opératoires mises en jeu, des processus mentaux qui les sous-entendent et des régulations qui les accompagnent”*.

Avant même l’émergence de l’automatisation algorithmique, certains travaux voient dans le passage à l’ère de l’automatisation informatique, l’émergence de nouvelles conditions de coopération (*travail en réseau*) entre les travailleurs. Ce travail relationnel repose sur l’utilisation des nouvelles technologies de transmissions d’informations et nécessite une compréhension de l’ensemble du processus de production, liée à sa flexibilité. Le caractère dynamique et adaptatif des industries flexibles passe par ces compétences de coordination et d’adaptation rapide au changement. Ces connaissances non transférables aux machines et pourtant essentielles au processus de production (*connaissances tacites*) sont par exemple liées à la surveillance (interprétation des signaux) ou la programmation des machines (interprétation du processus de production). Ce sont autant de domaines pour lesquels l’automatisation peut induire une requalification des ouvriers (*reconfiguration du travail ouvrier*) ce qui provoque *in fine* une *redistribution* des qualifications au sein de la chaîne.

198. À travers le cas de l’industrie allemande, Kern et Schumann (1984) constatent que l’automatisation peut provoquer un élargissement des compétences ouvrières, du fait de la nécessité, pour les ouvriers, de comprendre l’ensemble du processus de production.

199. *”La programmation, le lancement des programmes, la surveillance et le contrôle des usinages, sont autant d’espaces où se redéfinissent la connaissance et l’usage des machines-outils.”*

Comme nous l'avons expliqué, le traitement de la qualification repose sur une tension entre l'analyse "substantialiste" qui donne la primauté aux pratiques de travail et une analyse qui repose sur l'étude de la valeur sociale de l'emploi. Selon nous, l'analyse de la qualification sous l'angle des conventions sociales, ne tient pas nécessairement compte de deux phénomènes. D'abord, certaines tâches sont réalisables du fait de l'augmentation générale des compétences. Ensuite, certains emplois sont délégués à des populations minorisées, ce qui sert de justification à la dévalorisation de leur travail. Comme le montrent [Maruani et Nicole-Drancourt \(1989\)](#) dans une enquête sur la numérisation de la presse, l'arrivée de femmes recrutées à des postes similaires aux employés masculins du livre s'accompagne de salaires bien moins élevés. Les auteures soulignent l'existence d'un "mythe du travail facile". L'introduction des machines bureautiques permet de soulager les employées de certaines tâches, opérant alors un transfert de savoir-faire : les machines font pendant que "l'employée qui l'utilise est reléguée au rang d'exécutante si transparente qu'on en parle même pas" (p. 61).

Grâce à l'analyse du travail des données, nous proposons une analyse des processus de qualification et déqualification. Ces travailleurs impliqués dans le façonnage autant que dans le suivi de processus de production automatisé par algorithme illustrent alors plusieurs déplacements du travail, dans une direction qui n'est pas celle d'un travail déqualifié. Citant la littérature sur les compétences dans le secteur des services, [Beerepoot et Hendriks \(2013\)](#) souligne la complexité du travail et la variété des compétences (cachées) nécessaires dans les activités de service.

Notre précédent chapitre montre que les travailleurs étudiés dans cette enquête ont une place importante dans des circuits mondiaux de production des services informatiques reposant sur la mise à distance du travail, plutôt que sur l'automatisation. Nos enquêtés contribuent également de manière plus directe à l'intelligence artificielle en tant qu'"annotateur de données". Pour articuler la question de la qualification avec la place des travailleurs dans la chaîne de production, il faut pour cela s'intéresser à l'activité réelle de ces derniers. Qu'il y a-t-il derrière la fonction d'annotateur ? Comment ce travail, qui répond à des enjeux de qualité, est-il organisé ? Quelles sont les autres fonctions impliquées dans le processus de production ?

Dans une première section, nous analysons l'organisation du travail de l'annotation nécessaire pour atteindre les standards de qualité et de quantité requis par les entreprises d'IA. Nous verrons que ces dernières n'externalisent pas des tâches, mais bien des processus de production qui impliquent beaucoup d'autres fonctions que l'annotation. Dans une seconde section, nous nous intéressons aux pratiques quotidiennes de l'annotation en insistant sur la variété des projets et des techniques requises pour annoter. Au-delà d'un travail du doigt, nous verrons qu'il s'agit surtout d'un travail cognitif, loin d'être déqualifié. Derrière le travail du doigt, ces opérateurs qui visionnent et surveillent les données ne sont pas que des yeux humains de l'IA, mais exercent un jugement qui contribue à la *classification* des données et finalement à la *problématisation* des modèles.

7.1 Atteindre des standards de qualité par le management des travailleurs. Comment la formation et l'organisation du travail favorisent le développement d'une expertise dans le travail des données ?

L'automatisation basé sur les progrès de l'informatique, rend donc possible de nouvelles formes de collaboration en réseau. Cette collaboration relationnelle nécessite une compréhension globale du processus de production. L'automatisation de par sa nature incrémentale, notamment avec les nécessités d'adaptations des modèles, repose sur des compétences de coordination et d'adaptation rapides face aux changements dans la production. Ces compétences, non transférables aux machines mais essentielles au processus de production (connaissances tacites), englobent des domaines tels que la surveillance (interprétation des signaux) ou la connaissance des machines (compréhension du processus de production). Il s'agit de domaines où l'automatisation peut entraîner une requalification des travailleurs, entraînant *in fine* une redistribution des qualifications au sein de la chaîne de production.

L'objectif de cette première section consiste à analyser l'organisation du travail d'annotation à travers trois prismes. D'abord, la variété des rôles et des fonctions qui sont nécessaires au travail d'annotation. Au travers de la chaîne hiérarchique reconstruite à l'aide de nos entretiens, nous analysons comment s'organisent les projets d'annotations dans les entreprises malgaches. La formation est un autre élément déterminant dans la capacité des annotateurs à atteindre une qualité et une quantité suffisantes. Les travailleurs y développent leur expertise en abordant toute une variété de notions nécessaires pour produire des données : maîtrise des outils numériques, compréhension des projets et exercice de leur jugement sur la classification des données. Nous verrons que ce processus de production s'effectue en *réseau* (qu'il est socialisé) ce qui entraîne alors davantage un besoin de travailleurs qualifiés.

7.1.1 Au-delà du travail du clic : l'externalisation d'un processus de production intégré

L'organisation du travail d'annotation est très similaire selon les sous-traitants avec une variable importante : la taille de l'organisation. Par le biais de l'étude des entreprises de sous-traitances rencontrées dans le cadre de notre terrain, nous proposons au lecteur de mieux comprendre ce qu'induit le travail d'annotation "*intégré*" au sein de ces entreprises. Comment le processus de production est-il organisé ? Quelles sont les fonctions des travailleurs qui participent à l'annotation de données ? Comment ces travailleurs interagissent-ils avec les employés français des entreprises d'IA ?

De manière générale, ces organisations sont calquées sur la plupart des entreprises du secteur de l'externalisation de services numériques et impliquent donc plusieurs niveaux hiérarchiques. La capacité ou non des différentes couches hiérarchiques à dialoguer avec les clients finaux constitue une variable déterminante dans la division du travail d'annotation. La longueur de la chaîne hiérarchique va également dépendre de la

taille du projet. Dans le cas de projets qui comportent une centaine d'agents, il y a davantage des positions managériales intermédiaires. L'encadrement est principalement assuré par des responsables de production assurent le suivi des projets et l'interface avec les clients. Chacun d'entre eux dirige entre 20 et 30 agents et participe à la traduction des besoins des clients en annotation. Les "team leaders" sont équivalents à des "agents de maîtrise" et gèrent les projets au quotidien en interagissant de temps en temps avec les clients. Cela passe par des contrôles réguliers du travail effectué, ainsi que par un rôle d'aide, en particulier lorsque les travailleurs doutent d'une annotation. Ils dirigent entre 6 et 8 personnes et sont considérés comme plus experts. Les agents dits "simples" s'occupent de l'ensemble des tâches d'annotation. Ils sont amenés à la fois à labelliser des données et à vérifier des données précédemment annotées par leurs collègues ou eux-mêmes. Cette vérification de l'annotation peut également être déléguée au *team leader*, soit entièrement soit en complément, pour effectuer un deuxième contrôle sur les données.

Ainsi, les responsables de production doivent concrètement effectuer des tâches assez variées et qui reposent sur des compétences communicationnelles. C'est ce que nous raconte Jean qui a été en charge du suivi de notre enquête quantitative et qualitative chez le sous-traitant A, ce qui nous a permis d'échanger avec lui à de nombreuses reprises. Il s'occupe d'abord de recueillir les besoins des entreprises clientes au lancement d'un projet et d'organiser des réunions pour le suivi. Ces réunions permettent de faire remonter des difficultés d'annotation aux clients et, de son côté, de s'enquérir des éventuels changements du projet. D'autre part, les tâches de Jean consistent à communiquer les informations liées au travail quotidien aux annotateurs que ce soit sur les conditions quotidiennes de travail ou sur les projets²⁰⁰. Il a un rôle de gestion des ressources humaines qui consiste à évaluer le travail des annotateurs et doit construire les formations lors du démarrage des projets²⁰¹.

Chez le sous-traitant B, une entreprise plus importante, la production est cette fois divisée en deux pôles. Le premier emploie uniquement des annotateurs en *freelance* sous la supervision directe des "team leader" du pôle vise à prototyper les projets entrants. Dans les entreprises plus petites, cette phase est prise en charge par les dirigeants des entreprises avec l'aide des annotateurs les plus expérimentés. Les projets qui y sont menés sont soit des tests de faisabilité, soit des projets de courte durée (deux mois maximum). Cette phase sert à mettre en place un guide d'annotation en fonction du cahier des charges du client. Sur cette base, des documents de formation sont produits pour aider les annotateurs. Cette étape permet également d'estimer la charge de travail nécessaire à la réalisation du projet. C'est à ce moment que les échanges avec le

200. "je fais aussi la coordination des projets avec les clients. Je fais des suivis, des évaluations, et prochainement je vais former de nouveaux collaborateurs. Donc, je vais concevoir des formations, les délivrer et faire des suivis." (Jean - Conseiller RH - Sous-traitant A)

201. "S'il y a des informations importantes, par exemple la propriété au travail, je lance une alerte. J'envoie des messages sur des plateformes comme Facebook. On crée des groupes Facebook ou Hangouts (avec Google chat) et je fais des communiqués aux collaborateurs (...). Je collecte des informations et je fais en sorte que ces informations soient reçues par les cibles, que ce soient les collaborateurs ou la direction" (Jean - Conseiller RH - Sous-traitant A)

client sont les plus nombreux. Il convient notamment de préciser des éléments du guide d'annotation afin de minimiser les ambiguïtés lors de la production proprement dite. En effet, les donneurs d'ordres ne disposent pas nécessairement de guides d'annotations qui recoupent l'ensemble des situations contenues dans leurs données. Dans certains cas, ce pôle de test contribue donc de manière substantielle à la construction des guides en repérant des éléments non prévus dans les données. Il sert aussi à affiner le prix de la prestation, ainsi qu'à estimer les besoins sur le plan des ressources humaines, y compris sur certains projets pour lesquels il est nécessaire de procéder à des recrutements complémentaires.

Encadré n°48 : Les étapes avant le lancement d'un projet d'annotation

Z : c'est quoi du coup les différentes étapes de l'*onboarding* d'un projet ici ?

M : ils décrivent (...) ce dont ils ont besoin du coup nous on réfléchit (...)ce qu'on peut faire, est-ce qu'on a la plateforme pour le faire, est-ce que on a les ressources pour le faire, les personnes pour faire le travail ... parce que parfois ce n'est pas très évident. (...) Quand c'est faisable, on leur propose un prix, soit ils acceptent soit ils n'acceptent pas, et là on négocie. Si c'est okay on lance le pilote d'abord, et quand le pilote est okay (...) on lance la vraie production. Parfois ce sont des projets de quelques mois, parfois ce sont vraiment des projets en CDI, ça dépend des clients.

Mety - Chef de projet - Sous-traitant B

Cette phase pilote permet également de tester les agents et à estimer la faisabilité des tâches, en fonction des critères de quantité et de qualité requis par les clients. Elle sert également au recrutement d'annotateurs freelance, qui seront testés, formés et sélectionnés avant de participer au projet²⁰². Les projets prototypes validés par l'entreprise de sous-traitance et par le client, ont vocation à être pérennisés, et à basculer sur le pôle production. Cette phase implique donc le contrôle de la qualité des annotations pour éviter au maximum les problèmes de compréhension des guides d'annotations²⁰³. Dans ces cas-là, ils ont vocation à être complétés, afin de tenir compte des erreurs récurrentes et des difficultés de compréhension des annotateurs. Cette phase sert également à identifier les annotateurs les plus compétents, afin qu'ils effectuent le contrôle qualité du projet lors de sa mise en production. Ce contrôle est nécessaire tant il est fréquent que les clients aient des attentes irréalistes ou une idée vague du travail d'annotation. Le rôle de ces équipes-tests est également de faire circuler les informations importantes du cahier des charges aux autres pôles et

202. "Tout d'abord on forme les agents, on lance un avis de recrutement et on forme les agents. On forme tout le monde d'abord, après on fait une sélection pour les agents qui sont les plus compétents. Et quand on a vraiment des personnes qui ont travaillé dessus, ils commencent - on appelle ça l'enrichissement, le premier traitement en fait." (Meva - Team leader - Sous-traitant B)

203. "après (...) on fait du contrôle qualité [CQ], parce que normalement le pilote ne devrait pas durer très longtemps, juste quelques jours ou quelques semaines, un mois ou 2 mois maximum. Donc on fait du CQ au début surtout."(Meva - Team leader - Sous-traitant B)

notamment à celui qui est chargé de la formation²⁰⁴. C'est au sein de ce pôle que sont ensuite rédigés les documents et les présentations nécessaires pour former les agents aux projets.

Encadré n°49 : Le passage d'un projet en production sur le long terme

Z : Comment ça se passe quand vous faites une transition en CDI pour un projet ?

M : du coup c'est là qu'on forme les personnes dans le département de formation, parce que nous n'allons pas former les personnes qui sont en CDI, ça doit être le département formation (...). On leur passe des informations et ils vont constituer leur groupe, c'est ça qu'ils passent après aux gens qui vont vraiment travailler à long terme sur le projet. Dans un premier temps, on regarde encore le traitement des agents pour voir si ça correspond aussi, si les informations sont vraiment passées ... et ce n'est que après que on voit que tout est okay, qu'on décroche.

Mety - Chef de projet - Sous-traitant B

Au sein du sous-traitant G, une entreprise de 60 employés, un chef de projet s'occupe de réaliser des tests sur les projets. Comme dans le cas précédent, ce processus sert à déterminer la vitesse, la faisabilité de la tâche ainsi que le prix. C'est aussi lors de cette phase que les annotateurs découvrent les consignes du projet : *"[les clients] nous donnent les consignes en même temps pendant qu'ils expliquent à peu près le déroulement du projet (...) après ils nous envoient les consignes (et) ils (...), ils expliquent un peu comment on doit procéder, comment on doit annoter, comment on doit faire les choses."* (Simon - Chef de projet - Sous-traitant G). Au démarrage du test, les annotateurs ont également accès à un fichier sur lequel ils peuvent poser des questions aux clients²⁰⁵. De leur côté, le rôle des chefs de projet est de dialoguer avec les clients pour améliorer le guide d'annotation en fonction des retours des annotateurs : *"on revoit les trucs qui ne vont pas sur les consignes, les choses qui se contredisent ... Des fois les clients se contredisent (...), on est là au milieu, on essaie de synthétiser le tout."* (Simon - Chef de projet - Sous-traitant G). Cette phase de test prend au minimum une semaine et implique aussi les chefs de projet sur la phase de formation. Ainsi, Stéphane raconte qu'il prend du temps pour expliquer certains détails des guides d'annotations aux annotateurs tous les matins au démarrage des projets²⁰⁶.

204. *"quand il y a la montée en compétence des agents, il y a des agents qui se démarquent plus des autres, et là on passe à ces agents-là le CQ (...), et on fait juste du QA après, on analyse juste en fait après un CQ. En général c'est ça, on supervise. Au début on est vraiment dedans, on fait vraiment le CQ; on fait les corrections si besoin. Après, à la longue, quand ça va mieux, on supervise juste."*(Team leader prototype - Sous-traitant B)

205. *"Là, quand on va commencer, on crée les fichiers questions, ce qui est essentiel, les feuilles de temps aussi qui sont essentielles pour le projet."* (Simon - Chef de projet - Sous-traitant G)

206. *"Après cela on traite ... il y a à peu près 3 ou 4 jours où on pose beaucoup de questions, après on a tous les cas, on a à peu près tous les cas. Ce n'est pas obligé mais moi je le fais toujours, je crée un fichier PowerPoint pour bien illustrer, pour prendre tous les cas. On crée des slides, après on les renvoie aux annotateurs, on fait des briefs tous les matins, pas plus de 5 min, 5 à 10 min pour ne pas perturber la production."* (Simon - Chef de projet - Sous-traitant G)

Lorsque cette phase est mal réalisée ou qu'elle ne dure pas assez longtemps, les projets peuvent évoluer alors qu'ils sont déjà bien avancés. C'est le cas d'un projet de détection d'objets aériens que nous raconte Stéphane, dans lequel un élément important n'avait pas été identifié, ni par l'équipe d'annotateurs, ni par le client : *"Les consignes sont généralement un peu floues aussi. Ça ne recouvre pas toutes les situations, des fois si on ne pose pas les bonnes questions (...) sur [le projet] on a eu ces soucis-là. Donc là on a posé la bonne question à mi-chemin du projet ... même le client n'avait pas soulevé le point"* (Simon - Chef de projet - Sous-traitant G). Ces changements de consignes induisent un travail supplémentaire, puisque les annotateurs doivent alors refaire les tâches : *"bien sûr et ça retombe sur nous après, c'est toujours comme ça."* (Simon - Chef de projet - Sous-traitant G). Au sein des petites structures comme le sous-traitant G, le faible nombre d'agents dédiés aux tests des projets et à la formation des agents renforce justement ce genre d'incompréhension : *"Mais ici je suis en même temps chef de projet et reviewer aussi, je touche à tout. Donc c'est ça le souci pour un gros projet comme ça (...). Là ils ont expliqué que c'est pas un projet à long terme, donc on n'est pas obligé d'avoir un reviewer, mais les tâches sont vraiment très difficiles (...)"* (Simon - Chef de projet - Sous-traitant G)

La phase de test a donc plusieurs fonctions essentielles dans l'organisation du travail d'annotation : tester le projet, tester les agents, clarifier les consignes des clients. Ces fonctions sont réparties entre plusieurs travailleurs de la chaîne de production. En dehors des annotateurs qui testent le projet, les chefs de projet des entreprises les plus petites cumulent le travail de test, de recrutement et de formation des travailleurs. Dans d'autres cas, ce travail est réparti entre plusieurs agents dans les départements de prototypage et de formation. Les difficultés identifiées par Stéphane au sein d'une plus petite structure la variété des fonctions assurées par les entreprises d'annotation et qui dépassent le cadre de la simple tâche d'annotation.

Le secteur de l'annotation, comme le reste des activités de l'externalisation de services numériques, repose intrinsèquement sur la vente d'une force de travail capable de comprendre les demandes des clients en contribuant au mieux à la réalisation des finalités de la tâche finale. La nécessité d'aller vite pour produire, mais en le faisant avec un niveau de qualité suffisant est à l'origine d'une tension qui incite les organisations du travail à mettre en place une chaîne de vérification des tâches. Celle-ci implique des compétences beaucoup plus qualifiées : interpréter telle ou telle image, communiquer avec le client pour clarifier les consignes, gérer des annotations plus complexes que prévu en adaptant les supports de formation.

Nous venons de décrire l'organisation du travail d'annotation en analysant le processus de production qui est mis en place au sein des entreprises sous-traitantes. Ce faisant, nous avons commencé à esquisser une description des différentes fonctions des annotateurs. Jusqu'à présent, la littérature a identifié des rôles types selon la contribution de ces travailleurs à l'intelligence artificielle (Tubaro, Casilli, & Coville, 2020b). Nos résultats montrent que les entreprises n'externalisent pas une "tâche", mais bien un processus de production entier, ce qui implique également l'externalisation de toute une partie du travail de gestion de

projet et de gestion des ressources humaines à Madagascar. Ces fonctions sont au service d'une production centrée sur la qualité des annotations rendue nécessaire par la sensibilité des modèles aux données. L'annotation est alors séparée en deux phases. D'abord, le test et la conception du projet articulent les clients et les entreprises sous-traitantes autour de la définition des concepts qui s'appliquent à l'annotation et autour de la recherche de l'organisation du travail la plus pertinente. Ensuite, la phase de production, à proprement dite, qui concerne alors les annotateurs, leurs responsables ainsi que les interlocuteurs du côté du client. Au cours de cette phase, le projet est encore amené à évoluer, en fonction de ce que contiennent les données.

Ce résultat implique de mieux tenir compte de la dimension socialisée et collective de la production d'intelligence artificielle. Au cours de ces étapes, le modèle d'IA évolue en permanence, travaillé par les différents acteurs de la chaîne. Ce premier élément invite à reconsidérer la place des annotateurs dans la chaîne de sous-traitance : être au bout de la chaîne ne signifie par pour autant ne pas avoir d'impact sur le modèle d'IA. Bien au contraire, nous avons constaté que cela passe par une organisation de l'activité censée faciliter les remontées des annotateurs concernant les problèmes éventuels liés à la formalisation des annotations. L'annotation ne se réduit donc pas uniquement au travail du clic, dans le sens où elle comprend bien plus que des tâches uniques effectuées par des travailleurs isolés les uns des autres. Au contraire, le travail des données est une activité socialisée faisant l'objet d'une coordination assumée par des agents aux rôles bien définis. Selon nous, cela interroge quant à la manière dont l'IA est contrôlé. Si l'on se place dans la perspective de la régulation européenne sur l'IA qui prévoit un système de déclaration des choix de conception des modèles d'IA considérés comme risqués, comment documenter ces phases partagées entre plusieurs acteurs aux quatre coins du globe? Comment examiner des phases de production dont nous commençons seulement à découvrir l'organisation et les implications sur le modèle final?

7.1.2 Des rôles et des compétences diverses pour produire collectivement les annotations : une approche par les qualifications

La plupart des entreprises d'annotation sont organisées autour de la division entre les activités d'annotation et celles de contrôle, assumées par des personnes distinctes. C'est le cas chez le sous-traitant A sur un projet d'annotation de plats cuisinés : *"on se divise par deux, il y a les annotateurs et les contrôleurs et lorsque l'annotation est terminée, les contrôleurs vont contrôler est-ce que c'est vraiment ce type de plat."* (Mialo - Annotateur). On retrouve ici une distinction classique déjà largement identifiée dans le travail industriel, entre une fonction de production pure et un travail de contrôle. De manière générale, cette fonction de surveillance des tâches est plutôt occupée par des travailleurs plus expérimentés. Cependant, ce travail n'est pas systématiquement valorisé d'un point de vue salarial.

À cette division classique entre la production et le contrôle s'ajoute des postes spécialisés, notamment dans les projets plus complexes. Chez ce même sous-traitant, il existe ainsi une équipe de "spécialistes" ("équipe

VESTA”) sur un projet d’annotation de plateau-repas qui réalise un contrôle supplémentaire des erreurs réalisées par les autres travailleurs. Cette équipe est ”composée de 5 à 6 personnes et chaque personne a un tag respectif avec lequel ils se sont familiarisés, par exemple une personne est familiarisée au plat de résistance. Et ensuite on envoie les images en externe.” (Yanni - Chef de projet - Sous-traitant A). Comme les team leaders, ils servent également à résoudre des doutes en matière d’annotation.

Ces équipes se rapprochent de ce que d’autres entreprises plus grandes (sous-traitant B) appellent le contrôle qualité, généralement composé d’agents beaucoup plus expérimentés et formés qui est séparé de la production. Leur rôle est d’effectuer un contrôle supplémentaire des annotations. Quant aux annotateurs du bas de l’échelle, ils s’occupent, soit de produire des annotations, soit de les contrôler. Les ”team leader” ont la charge de contrôler certaines tâches et de conseiller leur équipe sur la réalisation des annotations. Ensuite, le département qualité réalise des contrôles manuels sur des échantillons de données²⁰⁷. Dans certains cas, les agents du pôle font des retours directement aux équipes d’annotations. Dans la plupart des cas, un contrôle final est réalisé par des personnes dédiées au sein des entreprises d’IA et se fait manuellement ou automatiquement.

Au cours de l’enquête, nous avons constaté l’existence de travailleurs qui, de manière informelle, deviennent des ”référents” sur les projets. Il s’agit d’agents plus expérimentés et plus actifs sur les groupes de discussion entre travailleurs. Ces groupes de messageries internes (Slack, Skype...) permettent aux agents de faire remonter des difficultés et éventuellement de demander l’avis de leurs collègues. Certains agents deviennent alors des référents sur ces messageries, car ils participent à résoudre des blocages. C’est le cas d’Ando, qui sera d’abord considéré comme un référent par ses collègues tout en restant annotateur simple, avant de devenir officiellement un ”team leader”²⁰⁸.

Encadré n°50 : Le rôle des annotateurs plus qualifiés

A : (...)si les agents se sentent en difficulté ou bien si ils ont un blocage (...), je peux les aider (...) et en parallèle, je gère des techniques d’approches plus efficaces plus précises pour faciliter leur identification (des problèmes) (...) En parallèle je reste toujours opérateur, mais à la fois je suis aussi devenu un leader de projet. Et en tant que chef d’équipe, comme j’ai dit précédemment, j’étais responsable du dispatch des tâches, suivi du traitement et avancement des tâches, du départ jusqu’à

207. ”Après [l’annotation] il y a des agents qui font les contrôles qualités, ce sont eux qui vérifient les classifications, si elles sont fausses ou pas et de trouver les oublis. Et après on les livre au client et ils font des QC aussi, automatiques ou manuels, sur des échantillons et ils nous donnent des retours, des remarques, des erreurs et après ils nous font des repasse sur le batch.” (Fidy - Cheffe d’équipe - Sous-traitant B)

208. Ici, Ando permet également de faire un rappel sur l’organisation du travail d’annotation au sein des grandes entreprises du secteur. Le projet sur lequel il travaille, en l’occurrence l’annotation d’éléments liés à la guerre sur des images satellites, emploie une centaine d’agents. Par conséquent, cela peut expliquer une plus grande finesse dans la chaîne hiérarchique (plus de strates) et l’émergence de rôles ”informels”

la validation de la livraison de la tâche demandée par le client

X : C'est quoi la différence avec chef d'équipe ?

A : oui c'est autre chose, c'est juste un terme qu'on emploie sur notre projet en fait

X : c'est informel en fait ?

A : oui (...)

Ando - Team leader - Sous-traitant B

Il existe enfin des annotateurs de "backup" qui interviennent sur plusieurs projets. Chez le sous-traitant B, ces travailleurs polyvalents servent à gérer les pics éventuels de production, par exemple, lorsqu'un client réclame soudainement à son sous-traitant de traiter plus de données. Cette polyvalence se traduit par la maîtrise de plusieurs techniques d'annotations : "détourer" (encadrer) des objets au *pixel près* ou placer des cadres sur les éléments d'une image... Cela nécessite également de maîtriser les spécificités de différents projets ce qui nécessite une compréhension fine des guides d'annotations ²⁰⁹.

Ce principe de polyvalence fait également partie de l'organisation du travail chez le sous-traitant G, mais sous une autre forme. Dans ce cas, c'est le dirigeant de l'entreprise qui indique préférer que les annotateurs changent régulièrement de projets afin qu'ils développent leur polyvalence et qu'ils montent en compétences de manière plus générale sur l'annotation. Le rôle des chefs de projet est ici plus direct en l'absence d'un département dédié au prototypage des projets, puisqu'ils s'occupent dans un premier temps de la prise en main des projets et se chargent de modifier les guides d'annotations et de former les agents ²¹⁰. Durant ces phases de démarrage, le rôle du chef de projet est déterminant de par sa capacité à traduire les enjeux des tâches à son équipe d'annotateurs. Il peut s'agir de détailler les tâches en expliquant un label ou un type d'annotation ou bien d'expliquer la finalité du travail. C'est ce que nous explique Angy à propos d'un projet d'annotation lié à la détection de plats et de nourriture : "Q : ça consiste à faire quoi chef de projet ? R : former les gens, pour qu'ils soient capables de bien comprendre [le] travail, (...) ce que le client attend de nous (...) comme AIAlim par exemple, avant ils ne connaissaient pas, « ah il faut juste retenir les noms des ingrédients et tout ça », mais ce n'est pas ça ! on fait de la base de données graphiques pour une application, je leur explique que c'est comme ça que l'on doit faire (...)" (Angy - Chef de projet - Sous-traitant G).

Une fois la phase de test réalisée, ils organisent le travail au quotidien et participent à la correction des tâches. De leur côté, les annotateurs simples produisent les données avant de laisser les contrôleurs vérifier les annotations : "On envoie les annotations à des reviewers, qui vont corriger les annotations. Ensuite on envoie au client, qui regarde si les annotations sont bien, tout ça. Si il y a des erreurs que le reviewer n'a pas vu, le client

209. "A : ici, backup c'est une personne polyvalente qui peut être ... qui peut aider en termes de quantité sur d'autres projets, en termes d'effectifs. Il doit être aussi à la maîtrise de plusieurs projets à la fois." (Ando - Team leader - Sous-traitant B)

210. "C'est le chef de projet qui prend le projet en premier, il fait les annotations tout ça, ensuite il va former les annotateurs, on fait les annotations tout ça." (Haga - Annotateur - Sous-traitant G).

renvoie les images en lui disant qu'il faut rectifier ça." (Haga - Annotateur - Sous-traitant G). Lors des projets, l'équipe managériale a également un rôle de coordination de la communication entre les annotateurs mais aussi entre les annotateurs et les clients. Angy explique que sa journée type consiste à utiliser des outils comme Slack ou sa boîte mail pour *"passer les consignes, les remarques, les retours du client envers l'équipe. Oui, c'est ça en premier."* (Angy - Chef de projet - Sous-traitant G). Ensuite, il doit également *"améliorer la production, revoir tous les fails [les échecs], les lacunes de chacun des annotateurs, pour être plus productif, en (termes de) qualité et quantité, vis-à-vis de ce que le client attend."* (Angy - Chef de projet - Sous-traitant G). Cette deuxième dimension de son travail se divise entre le suivi du travail des agents et la coordination des changements de consignes, surtout dans ce cas où : *"le projet est difficile parce qu'il y a des consignes qui changent tous les jours"* comme sur AIAlim (Angy - Chef de projet - Sous-traitant G).

Au cours de son entretien, il évoque bien la nécessité de garantir à la fois la qualité et le volume de données produites. Cette dimension a pour conséquence d'introduire plusieurs couches hiérarchiques, qu'on ne retrouve pas nécessairement sur les plateformes numériques de travail (PNT) de type "marché pur" comme Amazon Mechanical Turk. Ces fonctions servent à la fois à maîtriser les concepts liés aux données à annoter et qui ne sont pas systématiquement expliqués par les clients, à intégrer les nouvelles données, c'est-à-dire celles qui émergent pendant le processus d'annotation, et les ambiguïtés liées aux problèmes de compréhension des annotateurs.

Enfin, Angy illustre également son rôle à travers les interactions régulières qu'il a avec sa cheffe de projet française. Il discute régulièrement avec Anaïs qui gère depuis la France les discussions avec le client final ²¹¹. Cette dernière précisera d'ailleurs dans un entretien, que son rôle consiste à "filtrer" les demandes et les retours des annotateurs afin de ne pas *"embêter le client"*. Son rôle, assez rare en comparaison avec d'autres entreprises et est d'ailleurs lié à la difficulté de ce projet qui est illustré par plusieurs éléments : le nombre d'annotateurs, la durée du projet et la difficulté à détecter des aliments qui sont très variables en fonction des images. Ce travail communicationnel est donc distribué au sein de la chaîne de production. En revanche, plus le projet est important en nombre d'agents et difficile sur le plan des annotations, plus les chefs de projet vont concentrer ces tâches. Les annotateurs conservent toutefois la possibilité de transmettre des retours aux clients à travers des outils de messagerie.

Le sous-traitant G a également un centre de production hors de la capitale. Ce dernier, implanté à Fianarantsoa, une ville située à deux cents kilomètres au sud de la capitale, est une association qui emploie des travailleurs éloignés de l'emploi. Lors d'une réunion avec son directeur, ce dernier nous explique que ce centre n'a pas forcément vocation à traiter les tâches les plus difficiles, notamment parce qu'on y retrouve

211. *"C'est nous 2 qui faisons les échanges. C'est elle qui contacte directement le client. Et si j'avais des questions, je les pose avec elle, et elle, elle transmet au client. Et le lendemain, elle fait les retours sur toutes les choses que l'on a demandées"*(Angy - Chef de projet - Sous-traitant G)

des travailleurs moins bien formés au français et moins compétents en informatique. Les entretiens réalisés avec ces travailleurs ont d'ailleurs été particulièrement difficiles à mener, car ces derniers ne parlaient pas français et que notre traducteur ne comprenait pas intégralement le dialecte utilisé par ces travailleurs. Au moment de l'enquête, le sous-traitant G a pourtant été obligé d'avoir recours à ces travailleurs pour un projet considéré comme plus difficile, à cause d'un pic de production. Les travailleurs doivent y annoter des séquences vidéos image par image, afin d'identifier des avions dans le ciel. On y retrouve également des chefs de projet qui, outre le pilotage le travail d'annotation, réalisent également des tâches de correction²¹². L'objectif de ce travail de vérification est double : il s'agit de vérifier *"toutes les images"* en les traitant *"une par une"* (Haina - Annotatrice - AssoAnnotation), en identifiant les erreurs et à améliorer le traitement des données, en repérant les difficultés des annotateurs pour les aider *"à s'améliorer"* (Haina - Annotatrice - AssoAnnotation). Ce travail de correction est également dévolu à Lita, nouvellement nommée cheffe de projet au moment de l'enquête. Cette tâche nécessite une plus grande habileté, une meilleure compréhension des projets ainsi qu'une capacité à traiter rapidement les données en identifiant les éléments liés à la compréhension des consignes ou à leur incomplétude pour que le projet ne soit pas bloqué²¹³.

En dehors des chefs de projet, certains annotateurs simples deviennent aussi correcteurs s'ils atteignent un niveau suffisant de productivité en termes de qualité et de quantité d'annotations. Elle souligne d'ailleurs que sa présence sur le lieu de travail lui permet d'identifier les annotateurs les plus performants, ceux qui comprennent le mieux les consignes et qui font le moins d'erreurs possibles : *"moi je suis toujours avec l'équipe, je suis dans le même bureau (...) et je vois parfaitement quel est l'annotateur qui est bien à l'aise sur ce projet-là (...) qui arrive bien à annoter, rapidement et à avoir aussi une bonne qualité"* (Haina - Annotatrice - AssoAnnotation).

Le même mode de fonctionnement a été observé chez le sous-traitant D, une entreprise du secteur bien plus importante, qui compte trois ateliers à Madagascar pour un total de mille employés. Le processus qualité y est structuré de la même manière. Les annotateurs produisent des données au cours de la phase de test du projet. Puis, les chefs de projet identifient les annotateurs les plus compétents. Ces derniers deviennent contrôleurs pour le reste du projet. Ces contrôleurs doivent ensuite vérifier le travail effectué par leurs

212. *"je fais de la review, c'est à dire je vérifie tout le travail d'annotation que l'on fait et c'est vraiment ça le boulot qui est tellement difficile pour nous chef de projet reviewer"* (Haina - Annotatrice - AssoAnnotation)

213. *"le reviewer c'est spécial ... en fait c'est des corrections, pour moi. Mais le reviewer ici on travaille quand même les projets. Mais pour moi reviewer ça veut dire qu'on review les tâches qui ont déjà été faites (...). Mais ici ce n'est pas encore partagé comme ça, c'est juste à un moment donné"*(Lita - Annotatrice - Sous-traitant G)

collègues tout au long du projet ²¹⁴. Enfin, les chefs de projet réalisent une deuxième phase de contrôle ²¹⁵. Chez le sous-traitant D, comme dans d'autres entreprises du secteur, ce travail de contrôle sert à éviter les phases de "rework", au cas où le client n'est pas satisfait de l'annotation. Là encore, les chefs de projet cumulent une fonction RH, une fonction de communication avec les clients et une fonction de vérification des tâches. Leurs missions sont donc particulièrement variées ²¹⁶.

Dans cette deuxième sous-section, nous avons montré la variété des fonctions occupées par les annotateurs. Un premier élément mérite d'être souligné : les entreprises d'IA n'externalisent pas seulement une tâche, mais bien un processus de production dans son ensemble. Comme on l'a d'ailleurs constaté dans notre précédent chapitre, les entreprises d'IA sont conscientes de la pertinence d'avoir recours à ce type d'organisation plutôt qu'à des plateformes.

Pour que la production puisse fournir les volumes de données nécessaires au développement de l'IA tout en garantissant une qualité suffisante, il est nécessaire d'avoir accès à une organisation du travail qui permet un contrôle au quotidien des annotateurs. Dans la plupart des cas, les clients exigent entre 90 et 95% de qualité des détections, vérifiées soit automatiquement, soit par échantillonnage. La quantité dépend du projet. À titre indicatif, cela peut représenter entre 5 images, s'il s'agit de grandes images satellites avec plusieurs centaines d'éléments à identifier sur chaque image, et 500 images par jour, dans le cas de l'annotation de petites images simples avec seulement trois ou quatre objets à identifier. On constate donc que, même si les activités liées à la gestion des jeux de données reposent en partie sur des personnes dédiées dans les entreprises d'IA, l'organisation du travail prévoit aussi l'externalisation d'activités annexes à l'annotation simple, ce qui nécessite des travailleurs plus qualifiés.

La structure des projets est également déterminante dans l'organisation du travail. Ainsi, le fonctionnement par phases de test implique des travailleurs plus qualifiés pour prendre en compte les besoins des clients et les traduire du point de vue de l'organisation de la production. On y retrouve des chefs de projet et des personnes employées dans des départements recherche et développement qui sont également plus qualifiés

214. "ce sont les opérateurs, les chefs de projet aussi font des contrôles avant la livraison ... on a 2 étapes de contrôle. D'abord les annotateurs font les annotations, et sur ces annotations faites, nous, les chefs de projet, on se charge de regarder quel opérateur a le plus de qualité, et ces opérateurs-là, on les met contrôleur après. Donc du coup quand l'annotation est terminée, les annotateurs choisissent d'être contrôleurs, ils contrôlent une première fois les annotations qui ont été faites."(Clément - Responsable de plateau - Sous-traitant D)

215. "M : Et après cela, nous, les chefs de projet afin de livrer les annotations, ont fait une 2e vérification. Q :d'accord et ensuite le client en fait une ? M : oui, lui il regarde si les annotations sont satisfaisantes, alors le projet est validé. Et si ce n'est pas satisfaisant il nous envoie des reworks. "(Clément - Responsable de plateau - Sous-traitant D)

216. "une journée type, comme je l'ai déjà dit je pense, on commence avec l'organisation du travail pour les opérateurs, leur donner leurs tâches respectives « toi tu fais cette annotation-là, toi ça » ; vérifier les annotations des opérateurs ... pas en entier mais je fais des vérifications par échantillonnage ; communiquer au client si on a un doute ... par exemple si on atteint les 25% de la production, on doit aussi leur communiquer ça. Après ça, discuter avec les supérieurs hiérarchiques si nous avons un problème en interne par exemple. Et aussi faire le suivi de la production, le suivi des opérateurs." (Nivo - Chef de projet - Sous-traitant D)

que les annotateurs simples. Au quotidien, ces chefs de projets cumulent de nombreuses fonctions, souvent sur plusieurs projets. Les annotateurs "simples" réalisent également toute une variété de tâches et ce, sur plusieurs projets. Dans certaines entreprises, cette dimension est prise en compte par la mise en place de politiques de gestion des ressources humaines qui visent à favoriser la polyvalence des travailleurs.

Au-delà de la variété des projets, on retrouve également une diversité de postes liées directement à la production des annotations. Certains travailleurs "simples" deviennent ainsi parfois correcteurs et ont alors davantage de responsabilités ce qui nécessite plus d'exigence sur le plan de la maîtrise du projet d'annotation. Pour plusieurs projets, on a également vu se développer une spécialisation des annotateurs, à la fois de manière formelle et informelle. Certaines entreprises désignent ainsi des travailleurs comme étant des spécialistes de certaines classes du projet d'annotation. On pense notamment à l'équipe "fromage" du projet de détection automatique des plats. Cela se traduit également par des logiques formelles et informelles par lesquelles certains travailleurs devenant aussi "référents".

L'organisation du travail d'annotation révèle également la nécessité d'avoir des compétences communicationnelles, les travailleurs discutant régulièrement de l'annotation au quotidien. Ces derniers interagissent avec les clients à travers différents outils (par exemple : des messageries ou des fichiers questions). Ainsi, les données de notre questionnaire montrent que 54% des annotateurs demandent des précisions aux clients. Par ailleurs, en cas de doutes sur une annotation, les travailleurs demandent également à leur manager ou à d'autres annotateurs, que ce soit via des messageries internes ou bien directement à l'oral. Ici, l'analyse de séquences de résolution des problèmes liés à l'annotation montre une évolution entre les doutes qui apparaissent au tout début du contrat de travail d'annotateur et les derniers doutes en date au moment des entretiens. On constate qu'au fil du temps les annotateurs continuent à collaborer. On pourrait parler de "co-production" de l'annotation en suivant le travail de [Velkovska et Zouinar \(2020\)](#) qui analyse les usages des agents conversationnels et montre ce que *"que les [consommateurs] personnes font de plus de ce qui est prévu dans le script"*. De manière similaire, le travail humain réalisé à Madagascar constitue la part d'effort essentiel de l'automatisation.

Cette dimension collective constitue la seule manière de résoudre les problèmes qui restent stables tout au long des projets : 52% des séquences pour le premier doute impliquent des collègues contre 47,9% pour le dernier doute. En revanche, au fil du temps, les travailleurs font moins appel à leur manager (74,7% contre 53,4%), ils vont plus systématiquement consulter d'autres ressources en ligne d'eux-mêmes (33,3% contre 52%) et moins le guide d'annotation (46,7% et 21,9%). Ces résultats montrent une montée en compétences des agents, qui se détachent de l'aide managériale (chef de projet, guides d'annotations). Au fil des projets, ces derniers utilisent des ressources plus personnelles (des recherches internet plutôt que le guide d'annotation) et la communication entre annotateurs pour résoudre les problèmes ce qui témoigne d'une consolidation de leurs compétences professionnelles. Les compétences communicationnelles sont donc déterminantes dans

le processus de production et témoignent de postes de travail plus qualifiés qu'il n'y paraît et s'expliquent par la proximité entre les données et les modèles ainsi que par une organisation plus *intégrée* du travail d'annotation. Les chefs de projets concentrent une partie de ce travail communicationnel : discussion avec les clients et les annotateurs. Ils s'occupent également de la gestion du projet au quotidien à travers l'interprétation des consignes des clients pour former les annotateurs, remontées des problèmes aux clients, conseils aux annotateurs, vérification de la qualité.

Le travail des données comporte donc des tâches non routinières liées aux données. Elles relèvent parfois du *"sens commun"* et, dans d'autres cas, d'une spécialisation des annotateurs (*compréhension du problème à automatiser*). Toutefois, c'est l'organisation de la production en différentes strates de contrôle de la qualité qui impliquent le plus de compétences non-routinières. En effet, il s'agit alors de communiquer sur des problèmes (*compétences relationnelles*) concernant par exemple des données non prévues dans les guides (*interprétation des données*). On a également constaté que certaines entreprises développent d'ailleurs des mécanismes *"d'élargissement du travail"* pour former leurs annotateurs à d'autres techniques d'annotations, ou tout simplement pour tenir compte des pics de production. L'ensemble de ces éléments nuance l'idée d'un travail déqualifié. Il reste à voir comment est mis en place le processus de montée en compétences au sein des entreprises d'annotations au travers des actions de formation.

7.1.3 Le travail prescrit de l'annotation de données : la formation des travailleurs

L'autre enjeu de l'organisation du travail des données porte sur la transmission des consignes d'annotation auprès des travailleurs. Comme on a pu le constater dans notre partie précédente, ce travail de formalisation des tâches nécessite la codification d'un guide d'annotation. Il est retravaillé, y compris avec les travailleurs des données, notamment durant la phase de test. Ce guide est ensuite pris en main par des chefs de projets ou des annotateurs experts pour être transformé en documents et séquences de formation. Or, si le guide d'annotation permet de codifier les règles de l'annotation, la formation des annotateurs ne s'arrêtent pas uniquement à cet aspect du travail, car ces derniers sont également formés aux outils d'annotations, de communication et de gestion de projets. Avant de rentrer plus précisément dans les différentes phases de formation, revenons d'abord sur l'organisation de ce processus d'apprentissage.

Cette organisation diffère selon les entreprises. Ainsi, dans les plus grandes et les plus formalisées, un département formation est dédié à la réalisation des supports de formation, en collaboration avec le département de test et les responsables des projets. Dans des entreprises plus petites ou informelles, ce sont plutôt les *"team leaders"* qui vont occuper cette fonction, parfois en collaboration avec un formateur. Ils récupèrent les consignes et réalisent des formations, comme l'explique Yanni, chef de projet (sous-traitant A - Projet caisse de cantine) : (...) *Ce sont les team leaders qui vont former les annotateurs(...). Il y a des consignes ou des fichiers PDF envoyés par le client (...). Le team leader va dire ce que l'on fait [ici] : annoter des aliments pour*

les restaurants (...). [Il] explique qu'il y a deux ou trois types de traitement, annotations et reviews (contrôles), et qu'après on classifie les images, puis qu'une (autre) équipe finalise le travail avant de l'envoyer vers le client. (...)" (Yanni - Chef de projet - Sous-traitant A)

Outre la transmission des consignes, l'enjeu des formations est de sensibiliser les annotateurs à la posture attendue par les clients. Comme l'explique le responsable de l'équipe d'annotation d'une start-up française, un bon labelliseur, *"C'est quelqu'un qui est consciencieux, qui ne fait pas deux choses en même temps, et qui est soucieux de préparer un bon jeu de données pour obtenir de bons résultats."* (Clément - Chef de projet data - Klik). Au sein de cette entreprise qui ne travaille qu'avec des annotateurs français, la finalité du travail des données y est expliquée : *"on essaie aussi de les sensibiliser là-dessus. Et ne pas (...) aller vite, sans se soucier des répercussions que ça peut avoir avec les clients."* (Clément - Chef de projet data - Klik). Il s'agit également de faire comprendre aux annotateurs ce qu'il est attendu d'eux en termes d'arbitrage entre le niveau de qualité et la quantité de données annotées, et de leur permettre d'appréhender les conséquences de leurs erreurs ²¹⁷.

Selon Lin, qui dirige les activités d'annotations chez Automatik, les annotateurs ne sont *"se familiarisent avec le métier"*, même s'ils ne sont pas formés. Elle précise toutefois que la formation couvre plusieurs thématiques que nous avons identifiées dans l'ensemble des entreprises d'annotations. D'abord, les gestes de l'annotation et de la classification, c'est-à-dire : *"les règles de base de l'annotation. Par exemple, l'annotation de tous les éléments sur l'image, précision des boxes, vérification des catégories"* (Lin - Datasets manager - Automatik). La maîtrise des outils d'annotations et de leurs fonctionnements est également essentielle. Au-delà des outils de découpage et de classification, les annotateurs sont enfin censés maîtriser les outils de communication (messagerie) et de gestion de projet (tableur).

Le dernier élément de la formation comporte les éléments relatifs aux consignes du projet et sont expliqués par les chefs de projet *"(...) avec un document de formation explicite et des consignes précises"* (Lin - Datasets manager - Automatik). Comme on l'a observé précédemment, un test peut être organisé, afin d'évaluer les connaissances des annotateurs, d'identifier ceux qui sont le plus à l'aise, ou encore de trouver des difficultés communes à plusieurs annotateurs. Ces problèmes de compréhension illustrent des problèmes de définition de certains concepts du guide d'annotation ou des catégories trop ambiguës. La formation sert également à formaliser des pratiques d'annotations, en complément des gestes de base qui sont attendus des travailleurs de l'annotation. : *"il faut être minutieux, avec des capacités d'adaptation au fil et de vision des détails, patient, curieux. Très important, surtout pour [le projet caisse de cantine], il faut vraiment avoir un intérêt pour connaître la nourriture française et en étant curieux, (...) savoir chercher les infos sur Internet et creuser, et proposer."* (Lin - Datasets manager - Automatik).

217. *"(...) il faut que le labelliseur puisse quand même aller vite puisque le volume est aussi important que la qualité des objets détournés."* (Clément - Chef de projet data - Klik)

Pour faciliter la formation des travailleurs, certaines entreprises mettent en place des méthodes d'apprentissages par tutorat. Certains annotateurs se voient attribuer un tuteur censé lui expliquer le travail durant parfois plusieurs semaines. Au sein du sous-traitant G, Félix devait par exemple observer ses collègues pour perfectionner ses capacités d'annotations, à son arrivée dans l'entreprise ²¹⁸. Ces logiques d'apprentissage sont collectives et durent parfois sur l'ensemble d'un projet. Cela se traduit par des formations et des réunions pendant lesquelles les travailleurs échangent sur les consignes, les données et sur les bonnes pratiques d'annotations : *"il y avait beaucoup de formations (...). Avant chaque projet, (...) on devait faire un briefing, et puis on mettait notre connaissance en équilibre, pour ne pas fausser les données. (...) On se donne [aussi] des conseils entre nous, entre collègues pour mieux faire notre tâche (...)."*(Hejo - Annotatrice - Sous-traitant A).

Les travailleurs sont aussi spécifiquement formés à la réalisation des projets d'annotations. Cette étape démarre par la phase de test qui participe également à la formation des annotateurs ainsi qu'à la constitution des supports de formation. Cette période associe généralement des annotateurs plus expérimentés, un chef de projet et le client et sert à identifier les problèmes du guide d'annotation selon les annotateurs : *"Oui, on a (...) aussi fait un test, sur les catégories, par exemple sur les catégories crustacés, ou pâtes, ou riz, pour mieux connaître nos difficultés. Par exemple, test 200 images par catégories. "* (Naivo - Annotateur - Sous-traitant G). Une fois ce test réalisé, les clients *"envoient le résultat du test"* en posant des questions sur les raisons de certaines erreurs. Dans le cas d'un autre projet, Naivo raconte que ce test se base sur le guide d'annotation donné par les clients : *"Ils m'ont donné le test et des instructions sur des fichiers PDF. Et ils ont regardé le pourcentage de qualité pour savoir s'ils allaient nous donner le projet ou non. Du coup, ils ne m'ont pas appris directement, mais ils m'ont envoyé le process de la formation pour faire le traitement."* (Naivo - Annotateur - Sous-traitant G). Cette phase sert donc également à tester les annotateurs et le guide d'annotation. Du côté des entreprises d'annotations, cela peut également être utilisé pour sélectionner les annotateurs les plus compétents pour réaliser le projet ²¹⁹.

Suite à cette première phase, il s'agit d'établir des supports de formation à destination des annotateurs en adaptant les guides d'annotations. Comme l'explique Jean, alors ancien conseiller d'une dirigeante d'entreprise d'annotation, il est nécessaire de codifier des savoirs de manière à les rendre compréhensibles par les travailleurs : *"lorsqu'il y a des nouveaux projets et que le client n'a qu'un manuel. On va essayer de transformer ce manuel en savoir puis en savoir transférable pour que ça puisse être assimilé par les agents qui vont exécuter les tâches (...) On ne peut pas dire simplement « Lis ce manuel et exécute ». On va essayer de faciliter la vie*

218. *"F : après la formation, le chef a dit que je dois regarder les personnes qui travaillent et je dois suivre les méthodes. Parce que chacun a sa méthode, et je dois trouver la méthode pour faciliter mon annotation."*(Félix - Annotateur - Sous-traitant G)

219. *"les chefs de projet on se charge de contrôler la qualité sur la plateforme d'annotation, et on voit directement « tel opérateur [a bien réussi] ». (...) Et c'est là que l'on sélectionne des annotateurs qualifiés pour être des contrôleurs"*(Martin - Sous-traitant D - Chef de projet)

du collaborateur qui va exécuter la tâche. C'est un défi quotidien. ” (Jean - Conseiller RH - Sous-traitant A). Cette codification passe alors par un travail collectif réalisé en association avec des responsables de projets et en discutant avec les clients ²²⁰. Les formations sur les projets sont ensuite réalisées par des responsables formations. Elles dépendent du degré de structuration et de complexité des projets et peuvent durer entre quelques heures et un ou deux mois. En plus de ces périodes consacrées à la formation, les entretiens font ressortir une formation par l'apprentissage tout au long des projets.

Les périodes de formation restent cependant relativement courtes. Le rôle du formateur consiste à présenter les cas de figure potentiels. Ainsi, sur un projet de détection des vols dans les supermarchés, Michael (responsable - sous-traitant A) explique que : *”La formation a commencé par nous expliquer les différentes sortes de paniers/sacs et les différentes sortes de gestes que nous devons connaître. On regarde si la personne fait des mouvements bizarres, si la personne est douteuse, et là on doit avertir et cliquer sur une labellisation qui dit que c'est un geste bizarre. On nous a montré aussi les différentes sortes de sacs, de produits et faire la différence entre un client et une personne qui travaille dans la grande surface.”* (Michael - Responsable - Sous-traitant A). Après un test, les annotateurs suffisamment qualifiés rentrent en production, les autres ont accès à des formations complémentaires ²²¹. Dans les projets les plus compliqués, la formation est beaucoup plus conséquente et surtout, elle continue à se perfectionner au fil des projets et en fonction de ses évolutions. C'est le cas sur un projet médical, particulièrement difficile, qui est expliqué par Cédric en charge du jeu de données de la start-up.

Encadré n°51 : La formation sur un projet médical

”[Sur la formation]On a vraiment passé beaucoup de temps avec la première [annotatrice], qui a ensuite elle-même passé du temps avec les nouveaux, et on continue de passer du temps aujourd'hui avec eux. (...) On faisait de la relecture en s'organisant en consensus interne où on revoit des signaux qu'eux avaient revus pour leur faire du feedback (...) de temps en temps on va [aussi] annoter avec eux, ça permet de répondre à leurs questions, de vérifier qu'ils utilisent bien la plateforme, et pour l'instant ils ne sont que quatre donc ça va, mais on imagine qu'à l'avenir ça permettra de s'assurer qu'il n'y ait pas de gens qui n'annotent pas bien ou qui ne connaissent pas bien les outils. ”

Cédric - Lead data Team - AlMedic

220. *”Je lis et relis d'abord le manuel, je pose des questions au client, je fais des calls pour savoir comment on va installer les outils, comment on va créer des comptes pour avoir accès aux plateformes, et je prends note. Après ça, j'évalue le niveau des collaborateurs pour effectuer le travail, et je conçois la formation en fonction de ça. Je divise les étapes nécessaires et l'objectif de l'activité.”*(Jean - Conseiller RH - Sous-traitant A)

221. *”Oui, il nous donne des formations et après on a un test avec de nombreuses vidéos. Si on réussit à atteindre 80% [de bonne détection], on est apte à prendre le poste et à travailler. Mais si on rate le test, [le client] nous fait un appel vidéo et nous pose des questions.”* (Michael - Responsable - Sous-traitant A)

Ce qui va finalement nécessiter le plus de temps concerne davantage l'acquisition de savoirs et de savoir-faire qui dépassent le cadre de la compréhension de la finalité du projet (ce qu'il faut faire) : compétences liées à l'utilisation des outils, aux techniques d'annotation ainsi qu'à l'organisation du travail d'annotation (consulter les consignes ou communiquer avec les clients, les chefs de projet et entre annotateurs). Dans de rares cas, la formation des annotateurs intègre des éléments de compréhension sur l'intelligence artificielle.

Les formations dédiées aux outils d'annotations sont rendues nécessaires par l'existence de plusieurs outils disponibles sur le marché et utilisés par les entreprises d'IA. Cela explique également que certains annotateurs ayant déjà un passé professionnel dans le même type d'emploi aient tout de même besoin de formation complémentaire sur les outils. Mino (sous-traitant G) raconte ainsi qu'étant donné qu'il a *"déjà effectué des tâches d'annotations"*, son chef de projet n'a *"pas vraiment expliqué beaucoup de choses (sur l'annotation)"*. Sa formation s'est alors principalement concentrée sur l'outil utilisé au sein du projet d'annotation sur lequel il a été affecté après son recrutement²²². D'autres annotateurs découvrent entièrement le principe du travail d'annotation et sont généralement mis en difficulté par le maniement des outils et notamment par le détournement d'éléments sur des images. C'est le cas de Naivo (sous-traitant G), qui raconte toutefois qu'avec l'acquisition de compétences lors de la formation et au fil du projet, il a pu se sentir plus à l'aise sur le projet : *"(...) c'est toujours un peu difficile au début, mais après on est habitués. Et ça devient un peu facile."* Ces annotateurs soulignent d'ailleurs l'importance de la formation réalisée sur l'outil qui constitue la difficulté technique de l'apprentissage du travail d'annotation²²³. Étant source de difficulté, cet aspect est d'autant plus complexe que les outils et les techniques d'annotations changent selon les projets. Ando (sous-traitant A), explique d'ailleurs qu'en cas de nouvelles consignes d'annotations, il *"essaye de faire l'annotation plus lentement"* afin de perfectionner ses techniques de travail. Cette phase lui permet aussi de développer ses propres *"tours de main"* en cherchant *"les techniques et les raccourcis qui peuvent rendre la tâche plus facile, et accélérer la tâche"* (Ando - Annotateur - Sous-traitant A). Dans son cas, il cite l'exemple d'un outil sur lequel il a trouvé un *"un raccourci clavier"* lui permettant d'*"ajouter une box"* plus rapidement.

Les annotateurs doivent aussi maîtriser le reste des compétences techniques et relationnelles nécessaires à la bonne gestion des projets. Or, les outils informatiques font partie de ces compétences techniques qui ne sont pas nécessairement connues de tous les annotateurs, bien qu'il s'agisse d'une compétence demandée lors des entretiens. C'est le cas au sein d'une filiale du sous-traitant G qui est située à Fianarantsoa²²⁴ et

222. *"Il m'a appris à utiliser l'outil qu'on utilise ici. Et après la deuxième semaine, on m'a fait faire des tests d'annotation sur l'outil. C'est comme utiliser photoshop en fait. Il y a des images et on faisait des bounding box sur les trucs qu'on devait annoter. C'était des images de poteaux, poteaux électriques. Il y avait des fils, et des pinces, et on devait les différencier."* (Mino - Annotateur - Sous-traitant G)

223. *"(...)j'étais tout nouveau et je ne connaissais rien du process, je ne connaissais rien de leur interface (...) En général, on encadre juste des photos, mais la formation était vraiment basée sur l'outil, parce que l'outil est un peu difficile."* (Ihary - Annotateur - Sous-traitant A)

224. Il s'agit d'une association très reconnue qui propose des tâches à des travailleurs éloignés de l'emploi : tri de vêtements, fabrication de voitures et donc annotation de données

qui intègre des annotateurs éloignés de l'emploi et n'ayant pas le même niveau de formation, notamment en français et en informatique ²²⁵. Lors d'un entretien avec Haina, annotatrice au sein de cette filiale, son chef de projet intervient pour expliquer ses difficultés à gérer les projets d'annotation, notamment en ce qui concerne la formation : *"Au départ c'est un peu terrible, parce qu'on a prévu avec Pierre de faire une formation de 2 semaines(...), il nous a fallu un mois et plus pour les former (...) parce que nous on a pensé à la base que les annotateurs ont un peu de connaissances avec la technologie, internet, utiliser les outils en ligne ... parce que nous on utilise des outils en ligne, il y a le drive, il y a le slack, il y a tout ça. Moi j'ai pensé que c'est un peu facile de faire la formation, même avec les meet qu'on fait, slack, google meet, ou. Donc tout ça ... j'ai un peu sous-estimé (le temps de formation)"* (Haina - Annotatrice - AssoAnnotation)

Dans ces délais très courts prévus initialement avec le client, l'association devait donc former les annotateurs. Face à une population éloignée du numérique, ce temps a en fait tout juste permis de former les annotateurs à l'utilisation d'un "ordinateur", à "la familiarisation avec les technologies" et difficilement à la plateforme d'annotation. Or, le chef de projet de la filiale considérait que la formation à l'informatique était trop importante pour s'en passer. Par la suite, les phases entre les projets ont alors permis d'organiser de nouvelles formations sur les outils de gestion de la production : *"ça nous a permis de faire des petites formations pour assimiler les outils que l'on utilise"* (Haina - Annotatrice - AssoAnnotation). Ces formations servaient à la fois à expliquer l'utilisation des outils et à transmettre aux annotateurs des éléments de compréhension supplémentaires pour maîtriser les outils. Ces logiciels sont aussi variés que Slack qui sert comme messagerie interne, les tableurs sur Google Drive qui servent à poser des questions aux clients et les outils de visio-conférence. Ils sont pourtant tous essentiels à la production des annotations, ce qui rend ces formations et ces compétences techniques essentielles ²²⁶

Les compétences relationnelles et techniques qui permettent de communiquer, notamment avec les clients, sont régulièrement revenues dans les entretiens comme étant des compétences importantes. Nivo, chef de projet chez le sous-traitant D, explique par ailleurs qu'il forme les annotateurs *"sur la communication avec les clients"*. La particularité de cette entreprise est d'avoir mis en place un processus de production qui laisse peu de place à l'autonomie des travailleurs en cas de doute ²²⁷. En effet, si un annotateur "a un doute" il *"doit directement envoyer un mail au client pour l'informer"*. Outre la montée en compétences, ce travail d'annotation comporte d'autres aspects qui font également l'objet de formations. Maîtrise des outils de communica-

225. Ce qui montre, en creux, l'importance d'avoir accès à une main-d'oeuvre ayant un bon niveau de formation, ce qui est le cas des annotateurs de la capitale malgache.

226. *"Par exemple avant ils utilisent les fichiers questions, les Sheets en ligne. Donc j'ai juste dit comment s'en servir, comment utiliser ça. Mais si un jour je leur demandais de créer un fichier question, ils ne peuvent pas créer cela. Donc on a un peu essayé, j'ai un peu essayé de dire que ... j'ai fait un peu de formation avec eux, parce qu'il y a un peu de temps. Pareil avec ça aussi sur le meet -lorsque l'on fait un visio comme ça, les annotateurs, je prépare ça et les annotateurs assistent. Mais si je dis « on va faire un visio », ils ne savent pas très bien qu'est-ce qu'on fait pour avoir ça ... Donc tous ces petits outils qui sont indispensables dans le travail, ça, si on a un peu de temps, j'essaie aussi de faire un peu ces petites formations"* (Haina - Annotatrice - AssoAnnotation)

227. Ce point sera creusé lors de notre dernier chapitre

tion et de remontée d'information font partie des compétences attendues, ce qui se reflète également dans notre questionnaire, puisque 76% des travailleurs déclarent avoir eu une formation professionnelle (N=296). Parmi eux, ils sont 58% à avoir reçu une formation professionnelle aux outils informatiques (N=215). Certains travailleurs indiquent d'ailleurs avoir réalisé des formations professionnelles après quelques années passées à l'université, afin de maîtriser les outils informatiques pour intégrer le secteur de l'externalisation de services.

Au travers de l'analyse de la formation à l'annotation, se pose la question de la professionnalisation des annotateurs. Nos résultats mettent ainsi en évidence la continuité des pratiques de travail dans un même secteur. L'analyse des pratiques de formation montre que le développement des compétences procède d'une *"intention organisationnelle"* qui passe par l'acquisition de *"bonnes compétences"* au travers du recrutement et de la formation. On peut néanmoins s'interroger sur la réalité de cette *"professionnalisation"* tant la recherche et certains discours des acteurs de la chaîne mettent l'accent sur un travail répétitif ne nécessitant pas de qualités particulières. A ce stade de notre développement, nos résultats suggèrent au contraire une *"solidification des pratiques professionnelles"* qui résulte d'une volonté de l'organisation. Cette activité apparaît également marquée par la codification de pratiques à travers certains procédés ou les formations. Le point d'achoppement principal de cette analyse des pratiques d'annotations relève de l'articulation entre les individus et les organisations. Ainsi, nous constatons que cette *"solidification"* est liée à des impératifs de production qui s'imposent aux individus. Il est donc nécessaire de se questionner sur les activités réelles des annotateurs. S'agit-il d'une forme de *"développement professionnel"* au sens de Dubar (1991), comme construction collective de savoir-faire susceptibles d'améliorer la production? Surtout, dans quelles mesures cette dimension est reconnue ou non par les autres acteurs de la chaîne?

* * *

En comparaison avec littérature consacrée au travail des données réalisé via des plateformes numériques de travail, notre travail montre un aspect bien différent : les entreprises n'externalisent pas des tâches isolées, mais un processus de production. Ce mode de fonctionnement implique un certain nombre de postes qui permettent de gérer au quotidien le travail des données pour atteindre les impératifs de quantité et de qualité. Les attendus des donneurs d'ordres concernent également la gestion de projet et la capacité à communiquer autour du projet d'annotation. On voit émerger des postes plus qualifiés (équipes managériales à Madagascar) et une requalification du travail de saisie de données pour les annotateurs simples qui se traduisent par une forme de responsabilité sur la production. Les données de notre questionnaire sur les travailleurs confirment ces éléments :

- 93% des travailleurs pensent que leurs erreurs ont un impact sur la qualité du travail
- 76,4% des travailleurs pensent que leurs erreurs ont des impacts sur les coûts des projets.

Sur ce point, les résultats de cette première section sont dans la continuité directe des éléments présentés

dans notre chapitre précédent relatif à l'intégration des travailleurs des données dans le processus de production de l'IA. C'est aussi à travers cette intégration que se traduisent des enjeux de communication et de coordination. On retrouve ici des éléments propres à la critique de la déqualification du travail industriel et évoqué par Zarifian (1983) à propos de l'existence de "savoirs collectifs". Ces savoirs se construisent et se transmettent entre les travailleurs de part via leurs réseaux de discussion qui sont autant d'espaces d'apprentissages, permettant l'émergence d'une communauté de pratiques. A ce stade de notre développement, nous pensons que ces résultats indiquent une "solidification des pratiques professionnelles" qui est le fruit d'un choix des entreprises d'IA et de sous-traitance lié à la qualité et qui illustre le corpus de qualifications nécessaire à cette activité.

Signe que cette réalité n'est pas encore reconnue, les entrepreneurs de l'annotation de données rencontrés en entretien expliquent pourtant ne pas avoir besoin de travailleurs qualifiés. Quel sens donner à ces éléments, en apparence contradictoires? D'abord, nos constats limitent largement cette vision du travail des données en matière de formation initiale : 75% des travailleurs de notre échantillon sont allés à l'université. Par ailleurs, si les dirigeants d'entreprises d'annotation considèrent qu'il n'est pas nécessaire d'être très bien formé, ils exigent toutefois en général le niveau Bac pour les annotateurs simples, un niveau Bac +2 pour les équipes managériales et un niveau de français suffisant à l'écrit et à l'oral. Cela suggère finalement que les qualifications sont en partie invisibilisées, mais bien présentes au sein des entreprises d'annotations.

De notre point de vue, ces résultats illustrent une des limites de l'hypothèse de la déqualification. Ainsi, on peut y voir les conséquences de l'élévation générale des compétences (en langue, en informatique) évoquées par Gardey (2008) à propos du travail de saisie de données. Par ailleurs, étant donné que cette relation employeur-employé s'inscrit dans une dynamique relationnelle entre l'ancien pays colonisateur (la France) son ancienne colonie (Madagascar)²²⁸, cela pourrait également expliquer le processus de minimisation des qualifications, comme nous le verrons dans la partie quatre.

Dans la perspective de la sociologie de l'innovation, Renou (2006) parle quant à lui "d'espace de discussion". Avec ce concept, il désigne : "un espace de construction de procédures collectives d'interprétation et de résolution de problème". Ce résultat est en partie confirmé par notre questionnaire : 95,4% des travailleurs qui considèrent travailler en équipe et non de manière solitaire. Cette dimension du travail d'annotation a une implication directe sur l'IA : pour nous, c'est au sein de cet espace de conventionnement que se fabrique l'intelligence artificielle. Ce concept fait aussi référence aux enjeux étudiés en STS autour de la "localisation" des concepts qui s'appliquent à la classification (Bowker & Star, 2000). Dans notre prochaine section, nous verrons que ce travail de localisation nécessite également des qualifications. La suite de notre chapitre propose prolonge l'analyse d'une caractéristique importante de l'industrie de l'IA : cette industrie s'appuie sur la coordination de plusieurs acteurs qui interagissent entre eux pour développer le modèle. Ce mode de

228. Cet aspect de la chaîne de valeur est développé dans notre prochain chapitre

fonctionnement se rapproche des industries de process avec deux limites : la production de l'IA est continue du fait de la confrontation avec la vérité terrain, mais elle n'a pas besoin que les unités de production se situent sur le même lieu de travail.

7.2 Une solidification des pratiques professionnelles qui passe par la formalisation des qualifications requises pour optimiser le fonctionnement des modèles

L'industrie de l'externalisation à Madagascar s'est progressivement dirigée vers les tâches de production de données pour l'intelligence artificielle. Dans les années 2010, de nombreux travailleurs occupaient encore des postes dans les centres d'appels. Progressivement, certaines entreprises se sont lancées sur le marché de la catégorisation de données, en proposant des services comme la saisie de données pour constituer des fiches produits à destination des sites de commerce en ligne. En se spécialisant dans l'industrie de l'annotation des données, ces entreprises ont pu appliquer leurs connaissances à d'autres activités à l'intelligence artificielle : *"Pourquoi on fait de l'annotation de données ? Parce qu'à la base on faisait de la saisie de données. En réalité, l'annotation de données c'est de la catégorisation de données. Vous avez une donnée et vous dites, ça, c'est ça. Donc, c'est logique que l'on fasse ça. C'est une continuité dans ce que l'on faisait déjà à l'époque."* (Louis - CEO - Sous-traitant F).

C'est ce qu'explique la dirigeante d'une autre entreprise d'annotation qui a d'abord démarré son activité avec ces tâches de saisie de données et de centre d'appels avant de se spécialiser dans l'annotation : *"Ce qui est aussi intéressant, c'est qu'il y a beaucoup de boîtes qui font du BPO qui transforment ça maintenant en annotation de data. C'est vrai, les stages [étapes] sont les mêmes."* (Dinaso - CEO - Sous-traitant E) D'après elle, la différence principale réside dans la finalité des projets d'IA par rapport à ceux des entreprises d'externalisations de services (BPO) : *"C'est juste que l'output [n'est pas le même], (...) Pour un BPO, ça va dans le service Achats, le service marketing, ce n'est pas de la R&D."* (Dinaso - CEO - Sous-traitant E). Toutefois, plus les projets d'IA concernent un déploiement en production, plus ces projets vont être gérés avec les départements métiers, outre la R&D.

Dans cette seconde sous-section, nous analysons les différentes pratiques d'annotations de données. Nous montrons de quelle manière les travailleurs des données contribuent à la production, au déploiement et à la maintenance des modèles d'intelligence artificielle. Dans un continuum entre les activités les plus manuelles et les activités les plus cognitives, le "découpage" constitue une première étape, celle qui comporte le plus de gestes physiques. Ce travail cognitif se traduit par l'exercice d'un jugement et par la compréhension des catégories qui s'appliquent aux projets d'IA. Dans ces projets, nous verrons plus spécifiquement des logiques de spécialisation des travailleurs.

7.2.1 Découper, classer, saisir : le travail manuel et cognitif des travailleurs des données

Afin d'introduire cette partie, revenons très brièvement sur les enjeux liés à la conceptualisation du travail d'annotation. Longtemps, la littérature scientifique a traité cet objet sous l'angle d'un "micro-travail", *que tout le monde peut faire*, car il fait appel au sens commun et surtout peut être organisé *n'importe comment*. Le travail réalisé sur les plateformes numériques a largement encouragé cette conceptualisation, tout autant que l'utilisation du terme micro-travail. Ce dernier puise ses références à la fois dans le micro-crédit, en le détournant pour proposer plutôt des petites sommes contre des micro-tâches à des travailleurs de pays en voie de développement, ainsi que dans le travail "intersticiel", c'est-à-dire le travail réalisé sur des temps de pause, des temps de trajet, bref des "temps morts". Le travail de typologie réalisé par Tubaro, Casilli, et Coville (2020b) et qui prédomine aujourd'hui dans la littérature sur le sujet s'est focalisé sur la finalité des contributions des travailleurs des données. On trouve trop peu de travaux qui analysent les pratiques quotidiennes d'annotation. La réalisation de nombreux entretiens auprès de travailleurs ayant exercé sur beaucoup de projets différents et le fait d'avoir effectué une partie du terrain directement au sein des usines de production de données facilitent ce travail de reconstitution.

Le "travail du clic" ou du doigt peut aussi renvoyer à un "travail manuel", comme cela a pu être évoqué dans notre précédent chapitre par deux employés de MLDupli. Il n'est pas question ici d'opposer un travail prétendument purement manuel à un travail purement cognitif, mais plutôt de décomposer la pratique de l'annotation selon des étapes plutôt manuelles ou plutôt cognitives. Certains annotateurs, comme Fleur (AIEthic), soulignent ainsi le caractère répétitif du geste lié à la souris : « *Parce que dans l'annotation, on est un peu comme des robots, on clique, on clique. Ça ne pousse pas la créativité selon moi.* » (Fleur - Annotatrice - AIEthic). Joda, un travailleur du sous-traitant A compare son expérience dans un centre d'appels avec son travail d'annotateur : « *Par exemple, dans un call-center, pendant le travail, on met son casque, on active l'appel, et on parle tout le temps pour convaincre un client, durant les 8 heures.* » (Joda - Annotateur - Sous-traitant A). Dans son travail en tant qu'annotateur, il valorise l'utilisation de ses "mains pour cliquer avec la souris" en soulignant que le geste lui donne la possibilité de "discuter (...) de temps en temps." (Joda - Annotateur - Sous-traitant A). Ces extraits illustrent le premier niveau du travail d'annotation lié à la gestuelle et à l'utilisation de la souris. C'est également le cas de Viviane, travailleuse chez AIEthic et qui nous raconte un projet d'annotation de plaques d'immatriculation pour lequel toutes ses tâches ont été refusées par le client, l'empêchant d'être payée²²⁹. C'est alors qu'elle prend conscience du degré de précision nécessaire pour entourer des éléments, parfois relativement petits sur des images et donc de la nécessité d'utiliser, non pas le pavé tactile de son ordinateur, mais une vraie souris : « *en fait je n'annotais pas assez bien les box. Il fallait être assez précise sur les box et je n'utilisais pas ma souris. Parce que pour faire l'annotation de ces*

229. "c'était les annotations de plaques d'immatriculation je crois peut-être je n'avais pas bien compris le projet, et donc toutes les tâches que j'avais faites avaient pratiquement toutes été annulées, on a dû me retirer du projet. Et quand on vous annule la tâche, on vous retire encore l'argent, donc ça m'avait un peu déstabilisée ..." (Viviane - AIEthic - Annotatrice)

boxes, il faut impérativement avoir la souris (...) j'utilisais la souris de mon laptop, ce qui n'était pas évident» (Viviane - Annotatrice - AIEthic).

En plus de se voir parfois refuser les paiements des clients, il existe chez AIEthic une mesure de la qualité qui détermine si les travailleuses peuvent rester sur le projet ou non. C'est ce que Mafo a constaté en travaillant sur un projet de labellisation de skieurs. Elle se retrouve en effet enlevée du projet, car elle n'était pas assez précise dans le découpage des personnes sur les photos de pistes de ski : *« Il fallait labelliser les skieurs, et on nous a expliqué que c'était pour aider la machine à éviter les accidents. Il fallait dire si c'était un homme, une femme, un enfant. (...) c'était très difficile (...) j'ai été enlevée du projet, parce que mon taux de qualité n'était pas bon (...) à la longue que je me suis rendu compte qu'il te faut (...) te faut une souris laser quand tu fais la classification.»* (Mafo - AIEthic - Annotatrice). Derrière le geste anodin de bouger la souris, ce travail de découpage, qui correspond à la tâche la plus simple à faire pour les annotateurs requiert un degré d'habileté souvent nécessaire pour annoter correctement les données. Afin d'approfondir l'analyse des pratiques d'annotations, il est important de tenir compte de la variété des projets d'IA auxquels sont confrontés les annotateurs. Kajy, chef de projet chez Sous-traitant D raconte le projet qu'il dirige au moment de l'entretien : il s'agit de segmenter *”des personnes via des box 2D – aussi, avec ces boxes-là, on doit segmenter les parties du corps des personnes, c'est-à-dire encadrer la tête, le torse et tout le corps entier”* (Kajy - Chef de projet - Sous-traitant D). Sur ce projet, les annotateurs réalisent également *”l'acquisition des données”*, en plaçant des caméras au-dessus de leur atelier de production pour capter des images²³⁰. Reste ensuite à poser les *”box”* sur les bons endroits du corps sur des vidéos découpées en séquences, en tenant compte du mouvement des individus²³¹.

Le découpage est en fait un terme utilisé par plusieurs annotateurs pour désigner le travail d'encadrement sur des images ou sur du texte. Il s'agit du premier niveau de compétences de l'annotation : le travail de la souris. Dans les projets que nous présentons par la suite, l'objectif est d'identifier les difficultés posées par certaines tâches. Haino, chef de projet du sous-traitant F, nous détaille par exemple un projet dont l'objectif est d'encadrer des produits sur des rayons de supermarché. L'image comporte ainsi un rayon dans lequel sont rangés de nombreux produits. Pour lui, ce travail, *”est vraiment de l'annotation”*, dans le sens où il s'agit juste de placer des carrés autour des produits²³².

230. *”le client exige que les caméras doivent être placées au-dessus d'une porte, pour que l'on voie les gens qui passent (...)”* (Kajy - Chef de projet - Sous-traitant D)

231. *”F : (...) sur ce projet-là, la difficulté en fait c'est que la caméra détecte des vidéos de 30FPS, et sur l'outil d'annotation on doit les convertir en 5FPS, pour que l'on ait une annotation sur la vidéo de 30FPS, chaque 2 frames ouais ... et en fait la difficulté avec ce projet là c'est l'interpolation que les opérateurs font quand ils font l'annotation, par exemple quand une personne sort de la vision de la caméra, il faut déjà masquer le box avant qu'il soit complètement disparu, et en fait certains opérateurs ne respectent pas les consignes, et ils ne font pas ce truc-là du coup ... sur le 30FPS, les annotations ne sont pas correctes(...)”* (Kajy - Chef de projet - Sous-traitant D)

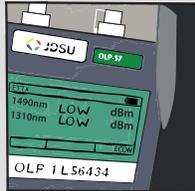
232. *”On a un client (...) et pour eux, on encadre des images. Il y a des images de rayons de supermarché (...)”* (Haino - Chef de projet - Sous-traitant F)

Lors d'un entretien, Mevo, ancien travailleur de l'entreprise C, nous explique les tâches d'annotations réalisées par ses collègues pour Amazon Mechanical Turk. Sur des images données par le client final, les annotateurs doivent annoter des éléments, comme sur les captchas que nous réalisons quotidiennement sur le Web : *"il y a des gars qui jouent avec des captchas toute la journée : « où sont les bateaux ? », tac tac tac ... Je sais pas quel est le but, mais il y a des gars qui font ça"* (Mevo - Responsable - Sous-traitant C). La référence aux captchas, qui constitue un autre moyen de collecter des données par le travail gratuit des internautes, révèle la facilité apparente de cette activité (*n'importe qui peut le faire*). D'autres projets nécessitent pourtant plus de précision pour réaliser le détourage. Ihary, annotateur chez le sous-traitant A, raconte un projet qu'il a réalisé pour une entreprise d'IA française qui vend une application de détection des problèmes de conformité des installations de fibre optique destinée aux techniciens-installateurs. Il doit encadrer différents éléments d'une image : la marque de l'entreprise qui a installé la fibre optique, les jarretières optiques qui permettent de connecter les câbles, et les valeurs des photomètres utilisés par les techniciens pour mesurer la puissance électrique des armoires (voir images ci-dessous). Pour ce faire, Ihary raconte qu'il doit regarder chaque détail de l'armoire de fibre optique ainsi que des photomètres, ce qu'il a trouvé difficile lors de son arrivée sur le projet : *"C'était ça le plus difficile, parce que ils étaient très stricts, qu'on recadre la photo bien juste. Parce que moi je n'étais pas habitué à ça, j'étais habitué à travailler sur des lettres, chiffres, mais ici, on travaille sur la souris, c'est la souris notre meilleure amie !"* (Ihary - Annotateur - Sous-traitant A)

7 réponses

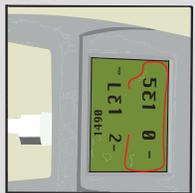
Travailleur sous-traitant A il y a 8 mois
@Lin – Datasets manager @Solution Architect Automatik concernant l'atténuation et la longueur d'onde, est ce que c'est correct que l'unité de l'atténuation est en **nm** ? car avant dans les autre projet l'unité pour l'atténuation est **dbm** et pour longueur d'ondes en **nm** (modifié)

Travailleur sous-traitant A il y a 8 mois
@Solution Architect Automatik Et pour les photomètre comme suit, il y a 2 valeur de nm. Est ce qu'on prend juste la première valeur ou les 2 svp?



il y a 8 mois Solution Architect Automatik
prenez les deux s'il vous plaît

Travailleur sous-traitant A il y a 8 mois
Merci . Et pour les valeur de cb (decibel) comme l'image ci dessous ?



il y a 8 mois Solution Architect Automatik
vous pouvez le mettre dans la colonne «puissance»

FIGURE 23 – Les règles d'annotations des armoires de fibre optique (Slack de Automatik)

Solution Architect Automatik 13 h 08
J'ai un commentaire @Team Leader sous-traitant A sur les annotations de demo-orange:

- dans le cas des jarretières: il faut annoter la case, et non la jarretière en entier !
- exemple dans l'image ci-dessous les jarretières ont été annotées, voici comment à aurait du être fait (screenshot CORRECTION)
- en particulier, si une case est masquée par une jarretière il faut quand même l'annoter (et on essaie de deviner si elle est avec un bouchon ou bien si il y a une autre jarretière)

Il faut donc repasser dessus pour les corriger

2 fichiers

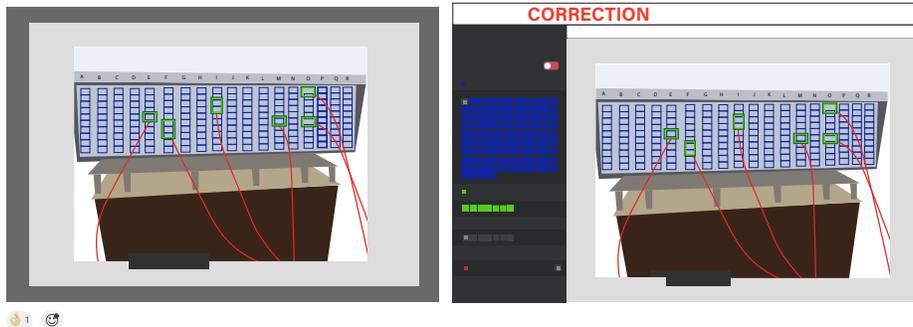


FIGURE 24 – Les règles d'annotations des armoires de fibre optique (Slack de Automatik)

Encadré n°52 : L'annotation d'une ville

D : C'était le projet le plus difficile jusqu'ici (...) ça consiste ... à annoter les véhicules mobiles et immobiles (...) les tâches sont tellement fatigantes (...) parce que ça consiste à vraiment regarder tous les détails qui s'affichaient à l'écran ... ça devait être parfait. Sur l'interface, (...) on reçoit une image satellite, sur l'image il y a plein de véhicules de parking ou d'une grande ville je dirais, ça consistait à prendre les véhicules un par un et compter (...) combien de véhicules il y a dans cette image, combien sont immobiles et combien sont mobiles (...). Donc il fallait mettre des bounding box

X : et ensuite dire « c'est mobile/ immobile/c'est un camion/c'est un piéton » ?

D : voilà (...) C'était la ville de Barcelone, et il y a des gros parkings à Barcelone. Il y avait aux environs 600 véhicules immobiles, et environ 200 mobiles

X : Ah ouais donc tu faisais combien d'images par jour ?

D : au début c'était simplement 2 comme ça ... à force d'être habitué à la tâche, j'ai commencé à acquérir beaucoup plus de vitesse, et après j'ai pu finir 5 images par jour, et c'était simplement en 2 mois que j'ai fini toutes les tâches du projet

Dafy - Annotateur - Sous-traitant G

Le type d'annotation est également susceptible d'entraîner des difficultés pour les annotateurs. Martin, chef de projet au sein de l'entreprise Sous-traitant D, nous explique le problème rencontré au cours du projet d'annotation d'ongles : *"sur (...) le projet qu'on a mis à mes débuts, sur les images de test « ah c'est facile, il faut juste à détourer », mais quand le projet est arrivé, dans un premier temps sur un lot de 1000 images, les 1000 images ont été toutes rejetées"* (Martin - Sous-traitant D - Chef de projet). Une première difficulté réside dans la différence entre les données utilisées durant la phase de test et les données de la phase de production. Les données de la phase de test étaient *"clean"* tandis que les ongles de la phase de production présentaient des irrégularités et notamment la présence de cuticules qu'il ne fallait pas annoter : *"sur l'image 3D il n'y avait pas de cuticule, parce que c'était tout clean, alors que sur les images réelles (...) il y a des cuticules qui dépassent, alors que les cuticules ne devraient pas être prises ..."* (Martin - Sous-traitant D - Chef de projet). Entre ces deux phases, un élément de consigne a été modifié par le client, occasionnant le rejet de toutes les images. Initialement, le travail consiste à entourer les ongles de manière plus grossière (avec un rectangle), mais le client a finalement précisé qu'il fallait entourer les ongles au *"pixel près"*, donc avec de nombreux points qui entourent les ongles (voir l'image ci-dessous)²³³. Les annotateurs se retrouvent donc à se questionner à propos des consignes initiales. Les évolutions des consignes et l'arrivée d'images qui ne sont pas prévues

233. *"Au début ils n'avaient pas précisé des pixels près sur la segmentation, donc sur la suite du projet il nous a demandé au pixel près."*(Martin - Sous-traitant D - Chef de projet)

dans les images de test perturbent alors la production ²³⁴.

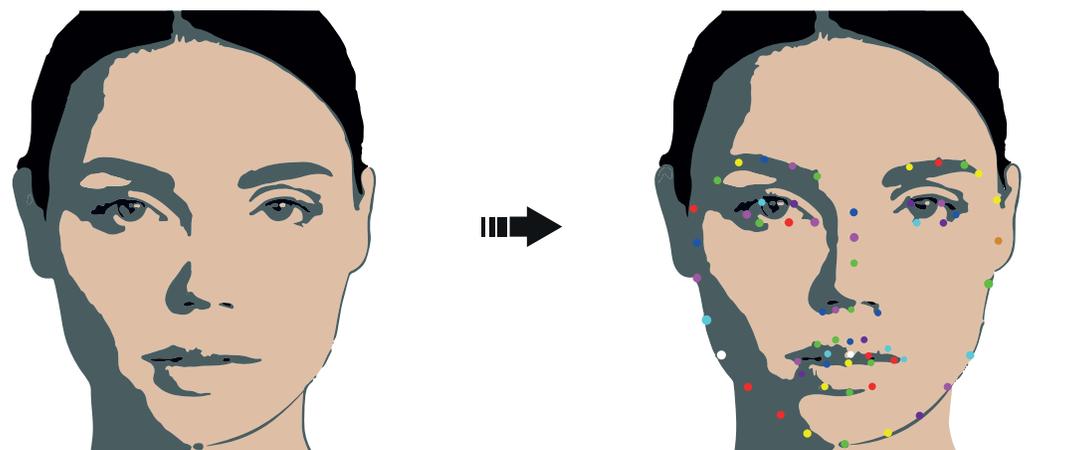


FIGURE 26 – Exemple d'annotation par point

On retrouve également cet aspect dans les projets d'IA qui sont déployés en production et pour lesquels de nouvelles images peuvent parfois être annotées par des travailleurs afin de perfectionner le modèle. Martin explique encore un projet de *"segmentation des pétales de fleurs de vanille"* censé permettre de détecter la période idéale de floraison pour leur collecte. Chaque annotation d'image comporte deux étapes, d'abord "l'encadrement" des pétales de fleurs, puis le fait de *"mettre des points sur chaque pétale"* (Martin - Sous-traitant D - Chef de projet). Se basant sur les consignes des clients, Martin a séparé l'annotation en deux groupes d'annotateurs, le groupe pour l'encadrement et le groupe pour les points comme *"le client"* l'a demandé. Toutefois, ces derniers constatent que ce système ne fonctionne pas. En effet, comme le raconte Martin : *"parfois les opérateurs n'ont pas le même niveau (...) un opérateur est plus doué que l'autre, et l'autre annote bien la segmentation, et l'autre fait les 10 points et il fait à sa manière et donc du coup les annotations ne se coïncident pas."* Associant deux techniques d'annotations, le projet est alors beaucoup plus complexe en termes d'organisation du travail. Il est également rendu difficile par la variété des images de fleurs de vanille et donc par les cas non prévus par le guide d'annotation ou par les images de test ²³⁵. Les annotateurs sont alors confrontés à de nombreux doutes et doivent être d'autant plus vigilants et communiquer entre eux. Le chef de projet se retrouve quant à lui à interagir de manière très régulière avec le client afin de mitiger ces ambiguïtés : *"c'était sur (la) semaine de production au début, qu'on a fait plus de fréquence (de réunion) à la demande du client. Mais (...) même en 3 mois de traitement, on avait toujours des doutes - du coup il y avait presque 2 ou 3 réunions par semaine avec le client pour lui demander « comment on fait pour ce cas-là ? comment on fait pour ce cas-là ? ». Parce qu'il y avait 25000 fleurs, il y a beaucoup de variétés de fleurs dans ces 25000 où tu n'as pas trouvé ça dans les tests."* (Martin - Sous-traitant D - Chef de projet).

234. Selon les projets, ces images sont extraites du Web (banque d'images) et comportent donc des exemples non réels

235. *"sur les fleurs de vanille, il y avait trop de cas que l'on n'a pas pu (annoter)"* (Martin - Sous-traitant D - Chef de projet)

Dans d'autres cas, les clients, conscients des évolutions du projet, changent les consignes pour y intégrer de nouvelles choses, ce qui est source de difficultés comme l'explique Yanni (sous-traitant A) : *"[ce que je] trouvais difficile, c'est qu'il y avait beaucoup de consignes qui changeaient sur une semaine ou deux (...) : le client dit que ceci doit être annoté comme ça, mais après, il dit qu'il faut le corriger en un autre tag."* (Yanni - Chef de projet - Sous-traitant A). Ces nouvelles consignes sont aussi liées au déploiement des projets en production qui apportent des nouveaux éléments à identifier. Dans le cas des plateaux repas, les smartphones ou les cartes de cantine que les clients déposent sur leur plateau n'avaient pas été anticipés durant la première phase d'annotation. Dans certains projets, ces évolutions de consignes sont liées à une complexification du modèle et donc des objets à identifier.

En lien avec la spécificité de certaines données à annoter, les vidéos entraînent également des difficultés et nécessitent plus d'habileté. C'est notamment le cas sur les projets liés à la voiture autonome pour lesquels les travailleurs doivent suivre des éléments en les annotant séquence par séquence, comme l'explique Joseph : *"en ce moment j'annote des voitures, vous voyez, si on prend l'image d'un embouteillage par exemple, il y a plein de voitures et on annote les voitures. Il faut faire des boxs, sur les voitures et après on suit la direction des voitures, (et) si la voiture avance il faut décaler petit à petit les boxs jusqu'à la fin."* (Joseph - Annotateur - Sous-traitant G) L'enjeu est ici d'arriver à suivre les éléments à annoter sur des séquences qui comportent de nombreux autres objets à ne pas prendre en compte.

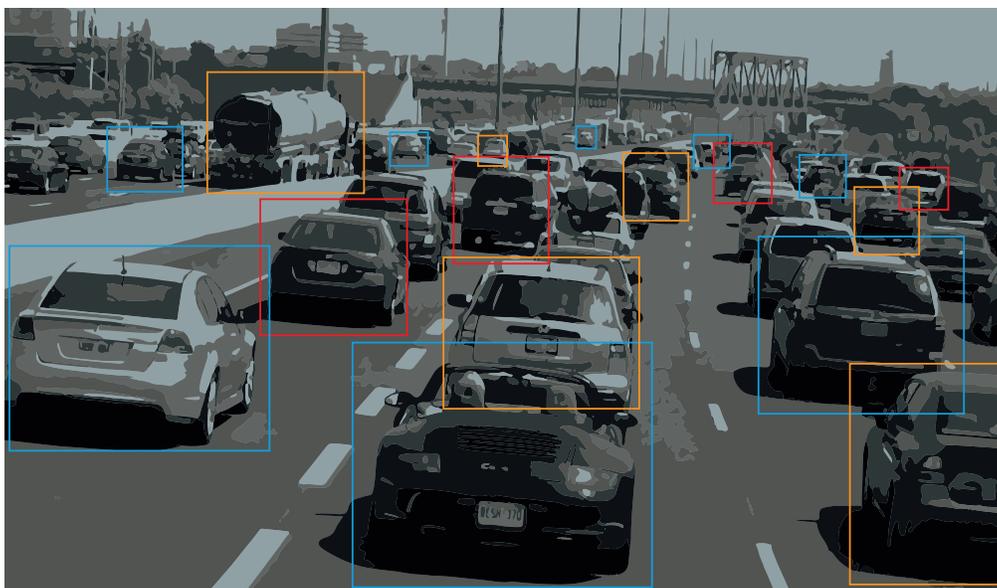


FIGURE 27 – Exemple d'annotation pour le développement d'une voiture autonome

Au sein de la même entreprise, d'autres annotateurs doivent également traiter des vidéos. Il s'agit ici d'entourer des éléments qui se déplacent dans le ciel. Comme pour les voitures en mouvement, l'interpolation

rend difficile la réalisation des encadrements : I : *”ce sont des annotations de séquences, il y a ce qu’on appelle l’interpolation des annotations, parce que l’on annote frame par frame l’image, on saute un peu l’annotation pour que l’interpolation soit respectée”* A cause du mouvement et du découpage de l’image, les objets sur l’image peuvent être déformés ou changer de couleur du fait de l’interpolation²³⁶. Par ailleurs, l’utilisation d’images infrarouges gênent également la visibilité des annotateurs²³⁷. La plupart des projets intègrent également un travail de *”classification”* de ces éléments. Par exemple, dans le projet StreetView, Holy, nous raconte cette autre phase du projet. Il s’agit alors de tracer les éléments, de les entourer, puis de les ranger dans des classes préalablement définies par le client : *”H : traitement d’images c’est le détournage. Il y a aussi le street, c’est comme Google Maps, mais ils ont tracé les rues, il faut annoter les poubelles, les voitures qui sont dans la rue, tous les poteaux, tout ça (...) il faut les tracer et après il faut marquer « ceci est une rue », « ceci est un trottoir », ...”* (Holy - Annotateur - Sous-traitant C). Notre précédente section nous a permis de constater la formalisation de ces classes au sein des entreprises d’IA, notamment à travers les guides d’annotation. L’image présente ci-dessous illustre également la manière dont sont renseignées les classes sur les outils d’annotations. Selon le projet et l’outil, ces classes sont directement présentes sur l’interface, à l’annotateur ensuite de ranger les éléments. Ce type de projet reste plutôt simple du fait des éléments à annoter. Il s’agit d’éléments présents sur des routes et qui sont relativement communs (un poteau, une route, un arbre).



FIGURE 28 – Exemple d’une annotation par zone de l’image

236. En traitement d’images, l’interpolation est le mécanisme par lequel les logiciels vidéos calculent un pixel en fonction des pixels voisins. Par conséquent, le découpage de la séquence et du mouvement peut changer la couleur des objets présents sur l’image, voire même ajouter des objets non présents sur les images précédentes

237. *”Parce que là c’est des images infrarouges, donc là, il y a des images visibles ... il y a 2 batchs, 2 sortes d’images, les images visibles et les images infrarouges. L’image visible c’est toujours en noir et blanc (...) ce n’est pas trop intelligible (...) je dirais que c’est très difficile, c’est vraiment très difficile”* (Simon - Chef de projet - Sous-traitant G)

D'autres projets font appel à un travail de catégorisation plus complexe. C'est le cas d'un projet du sous-traitant G, qu'évoquent Beby et Dafy dans un entretien. Il est ici impossible de connaître l'utilité finale du projet, aussi nous faisons l'hypothèse qu'il s'agit d'un modèle de détection de la signalisation maritime, potentiellement utilisé pour la navigation autonome ou pour la surveillance des océans. Les travailleurs annotent *"les balises latérales mises en mer (...)"* et indiquent leurs catégories : *"balises latérales rouges, des balises latérales vertes ..."* (Beby - Annotateur - Sous-traitant G), c'est-à-dire à identifier les amers, des points de repère servant à la navigation maritime (phares, balises...). Il explique également les différents niveaux de catégorisation qui existent dans le projet. Un premier niveau concerne la catégorie de l'objet, une balise ou un phare, un second la couleur de la balise, un dernier la direction de l'objet (Ouest, Est...).

Encadré n°48 : L'annotation des "amers maritimes"

X : c'est des cartes maritimes, c'est-à-dire ?

D : les cartes maritimes et les AMER (...) maritimes aussi - ce sont (...) les balises de signalisation maritime, (...). Ça consistait à prendre dans une boîte englobante les catégories, les différentes catégories des différentes AMER, les différentes balises de signalisation maritime, et dire de quelle couleur elle est, verte ou rouge, de quelle catégorie elle est, est-ce que c'est une balise latérale verte ou rouge ? une balise cardinale nord/sud/est/ouest, est-ce que c'est un phare ? (...) la signalisation maritime voilà

Dafy - Annotateur - Sous-traitant G

Nous avons déjà évoqué des projets d'annotation de vêtements. Sur un projet similaire, Sahotay se retrouve à classer des images relativement simples : chaussures, pantalon. Toutefois, le projet évolue vers une complexification des catégories, ce qui déclenche alors de nouvelles phases de formation. C'est le cas sur un projet similaire traité par Lita : *"mon travail, mes tâches, c'était de déterminer si par exemple il y a une robe"*. Le deuxième niveau consiste à identifier la forme du vêtement : *"on détermine si cette robe est longue, si il y a des manches, les cols ... des trucs comme ça"* (Lita - Annotatrice - Sous-traitant G). Il peut aussi s'agir d'identifier la couleur ou la coupe du vêtement (un col rond ou un col Mao). Dès que les projets intègrent différents niveaux de catégorisation, les annotateurs font face à davantage de doutes et se retrouvent alors faire appel au guide d'annotation, à leur manager ou même directement au client. Ces deux dimensions, variété des classes et *"co-production"* des annotations renforcent la part d'interprétation possible et peuvent donc affecter le modèle.

Certains projets de classement renforcent encore davantage cette part d'interprétation du travail d'annotation. Sur un projet d'automatisation des dommages des poteaux électriques, les annotateurs du sous-traitant G doivent encadrer la rouille, puis attribuer un classement aux poteaux, selon le niveau de dégrada-

tion ²³⁸. Pour réussir à différencier le niveau d'altération des poteaux, Dafy a accès à un PDF qu'il a examiné "avant de commencer à travailler" pour vérifier toutes les catégories existantes et apprendre les consignes. Sur ce type de projet la part d'interprétation est très importante, du fait de la difficulté d'apprécier le niveau de dégradation des poteaux : "[Il faut encadrer la rouille] il y a différentes catégories, genre il y a la catégorie A qui est la simple rouille, il n'y a plus de peinture sur le poteau, ou bien la catégorie B, la rouille commence à s'enfoncer dedans, la catégorie C est un peu plus grave, la catégorie D ainsi de suite jusqu'à la catégorie E, jusqu'à ce que le poteau soit complètement détruit" (Dafy - Annotateur - Sous-traitant G).

D'ailleurs, Dafy raconte une séquence pendant laquelle il a eu de nombreux doutes concernant des annotations au tout début du projet. C'est dans ces moments-là que les annotateurs discutent entre eux, avec leur chef ou avec le client, afin de se mettre d'accord. Ainsi, Dafy raconte qu'il s'est retrouvé à faire l'annotation en compagnie de son chef de projet à qui il a demandé de l'aide : "on était tous les 2 dans le doute parce que le projet vient de commencer ... on s'est dit « viens, on va regarder une à une ces rouilles », il y avait toutes les catégories, et on a commencé à trier une à une, que celle-ci c'est une catégorie A, B, C ... comme ça." (Dafy - Annotateur - Sous-traitant G). Toutefois, le fait d'être à deux peut aussi être source de dissensus sur l'annotation. Dafy explique qu'"il y avait certains moments où on ne s'est pas mis d'accord, parce que c'était trop difficile de juger les rouilles".

Malgré les exemples donnés dans le guide d'annotation, ce dernier n'intègre pas forcément tous les cas de figure présents dans les images à annoter. Par ailleurs, les différences entre les catégories de rouille sont parfois ténues. C'est avec la montée en compétences au fil du projet et l'habitude de visionner les images de rouille et de naviguer entre les catégories que le projet est devenu plus facile pour Dafy, ce qui n'empêche pas la présence de cas particuliers pour lesquels il est nécessaire de demander au client. Dafy raconte le cas d'un poteau pour lequel il a posé la question au client. Or, ce dernier s'est lui-même révélé incapable de donner la bonne solution, tout simplement parce qu'elle n'existe pas : "ils nous ont juste répondu que ces trucs sont faits pour être jugés par l'œil humain parce que l'annotation ça consiste à aider l'intelligence artificielle, et ils ont juste répondu « jugez les par vous-même », c'était comme ça, c'est ce qu'ils nous ont répondu" (Dafy - Annotateur - Sous-traitant G).

Ce travail de classement est parfois rendu difficile par la différence culturelle entre les annotateurs de Madagascar et les entreprises d'IA françaises ou américaines qui demandent des tâches contextuelles, c'est-à-dire relevant de l'environnement d'un pays occidental. Un projet illustre particulièrement cette dimension : l'automatisation de la modération de contenu vidéo d'un réseau social déjà évoqué dans notre chapitre précédent. Les annotateurs catégorisent des vidéos signalées par les utilisateurs, à l'aide de labels comme

238. "aussi il y a le projet sur lequel je travaille en ce moment (...) ça consiste à voir les rouilles sur les poteaux, sur les pylônes, et de différencier par rapport aux dégâts, les dégâts qui s'affichent sur les pylônes, genre si il y avait beaucoup de rouille, d'érosion ... " (Dafy - Annotateur - Sous-traitant G)

”opinion”, ”choc”, ”sarcasme”, ”race”, ”Genre”... Ces vidéos proviennent toutes d'utilisateurs francophones et contiennent des éléments politiques ou sociaux relatifs à ce contexte. Les difficultés des annotateurs proviennent alors de la différence culturelle : comment juger un contenu sexuel? Comment juger le racisme? Comment identifier le second degré? Un annotateur raconte qu'il : *”Il faut vraiment être quelqu'un qui visionne beaucoup de films (étrangers) pour reconnaître directement que cette image est peut-être caractérisée dans des critères que le client a demandé. Et chez nous, dans le pays, la culture est assez limitée pour ce genre de choses.”*

Ce travail de catégorisation contient donc une part importante d'interprétation et de jugement basée sur la capacité des annotateurs à faire la synthèse entre les consignes du client (guide d'annotation) et les images qu'ils visionnent au quotidien (les données). Cette dimension implique et favorise la nature collective du travail d'annotation, les travailleurs étant tenus de communiquer entre eux. Certains projets sont certes plus simples, car ils comportent moins de catégories. Pour les autres, la longueur du guide, le chevauchement des classes ou bien leur caractère plus spécialisé nécessitent une formation initiale quasi constante, au fil des évolutions du projet. Les annotateurs présents sur ces projets y expriment leur montée en compétences. Comme pour le classement, le travail de saisie peut être considéré comme le second niveau de compétences, pour des raisons différentes. Cette dimension du travail d'annotation existe spécifiquement dans les projets liés au texte ou au son. Ce travail n'est pas spécifique à l'annotation et existe également sur des projets de ”numérisation”, comme la saisie de chèques ou la retranscription d'échanges enregistrés pendant des réunions.

Appliqué à l'automatisation algorithmique, ce travail de saisie concerne des projets comme l'automatisation de la détection d'éléments sur des factures. Sur ces projets, les travailleurs doivent par exemple identifier les montants présents sur les factures (montant hors taxe, montant de la TVA...) et les retranscrire sur une interface. Pour des assistants vocaux, les annotateurs doivent cette fois écouter des extraits sonores et les retranscrire. Les entreprises se servent de ce travail pour comparer les résultats de leurs modèles avec la compréhension humaine. Les compétences nécessaires sont plus élaborées. Il s'agit d'un ”travail de clavier” qui nécessite une maîtrise importante de la langue du donneur d'ordre à l'écrit, ainsi qu'une importante concentration : *”Un bon annotateur, pour moi c'est les gens qui ne valorisent pas leur poste. [sic] Il y avait déjà des gens qui ont été licenciés pour ce statut, ils ne se concentrent pas sur leurs tâches alors que le travail a besoin d'une profonde concentration, d'un profond travail de choix”* (Maro - Annotatrice - Sous-traitant A)

Dans son livre, [Braverman](#) (1998) analyse le développement du travail bureaucratique comme étant une forme aboutie de travail manuel. Selon lui, le *”travail est toujours effectué dans le cerveau, mais le cerveau est utilisé comme l'équivalent de la main du travailleur de détail dans la production, saisissant et relâchant un seul morceau de ”données” à maintes reprises.”* (p.319). Le travail des données est en apparence une bonne illustration de cette fin du *”travail mental”* synonyme d'une forme extrême de déqualification. Dans un entretien,

Y., data analyst chez MLDupli, parle d'ailleurs de "manual work" pour désigner les personnes qu'elle dirige pour produire des données. De manière prémonitoire, Braverman (1998) dit encore que l'étape "suivante est l'élimination complète du processus de pensée - ou du moins dans la mesure où il est éliminé du travail humain - et l'augmentation des catégories d'employés de bureau dans lesquelles seul le travail manuel est effectué." (p.319) soutenant l'hypothèse de Stephen Mallet sur la disparition des compétences ouvrières au profit des machines, la fin du rapport de l'homme à l'objet final de son travail. Pourtant, si les annotateurs ne sont pas nécessairement au courant de la finalité algorithmique de leur travail, ils sont pleinement engagés dans un travail de classification d'éléments qui implique des prises de décisions et une interprétation des consignes, en fonction de la "réalité terrain". Le travail du clic se transforme alors en travail de la classification qui s'accompagne de processus de gestion et de remontée des doutes.

7.2.2 Comprendre, identifier, interpréter : des jugements experts orientés métiers

La section précédente nous a permis de décomposer les gestes et les compétences nécessaires à l'annotation. Le travail d'encadrement repose sur un travail de la souris, très varié selon les projets. Le travail de classification implique également cette gestuelle, mais passe par d'autres processus. Les annotateurs doivent catégoriser des éléments, en fonction des consignes des clients, consignes qu'ils doivent interpréter au regard de la réalité des données.

Dans les projets d'IA que nous abordons dans cette section, nous avons constaté des mécanismes de montée en compétences et de développement de l'expertise sur les domaines concernés par les projets d'annotation. Sur ce point, notre première partie a déjà traité du caractère incrémental de l'entraînement de l'IA et de l'annotation qui justifie des annotateurs présents sur du plus long terme. Par exemple, le projet d'automatisation de la détection des ECG est réalisé par des médecins ou par des techniciens en radiologie formés à la reconnaissance des électrocardiogrammes du coeur. Si l'on place ces projets le long d'un continuum "annotation de sens commun" / "annotation experte", ces derniers sont situés du côté de l'expertise appliquée au domaine de l'annotation ("*domain expertise*").

D'autres projets reposent également l'expertise des annotateurs. C'est le cas de celui réalisé par Tefy (sous-traitant G). Comme il l'explique, ses connaissances en agriculture justifient qu'il travaille sur ce projet ²³⁹. Ce dernier est en effet diplômé d'une licence professionnelle en agronomie et il a déjà exercé comme technicien agricole au sein d'une exploitation. Après une première expérience en tant qu'"opérateur en ligne" au sein d'une grande entreprise du secteur de l'externalisation, il s'est retrouvé à travailler chez le sous-traitant G sur un projet d'annotation agricole. Le projet consiste à identifier la qualité des produits agricoles en

239. Deux ans avant la réalisation des entretiens, le sous-traitant G a changé de direction, suite à la mort de l'ancien dirigeant. Plusieurs clients ont alors voulu partir, dont le client du projet de Tefy. La seule condition imposée par ce donneur d'ordres pour conserver son contrat a été de conserver Tefy sur le projet, car il le trouvait performant. Ce dernier s'est donc retrouvé à réaliser tout seul l'annotation de ce projet agricole qui consiste à annoter des productions agricoles (grains, raisins...).

déterminant un taux ²⁴⁰ : les grains sont-ils en bon ou en mauvais état. Le produit d'IA fonctionne avec des photographies prises par les utilisateurs : *"par exemple de l'orge sur un plateau, et donc là tout de suite à partir de l'application, on voit qu'il y a 100% de bon état ou ... il y en a d'autres c'est 50% en bon état"* (Tefy - Annotateur - Sous-traitant G).

L'annotation consiste à placer quatre points sur le raisin ou la céréale, puis à le catégoriser. Les classes sont composées de plusieurs niveaux : présence d'impuretés (feuilles) et état des grains. Il s'agit de les identifier : *"si les grains sont cassés, si les grains sont normaux, si il y a des feuilles (...) sur la table ou sur le plateau, si il y en a qui sont mures ou (...) pas mures ... il y en a aussi beaucoup des impuretés, par exemple il y a des herbes sur le plateau parce qu'il vient de la récolte (...)"* (Tefy - Annotateur - Sous-traitant G). Tefy considère que son expérience agricole lui sert dans son travail d'annotation. Le responsable français de la production de ce sous-traitant précise que c'est davantage l'autonomie de Tefy et ses compétences en français qui l'ont poussé à le placer sur ce projet, où il est nécessaire qu'il soit en contact direct avec le client, étant seul sur le projet. Son expertise en agriculture lui permet d'avoir des échanges plus détaillés avec le client, notamment en cas de doute ou d'image ambiguë. Tefy donne l'exemple d'un grain qu'il ne sait pas annoter, car ce dernier est légèrement fendu : *"J'avais des doutes parce que j'ai dit « est-ce que c'est un grain normal ou bien est-ce que c'est un grain cassé ? sur le bout du grain, c'était un peu coupé."* (Tefy - Annotateur - Sous-traitant G). L'extrait d'entretien ci-dessous résume le processus de résolution du problème sur ce cas. On constate que Tefy a dialogué avec le client par messagerie et, malgré un doute persistant, a accepté la réponse du client (il s'agit d'un grain normal), notamment car en tant qu'agriculteur malgache, il n'a pas eu l'occasion de voir des grains de blé. Par ailleurs, il est pris par l'impératif de quantité d'images à annoter, ce qui a pour conséquence de le limiter dans le temps qu'il peut consacrer à argumenter avec le client.

Encadré n°53 : L'annotation d'un projet agricole 1

X : quand vous voyez cette image de grain coupé, la première réaction c'est quoi ?

T : je fais une capture, et j'ai demandé l'avis du client, et j'ai posé des questions « est-ce que c'est un grain normal où est-ce que c'est un grain cassé ? »

X : vous étiez d'accord avec la réponse du client ?

T : [rires](...) en tant que agriculteur, moi je n'ai pas encore vu un grain de blé après la récolte, c'est ça le problème. Parce que ici c'est le riz, c'est ça le problème. Donc j'accepte la réponse du client. Mais à mon avis c'est coupé (et) si le grain est coupé c'est cassé sur la classe

X : donc vous n'étiez pas d'accord !

T : mais c'est le client, il a dit que c'est un grand normal, donc c'est normal!

240. "nous avons beaucoup d'images, dans le champ, des raisins dans le champ, des raisins qu'ils vont mettre sur des plateaux ... mais les images des raisins c'est arrêté, maintenant on est plutôt focalisé sur les céréales, orge, blé, maïs, pois ... beaucoup beaucoup" (Tefy - Annotateur - Sous-traitant G)

X : vous avez essayé de débattre avec lui ?

T : des fois ... ce n'est pas débattre mais on échange des discussions. Mais ce jour-là, j'avais beaucoup d'images à faire, donc celui qui a dit que c'est normal donc c'est normal ! parce que si je perds beaucoup de temps sur une image, avec une petite graine, ça va me retarder(...)

Tefy - Annotateur - Sous-traitant G

Au cours de l'entretien, il nous explique que si sur le cas précédent il n'a pas spécialement débattu de son jugement, il arrive aussi régulièrement que le client accepte son avis. Il donne l'exemple d'une annotation qu'il a réalisée très récemment au moment de l'entretien. Sur l'image il a identifié des impuretés sur l'image : une tige et une feuille collées ensemble. Sur ce cas, il a proposé au client d'annoter tout d'un coup, plutôt que de faire la segmentation des deux éléments, ce que le client a accepté.

L'exemple de Tefy illustre la capacité des annotateurs à comprendre la complexité des projets d'annotations, lui permettant d'identifier plus facilement des choses à améliorer ou bien des ambiguïtés difficiles à traiter. C'est par le biais de la messagerie que ces discussions sont traitées en direct avec le client. Son cas montre également la pertinence pour les clients d'avoir accès à des annotateurs compétents dans le domaine sur des projets spécialisés. La conséquence de ce mode de fonctionnement est qu'outre l'annotation pure (entourer, classifier), Tefy se retrouve à interpréter les consignes et les images, à tel point que le client intègre parfois sa vision au projet d'annotation.

Après avoir évoqué un projet qui porte sur le monde agricole, prolongeons la réflexion sur l'expertise des annotateurs en évoquant les modèles de reconnaissance de nourriture. Ce type de projet, déjà abordé dans notre première partie, est revenu plusieurs fois au cours de l'enquête. Nous avons déjà introduit le cas d'une start-up française d'IA opérant pour le compte d'une grande entreprise de restauration collective pour développer une caisse de cantine automatique. Le même principe est développé par une autre start-up française, cette fois uniquement spécialisée dans la détection d'aliments à travers deux produits : *"une application qui permet de connaître la valeur nutritionnelle des aliments"* (Élise - Annotatrice - Sous-traitant G) et une caisse automatique de cantine. Le rôle des annotateurs consiste cette fois à *"classifier tous les aliments dans des photos"* (Élise - Annotatrice - Sous-traitant G). Plus précisément, les annotateurs doivent entourer les ingrédients d'un plat (par exemple : des haricots verts), y compris lorsqu'ils sont mélangés (par exemple : un ragout de boeuf). L'image ci-dessous illustre les objectifs du projet : chaque élément y est annoté et nommé, ainsi que les aliments composants le plat principal.

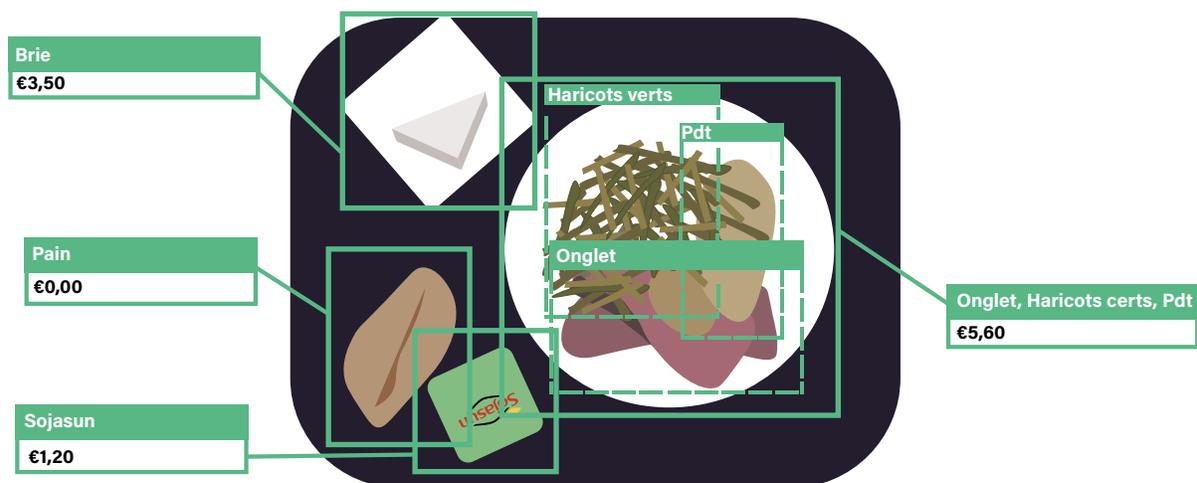


FIGURE 29 – L’annotation d’un plateau repas pour une caisse de cantine automatique

Le travail se décompose donc en deux tâches principales, toutes deux réalisées par les mêmes annotateurs : *”on fait la géométrie, puis la classification. D’abord on annote tout ce qui est dans l’assiette, puis on fait la classification”* (Naivo - Annotateur - Sous-traitant G). Comme l’explique Naivo, la difficulté du projet est plutôt liée à la classification qui *”très très détaillée, parce que tout ce qui est dans l’assiette, on doit le classifier”*. Il existe ainsi plusieurs niveaux d’annotations : d’abord les éléments du plateau repas (pain, yaourt), puis le descriptif des éléments des plats (ingrédient mélangé ou non) et la labellisation des ingrédients (purée ou viande d’agneau). Cette difficulté est ressentie par plusieurs annotateurs, comme Élise qui considère que *”[ce] n’est pas vraiment facile, y a des tonnes de choses à connaître (il faut) retenir le nom des choses.”* Chaque annotateur réalise dans une même semaine, soit la *”géométrie”*, soit la classification, comme l’explique Naivo : *”C’est le client qui nous a donné l’emploi du temps. Le lundi mardi mercredi, vous allez faire ça, et après on travaille là [soit géométrie, soit classification].”* Une douzaine d’annotateurs sont impliqués sur ce projet : huit personnes sont des annotateurs simples, il y a aussi trois correcteurs et deux chefs de projet, dont un en France.

Les travailleurs s’appuient sur plusieurs outils, mis en place de manière collaborative entre le client final, l’équipe managériale en France et le chef de projet à Madagascar qui servent à expliquer les différentes classes. L’utilisation de ces supports est permanente, étant donné la présence de nombreuses classes : *”Classe viande, classe fromage, cuisine asiatique, cuisine du monde. Il y a plus de 3000 ingrédients à mémoriser.”* (Pota - Annotateur - Sous-traitant G). Une base de données regroupe les images de référence pour l’annotation. Il s’agit d’un tableur qui regroupe d’un côté des images d’exemple et de l’autre la classe à y associer. Un autre support sert quant à lui à *”connaître la classe : « dessert emballé », « aliment simple », « mélange »”* (Angy - Annotateur - Sous-traitant G). Il existe également plusieurs supports de formation, un par classe d’aliment :

”(...) *Nous avons un support de viande, concernant les viandes, les soupes, fromage, pain ... il y en a pleins.*” (Angy - Annotateur - Sous-traitant G). Angy indique également que ces supports ne sont pas complets, c’est-à-dire qu’ils ne prévoient pas l’ensemble des cas potentiels ²⁴¹. Son travail consiste alors à aider Anaïs, cheffe de projet en France et également interrogée dans le cadre de cette recherche, à compléter cette base de connaissances ²⁴².

Dès que la classification est complexe, cela nécessite un travail plus important de la part des annotateurs, notamment sur le plan de la vigilance. Ce cas illustre également les limites du travail de classification. Bowker et Star (2000) nous avertissent bien sur ce point. Pour être fonctionnelle, une classification doit se traduire par la *”couverture totale du monde qu’il décrit”*. Or, il paraît vain, du moins difficile, de réussir à classer l’ensemble des plats et ingrédients du monde entier (variété des données). Par ailleurs, l’alimentation a vocation à évoluer, dès lors comment se passer de l’annotation (*volatilité de la vérité terrain*) ? Les annotateurs produisent ainsi une *”vérité terrain”* qui est susceptible d’évoluer constamment, à mesure que le projet intègre de nouvelles données et donc de nouvelles classes.

Au quotidien, cela se traduit également par des compétences de résolution de problèmes, lorsqu’un cas n’est pas prévu. Ces phases de résolution de problèmes sont particulièrement visibles lorsque les annotateurs doutent d’une catégorisation. Angy explique ainsi avoir eu du mal à distinguer différentes marques de yaourt sur un plateau et devoir systématiquement consulter le document de formation pour annoter correctement. Cette dimension du travail est d’ailleurs renforcée par les évolutions des consignes d’annotation : *”Parfois s’il y a des consignes qui changent, par exemple, notre client(...) s’il donne des idées de (consignes) s’il faut changer ça, ça change.”* (Eve - Annotateur - Sous-traitant G). Ainsi, depuis 2015 et le démarrage de cette start-up, les campagnes d’annotations se sont succédé en fonction des évolutions de leur produit. Comme dans le cas de Tefy, les annotateurs peuvent également faire remonter des idées ou des problèmes à corriger comme l’explique Eve : *”Mais si on a des idées, on le propose et s’ils acceptent ils changent, nous aussi on peut donner notre idée s’il y a quelque chose qui ne fonctionne pas, et comme ça c’est eux qui valident”*.

La mise en oeuvre de ces différentes couches de travail humain nécessaires à l’interprétation des données, y compris en identifiant les cas non prévus initialement, sert à maximiser la qualité et la quantité des données. Pour la nourriture, cet objectif est difficile à atteindre selon Angy : *”je pense pas que on peut arriver à 100%, ce n’est pas possible, parce que même il y a ... même moi je ne sais pas. Il y a quelques ambiguïtés dans le projet”* (Angy - Annotateur - Sous-traitant G) En prenant l’exemple des fromages frais, il explique qu’il est difficile d’atteindre les 100% de détection, tant les ingrédients se ressemblent. Comment identifier très clairement

241. *”Mais c’est encore un peu difficile (...) il n’est pas encore complet à 100% pour que l’équipe ne manque de rien. C’est Anaïs qui crée les supports”* (Angy - Annotateur - Sous-traitant G)

242. *”moi j’aide tout simplement Anaïs à insérer quelques images qui ne sont pas dans (...) on s’entraide comme ça.”* (Angy - Annotateur - Sous-traitant G)

une purée de lentilles vertes d'une purée de brocoli sur la base d'une seule image? Lorsque les ingrédients sont mélangés ou recouverts de sauce, il est également complexe d'isoler chaque élément séparément.

Encadré n°54 : Les ambiguïtés de l'annotation de nourriture

X : c'est quoi les ambiguïtés ?

B : parfois, quand on voit du fromage par exemple, le fromage frais ou la ricotta dans une salade, c'est ... ça c'est vraiment dur ! on ne peut pas se permettre de dire n'importe quoi, mais on doit reconnaître c'est pour ça que peut être des fois on n'est pas à 100% (...)

X : du coup l'application elle ne va pas le reconnaître si vous n'y arrivez pas ?

B : voilà c'est ça (...) si toutes les images étaient simples on pourrait faire 100%, mais vu qu'il y a des images complexes ... surtout quand c'est mélangé, on arrive plus à détecter c'est quoi. Parfois les viandes sont recouvertes de sauces, on ne reconnaît pas si c'est du poulet ou de la dinde ...

Tefy - Annotateur - Sous-traitant G

Le projet est également rendu complexe du fait de la variété des cultures gastronomiques représentées dans les images. A l'époque de sa première semaine sur le projet, Angy est ainsi obligé de faire des recherches sur Google pour identifier certains types de plats ou de sauces. Il explique aussi qu'il a eu du mal à reconnaître certains plats provenant d'autres cultures gastronomiques ("*les plats typiques maghrébins*"). Cette méconnaissance est liée à la localisation de la production, ces travailleurs ayant une culture gastronomique très différente de la culture française. D'autres types de cuisine aujourd'hui bien représentées dans les cantines françaises (indiennes, chinoises...) sont également présentes dans les images : "*Là, sincèrement c'est pas facile parce que 80% des projets, ce n'est pas dans notre culture, c'est pour ça que c'est vraiment difficile. La gastronomie mexicaine, la gastronomie européenne, chinoise, indienne ...*" (Angy - Annotateur - Sous-traitant G)

Il cite également l'exemple des fromages français, comme étant particulièrement difficiles à retenir et à différencier : "*Le problème c'est les ingrédients, par exemple les fromages, chez vous il y a plein de types de fromages ! là vraiment c'est galère de le faire ! nous sommes dans le projet fromage là, et c'est vraiment difficile à maîtriser*" (Angy - Annotateur - Sous-traitant G). Sur ce projet, les travailleurs doivent également isoler les ingrédients d'un plat, opération particulièrement complexe lorsque les plats sont composés de nombreux éléments : "*les viandes et tout ça ... parce que nous les malgaches on n'a pas trop de gastronomie, à part le romazava, des trucs comme ça ... mais vous, vous avez plein de plats gastronomiques ! c'est difficile à maîtriser ! il y a des sauces c'est la première fois que j'ai connu les sauces ... des sauces tartare ... il y en a plein, sauce à la crème, béarnaise, hollandaise ! c'est difficile vraiment à maîtriser !*" (Angy - Annotateur - Sous-traitant G) Ce projet "très difficile", selon les mots de leur dirigeant, a poussé ce dernier à changer l'organisation du

travail en cours de projet, en recrutant une cheffe de projet présente en France pour faire l'interaction entre le processus d'annotation et le client final : *"Avant quand il n'y avait pas Anaïs j'étais en contact direct avec les clients, mais il n'a pas trop [le temps] c'est pour ça que Pierre a mis Anaïs, parce qu'elle a plus de connaissances que nous sur les (plats), elle a plus de savoir par rapport à nous, c'est pour ça qu'on a besoin d'elle"* (Angy - Annotateur - Sous-traitant G). Son rôle est de régler le projet en aidant les annotateurs à gérer les problèmes d'annotation.

Comme l'explique Romane dans l'extrait d'entretien qui suit, cette chaîne hiérarchique de gestion des ambiguïtés intègre différentes strates. Les annotateurs peuvent d'abord regarder le support et demander à leurs collègues ou à leur chef de projet à Madagascar. Dans le cas de Romane, il a également fallu faire remonter l'information à Anaïs : *"J'ai d'abord regardé le support, mais la photo ne correspondait pas, donc j'ai demandé au chef de projet. Donc pour ce cas-là, c'était un morceau de cervelas, j'ai regardé le support, mais y a eu hésitation donc je demande directement à Anaïs par Slack."* (Romane - Annotatrice - Sous-traitant G)

L'organisation du projet, sa durée et ses difficultés conduisent les dirigeants de l'entreprise d'annotation à constater la montée en compétences des agents. Au fur et à mesure, ces derniers se perfectionnent dans la détection de données très variables et des anomalies. Il s'agit d'une dimension déjà identifiée dans le cas d'une autre entreprise d'IA française évoquée dans notre premier chapitre et qui considère ses annotateurs comme *"ses propres employés"*. Dans l'extrait qui suit, le dirigeant du sous-traitant G ajoute que c'est ce qui différencie ce mode d'organisation avec le "marché pur" (PNT) et dans lequel il existe un *"risque de perte de compétences, parce qu'on ne sait jamais sur qui on va tomber, et si les gens vont rester sur la durée"*. Sur ce projet, il précise aussi qu'il *"a quand même fallu du temps pour que nos annotateurs et les revieweurs développent (leurs compétences) AIAlim c'est autour de (plus de 1000) classes de nourriture aujourd'hui (...) AIAlim, il y a une vraie complexité de l'annotation (...)"* (Pierre - CEO - Sous-traitant G). Il est également nécessaire de mettre en place une organisation du travail pour favoriser la gestion de cette complexité : *"Pour certains types de projets, il faut que les gens soient tout le temps sur place, soient encadrés, qu'il y ait des chefs d'équipes qui comprennent mieux, qu'ils puissent vraiment leur expliquer ce qu'il faut faire."* (Pierre - CEO - Sous-traitant G).

Un autre projet a retenu notre attention pour les dynamiques de montée en compétences qu'on a pu y observer. Il est également réalisé du sous-traitant B, une entreprise formelle, qui fait partie des acteurs relativement anciens du secteur de l'externalisation (depuis 2008). Sur ce projet, c'est plus d'une centaine de travailleurs qui font de l'annotation. Pendant une semaine, nous avons pu interroger une partie de leurs travailleurs directement depuis leurs locaux. On y a constaté la modernité de cette entreprise : il s'agit d'un plateau sécurisé et très bien équipé, puisque les annotateurs disposent de deux écrans. Ce dernier détail du décorum a son importance, car le projet consiste à annoter des images satellites (voir : l'image ci-dessous).

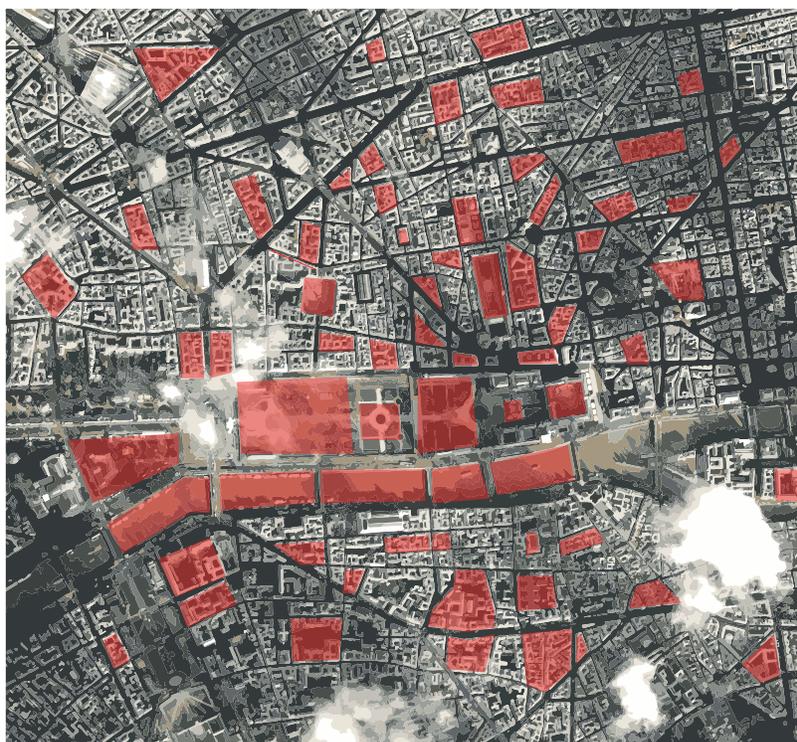


FIGURE 30 – Exemple d'un projet d'annotation d'une image satellite

Ces images de très grandes tailles comportent souvent des centaines d'éléments à identifier. Ces éléments sont de différentes natures, comme le précise Asy, annotateur chez ce sous-traitant : *"Le projet consiste en des annotations d'images satellites. Des images de voitures, de bateaux, de bâtiments, de routes."* Initialement, le projet ne comportait qu'une phase de détourage des éléments comme l'explique Fidy : *"Avant on ne classifie pas (...) on fait tout simplement des annotations. Et nous n'avons pas de classification à cette époque."* Par la suite, le projet s'est élargi à la classification de ces éléments. D'après le jargon utilisé par les annotateurs, il s'agit d'identifier des *"observables (...) dans une zone donnée (...) Donc ils nous envoyaient des images satellites. Et après, nous on va juste dire « là y'a un bateau », on va dire quel type de bateau c'est, y'a tout un process (...)"* (Ando - Annotateur - Sous-traitant B). Le projet sert en fait à entraîner un modèle de détection qui s'applique à la surveillance de zones en tension (par exemple : zones de guerre). L'image ci-dessous illustre le processus d'annotation : les annotateurs doivent identifier les éléments, puis y apposer un label. La différence, c'est que les images traitées par les annotateurs sur ce projet représentent un territoire beaucoup plus vaste. Certains éléments deviennent alors beaucoup plus difficiles à repérer.

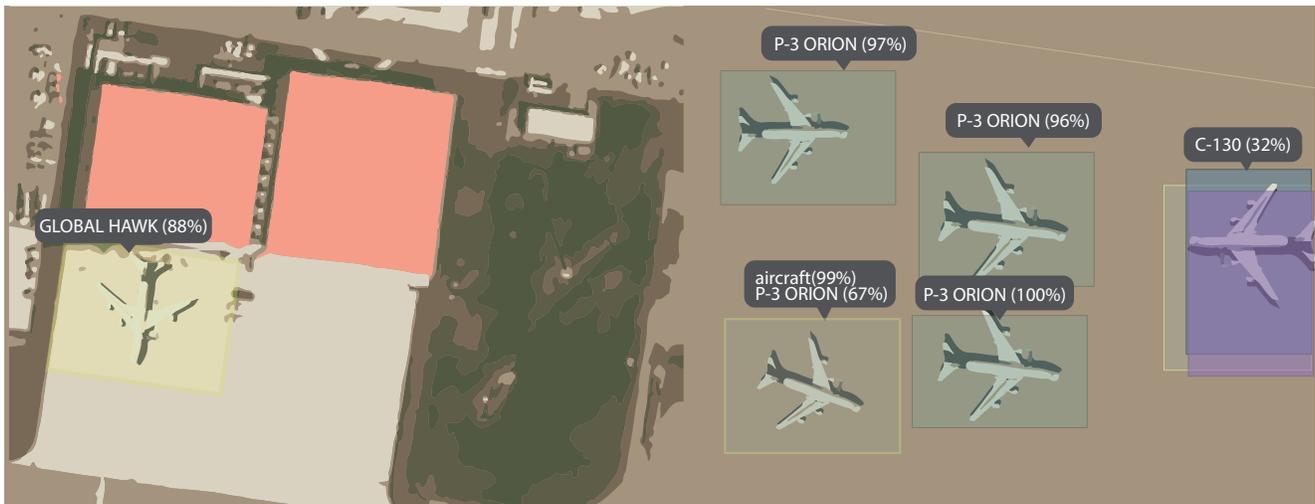


FIGURE 31 – Exemple d'un projet d'annotation pour la classification d'avions

Au fur et à mesure, la partie "classification" du projet est devenue de plus en plus complexe en intégrant de plus en plus de niveaux d'annotations ce qui implique la montée en compétences des agents et une organisation du travail de gestion des doutes. Les niveaux de classification vont de un à quatre, en fonction de la précision de l'élément à classifier. Le niveau un consiste à labéliser l'élément comme étant un véhicule et le quatre à identifier un élément présent sur le véhicule : *"Ça a évolué. Avant, quand on tague des véhicules, ça reste tout simplement sur level 1 véhicule. Mais après on nous a donné petit à petit les Levels jusque 4, level deux par exemple pour les véhicules civils, et level trois si c'est des semi-remorques, ou des utilitaires. Et level 4, c'est par rapport ... par exemple sur les semis, il y a des vannes (sur) les pétroliers et tout"* (Fidy - Annotateur - Sous-traitant B). Par ailleurs, le projet s'organise en fonction de différents types d'annotations, ce qui ajoute encore des difficultés. Régulièrement, les "observables" changent. Des véhicules sur route, les annotateurs passent ensuite aux avions ou aux bâtiments en fonction des demandes des clients. Cette rotation des agents se base également sur la spécialisation de ces derniers : *"Il y a plusieurs batches, et dans un batch il y a plusieurs sites, et plusieurs dates. Si le client demande on ne veut que des véhicules par exemple, et le client dit on veut 15 agents, 20 agents, le chef d'équipe regarde qui est fort en camion, et c'est comme ça (...)"* (Asy - Annotateur - Sous-traitant B)

Un autre aspect problématique du projet concerne la nécessité de répondre correctement aux exigences des clients, au regard de la finalité de l'automatisation. Comme l'explique Asy : *"Chez AIWar, on rencontre toujours des difficultés, donc on a toujours des doutes sur tout, parce que le client est très exigeant, parce que c'est le travail qu'on fait qui va aider l'algorithme dans ce qu'il fait. Si on fait des erreurs, ça passe pas."* Cet aspect, généralisable à l'ensemble des projets d'IA, concerne le degré de précision attendu pour le modèle et la variété des cas de figures qu'il doit intégrer. Ainsi, l'entraînement de voitures autonomes doit aboutir

à un très haut pourcentage de détection des éléments présents dans l'image. Par ailleurs, comme pour le cas de la nourriture, les cas de figure rencontrés sur les routes seront très variés.

En fonction de ces deux éléments, le travail d'annotation sera plus ou moins exigeant et nécessitera plus ou moins de compétences. La nourriture illustre par exemple la variété des données nécessaires, même si le degré de précision attendu pour le modèle n'est pas aussi élevé que sur AIWar (*sensibilité du projet*).²⁴³ Ici le projet requiert un haut niveau d'exigence, car il s'applique car un domaine extrêmement sensible : la surveillance des zones de guerre. Au-delà de ces deux aspects, la qualité des données peut également rendre plus complexe le travail de détection. Sur ce projet, les travailleurs ont accès à des images satellites représentant de grandes zones terrestres. Les éléments à détecter sont parfois très pixelisés. Aussi, si les deux premiers niveaux de la classification sont aisément identifiables, ce n'est pas nécessairement le cas pour les niveaux trois (type de véhicule) et quatre (éléments sur le véhicule) : *"l'image est très mauvaise, on distingue presque rien, mais on sait que c'est un véhicule militaire. On voit, mais ses caractéristiques (c'est plus compliqué)." (Asy - Annotateur - Sous-traitant B).*

La difficulté de ce projet provient, de la quantité de consignes à maîtriser : *"Il y a beaucoup de pdf qu'on doit apprendre (...) A peu près 60 pages, ou plus [pour les guides]" (Raja - Annotateur - Sous-traitant G).* Chaque catégorie d'annotation comporte ainsi son propre guide qui détaille l'ensemble des objets à identifier. Il existe par exemple des guides entièrement dédiés aux hélicoptères ou aux bateaux. Comme l'explique Fidy, un annotateur rencontré sur son lieu de travail, les PDF contiennent de nombreux exemples. Cependant, les travailleurs doivent également faire des recherches complémentaires lorsque ces exemples ne sont pas suffisamment bien expliqués, ou qu'ils ne sont pas présents dans les guides²⁴⁴. Plusieurs annotateurs de ce projet expliquent que leur démarche d'identification des éléments nécessite qu'ils fassent des recherches sur des archives d'images satellites tirées de Google Maps. Cela leur permet par exemple de valider la présence ou non d'un bâtiment qu'ils auraient pu confondre avec un autre élément.

Ces difficultés amènent aussi les annotateurs à développer des techniques d'entraide. Comme l'explique Raja : *"Il faut s'entraider à demander aux teams leaders, aux collègues"* (Raja - Annotateur - Sous-traitant G). Ces mécanismes de coopération se déroulent aussi sur des messageries internes. Fidy nous explique ainsi sa méthode de résolution d'un problème concernant l'identification des tentes. Dans ce cas, il a montré son problème dans une conversation de groupe sur Skype et qui comporte les annotateurs dédiés à l'identification de cet "observable" (les tentes). Il explique également que chaque catégorie d'annotation dispose d'un

243. Cette distinction a été abordée par le dirigeant du sous-traitant F, qui, dans un entretien, comparait la voiture autonome avec des projets d'identification de produits mal rangés dans un supermarché. La plus grande complexité de l'annotation s'expliquant par la criticité de la finalité du projet d'automatisation

244. *"F : Les clients nous donnent les PDF sur les classifications, il y a beaucoup d'exemples, mais il faut faire beaucoup de recherches (en plus)" (Fidy - Annotateur - Sous-traitant B)*

groupe d'entraide spécialisé²⁴⁵. Les annotateurs prennent donc parfois des décisions collectives, de manière à éviter de "déranger le client". On retrouve ici une tension entre le niveau de qualité espéré et le fait que le client passe par un service externalisé. Ce dernier n'a pas vocation à vérifier chaque annotation, surtout dans ce projet qui comporte plus d'une centaine de travailleurs. Flore, une des rares femmes annotatrices rencontrées durant notre terrain, nous explique une difficulté qu'elle a eue sur un observable. Il s'agit d'une image de bateau "très floue" et qu'elle n'arrive donc pas à classifier. Elle a donc décidé de faire voter ses trente collègues via un groupe dédié à cet observable, puis de vérifier sur Google Earth et enfin de confirmer auprès de son manager²⁴⁶.

Tsiky, une autre annotatrice, nous explique également sa méthode de résolution de problème sur le même projet. Dans son cas, elle est en difficulté après avoir "changé de chef d'équipe", ce qui l'oblige à se "débrouiller toute seule" sur l'annotation. Sur cette image, son problème est lié à l'identification des catégories d'un bateau. Elle commence par chercher l'image d'un bateau similaire sur Google : "On regarde le contexte de l'image, et si c'est un bateau militaire, on cherche bateau militaire, on regarde un à un et s'il y a des images similaires on regarde et on lit les articles. C'était pas très nette la réponse et donc j'étais forcée de demander ce que c'était." (Tia - Annotatrice - Sous-traitant B). Sans réponse claire, elle s'est ensuite résolue à demander au groupe. Au sein de cette discussion, ce sont douze annotateurs qui lui répondent et lui demandent d'explicitier le cas : "ils me demandaient d'abord les mesures, la taille plutôt, les contextes comme je l'ai dit et si le [bateau] transportait quelque chose." (Tia - Annotatrice - Sous-traitant B). Après cette phase de discussion, un vote a lieu et indique de ne pas annoter l'image, car il s'agissait en fait d'un bateau en construction²⁴⁷.

Cet exemple montre que quand des données nouvelles émergent, il existe des mécanismes de gestion au sein des entreprises d'annotations. Les annotateurs doivent faire un travail de recherche, se coordonner entre eux et en dernier ressort demander aux clients. Ces méthodes, codifiées par l'entreprise d'annotation, illustrent les décisions que prennent les annotateurs face aux données ambiguës ou imprévues. Ces décisions concernent à la fois le degré de difficulté d'une image et la manière de résoudre le problème, si possible sans "déranger le client", mais aussi l'objet présent sur l'image. Flore explique également un autre doute, qui concerne cette fois les tentes militaires.

245. "Q : [Votre doute] (...) c'était plus ou moins sur quoi ? R : je travaille sur les tentes, les abris, et je n'ai pas encore bien étudié les classifications et les contextes, je suis un peu confuse (...) on a des groupes Skype pour chaque observable, quand on a des doutes, on demande sur Skype" (Fidy - Annotateur - Sous-traitant B)

246. "La première fois c'était sur une image très floue (de) bateau. Elle a du faire un petit vote, pour voir combien de personnes ont répondu quelle réponse, puis regarder sur google earth pour prendre la décision. Puis elle a été voir l'encadreur pour confirmer. Oui, il a confirmé ce que la majorité avait dit sans aller déranger le client." (Flore - Annotatrice - Sous-traitant B)

247. "(...) on prend ce que la majorité des personnes voyait, on le classifiait en tant qu'inconnu, oui parce que c'était un bateau en cours de construction et à la fin le client a dit qu'on prenait pas en compte ce genre d'images." (Tia - Annotatrice - Sous-traitant B)

Encadré n°55 : Les ambiguïtés de l'annotation des tentes militaires

Les tentes(...) militaires c'est difficile à décrire, à cause de la couleur de l'image. On devait regarder le contexte [l'environnement], si c'est une zone militaire ou une zone civile, et des fois dans une zone civile il y a soudainement une tente militaire qui se monte. Et si une tente était dans une zone civile, c'était par la couleur de la tente qu'on distinguait. C'était en Lybie. J'ai fait des recherches, et après quand j'avais des doutes, j'ai demandé à mon team leader, et au chef d'équipe. Sur skype. Oui, il a répondu tout de suite. Lui aussi il a fait des recherches, pour ce cas-ci il a fait des recherches puis il a conclu que c'était une tente civile vu que c'était un hangar d'entretien des véhicules.

Flore - Annotatrice - Sous-traitant B

Ce projet illustre donc nos arguments en faveur d'une analyse des pratiques d'annotations comme relevant d'un travail plus qualifié qu'en apparence. Un premier argument repose sur la nature du projet. Lorsque ce dernier comporte des images difficiles, variées et un nombre important de catégories, cela renforce le travail d'interprétation réalisé par les annotateurs. C'est alors leur capacité à résoudre le problème à travers des recherches et de l'entraide qui entre en jeu et non uniquement la compréhension des consignes, par définition incomplètes des clients. L'annotateur a donc un rôle à jouer pour prendre des décisions concernant chacune des données traitées. D'abord, sur l'évaluation du degré de difficulté d'une image : "dois-je faire remonter ce doute à mes collègues, au chef de projet ou au client ?" Ces décisions peuvent alors être individuelles lorsque l'annotateur décide d'annoter quand même l'image, ou bien collectives lorsqu'il associe un groupe de discussion, voire hiérarchiques lorsqu'il associe son chef de projet ou le client.

On constate également que les entreprises d'annotations ont codifié des pratiques d'annotations, de manière à ce que les annotateurs cherchent le plus possible les solutions à apporter aux données inconnues, sans avoir à solliciter le client final. Cela passe par la codification d'un processus de production qui implique une communication soutenue entre les annotateurs, à travers des groupes de discussion. Ces groupes servent aussi à faire voter les annotateurs pour faire ressortir un jugement majoritaire sur ces cas complexes. Sur ces groupes, certains travailleurs prennent d'ailleurs un rôle de référent informel en conseillant régulièrement leurs collègues.

Les données tirées de notre questionnaire tendent à confirmer que la variable de complexité du projet est une variable déterminante dans l'évaluation du poste de travail en termes de compétences communicationnelles et de variété des tâches à réaliser. Un premier tri croisé confirme que c'est chez les sous-traitants G et B, les entreprises qui gèrent les deux projets précités, que les annotateurs demandent des explications aux clients. C'est au sein de ces entreprises qu'on a rencontré le plus de variété de projets et les projets les plus complexes. Ce résultat est confirmé par le tri croisé entre le sentiment d'incompétence des travailleurs

et leur appartenance aux entreprises de notre échantillon. À nouveau, ce sont les travailleurs des sous-traitants G et B qui se sentent le plus incompetents. De notre point de vue, cela soutient l'hypothèse d'une spécialisation du travail d'annotation, liée à l'évolution du secteur de l'IA vers des applications beaucoup plus spécifiques, notamment dans le cas de la recherche scientifique et des IA génératives comme ChatGPT.

	Oui	Non ¹
Sous-traitant G	74,5	25,5
Sous-traitant B	53,8	46,2
Sous-traitant A	65,4	34,6
Ensemble	48,6	51,4

¹Réponses en pourcentage, p-value : 0.00005956

TABLE 9 – Appartenance à une entreprise de notre échantillon croisé avec les demandes de précision aux clients

Par ailleurs, il est important de souligner que les annotateurs sont 76,4% à se sentir autonomes dans leur tâches, ce qui confirme l'hypothèse d'une forme d'agentivité de ces derniers. En revanche, cette agentivité est fortement dépendante de l'entreprise cliente et donc des projets sur lesquels les annotateurs travaillent. Notre travail s'inscrit de ce fait dans l'attention au travail réel. Ainsi, malgré un discours managérial fortement empreint de contrôle, on a pu constater que les tâches ne s'imposent pas d'elles-mêmes aux annotateurs, d'autant plus si le projet est complexe. C'est d'autant plus vrai, que, comme on le verra dans la partie suivante, le travail d'annotation ne s'impose pas non plus nécessairement aux donneurs d'ordres. Les annotateurs sont ainsi amenés à prendre des décisions sur le fait d'annoter, sur comment annoter et sur le fait de faire remonter des éléments étranges (*outliers*) à leur chaîne hiérarchique.

	Oui	Non ¹
Sous-traitant G	80,9	19,1
Sous-traitant B	82,8	17,2
Sous-traitant A	60,3	39,7
Ensemble	74,3	25,7

¹Réponses en pourcentage, p-value = 0.00233

TABLE 10 – Appartenance à une entreprise de notre échantillon croisé avec le sentiment d'autonomie

Dans cette thèse, nous avons jusqu'à présent utilisé plusieurs termes pour désigner les travailleurs étudiés lors de notre terrain à Madagascar. "Micro-travailleurs" ou "travailleurs des données" sont deux catégories

conceptuelles qui renvoient, d'un côté à une définition basée sur l'organisation du travail d'annotation par les plateformes numériques de travail et de l'autre au travail tel qu'il est envisagé du point de vue des entreprises d'annotation de données. Dans les différentes sections de ce chapitre, nous avons souligné les qualifications nécessaires pour exercer ce travail. Ces savoir-faire sont construits à travers la formation initiale des annotateurs, souvent passés par l'université, mais aussi à travers une formation *in situ* dispensée au sein des entreprises d'annotations.

Dans ces sections, on a également pu constater que l'annotation implique une variété de postes de travail, certains avec plus de responsabilités et dont le rôle est essentiel dans la gestion de la qualité des annotations. Le rôle de simple annotateur est également plus complexe que le seul sens commun qui permet d'identifier un chat d'un chien en quelques secondes. Certains projets sont plus compliqués que d'autres et nécessitent alors une littératie en français comme en informatique et parfois même une habileté technique. Au travers de l'organisation du travail qui est mise en place, ainsi que de l'acquisition des compétences, on s'aperçoit davantage ce qu'implique d'annoter des données avec des impératifs de qualité et de quantité.

Ces objectifs de production dépendent très largement de la capacité des annotateurs à voir le plus correctement possible et sans biais ce que contiennent les données. Cette pratique implique un travail d'interprétation des consignes et de ce que les données comportent, surtout lorsqu'apparaissent des données imprévues. Par ailleurs, tout au long de cette section, nous avons pu constater la variété des projets et des implications des travailleurs des données. L'analyse des pratiques du travail d'annotation montre, dans ce contexte, que le développement des modèles d'intelligence artificielle s'appuie sur une force de travail dont les qualifications se solidifient.

Sur cette base, on peut donc légitimement se questionner sur le rôle réel de ces travailleurs, au regard de la chaîne de production des modèles d'IA. On peut encore s'interroger sur la manière de conceptualiser cette activité. Selon nous, le travail des annotateurs intègre, en plus de la classification des données, un travail de problématisation des modèles d'IA. C'est à ce moment que le travail séminal de (Bowker & Star, 1999) peut nous permettre de comprendre le caractère véritablement essentiel de l'annotation de données et qui consiste à compléter le système de classification en identifiant les éléments auparavant *"laissés dans l'ombre"*. Notre hypothèse est que, ce faisant, ils participent à cette *"succession de pratiques collectives visant à définir les termes d'un problème à résoudre"*, autrement dit à la problématisation des modèles d'IA (Jaton, 2017).

7.2.3 Classifier et problématiser : quand le travail des données devient un travail de l'IA

Dans notre précédente partie, nous avons analysé le processus de construction des modèles. On y a constaté les nombreux allers et retours entre le travail de modélisation et de production de données. De ce point de

vue, l'industrie de l'IA se caractérise par son fonctionnement *incrémental*. L'instabilité des processus de production, des modèles d'IA et donc des méthodes d'annotations est une caractéristique majeure de cette industrie. Contrairement à Sambasivan et Veeraraghavan (2022), nous pensons donc que l'IA n'aboutira pas à la disparition des travailleurs spécialisés et par conséquent à une incorporation complète du travail dans les modèles, mais plutôt à une redistribution du travail expert tout au long de la chaîne de valeur. Dans cette section, nous approfondissons cette hypothèse, à travers une réflexion sur le lien entre le fonctionnement de cette industrie et la distribution du travail cognitif de classification et de problématisation le long de la chaîne de valeur.

Nous souhaitons montrer que ces travailleurs participent à l'essence même de la construction des modèles d'IA, ce que nous appelons le *"travail de problématisation"* (CALLON, 1986; Jaton, 2017). Derrière ce terme que nous empruntons à la sociologie des sciences, il s'agit d'analyser les pratiques collectives de définition du traitement des problèmes abordés par les modèles. Autrement dit, lorsqu'une entreprise souhaite créer un modèle pour reconnaître des aliments dans un plateau-repas, comment le problème est-il construit par les acteurs qui participent à sa fabrication? Dans cette partie, nous montrons justement que ces pratiques collectives interviennent à la fois tout au long du projet et tout au long de la chaîne.

Pour expliquer plus en détail ce type de travail, reprenons les deux études de cas déjà abordées dans notre précédent chapitre : sur la détection automatique d'aliments et sur l'identification des vols dans les supermarchés. Ces derniers proviennent de deux entreprises d'IA différentes, mais sont exécutés par le même sous-traitant à Madagascar. C'est au cours d'un entretien avec Jean, conseiller de la dirigeante qu'il précise ses fonctions et qu'apparaît plus nettement la fonction de *"remontée des annotateurs sur les modèles"*. Ainsi, il explique être en contact régulier avec les clients *"parce qu'il y a toujours des changements (...), des mises à jour sur la plateforme, et de notre côté, on fait de notre mieux pour y participer, pas seulement en tant qu'exécutant, mais en tant que participant"* (Jean - Conseiller RH - Sous-traitant A). Ces remontées d'informations concernent, par exemple les labels, c'est-à-dire des éléments nouveaux qui apparaissent dans les projets d'annotations : *"Par exemple, il y a des clients qui volent dans les supermarchés et au lieu d'utiliser les sacs, ils cachent les produits dans leurs vêtements. Il y a une dernière mise à jour qui a nommé cela « body theft »"* (Jean - Conseiller RH - Sous-traitant A - Projet AIView). Conscient du caractère crucial des procédures de remontée d'informations, il a proposé de faciliter ce travail en créant une *"plateforme de communication"* avec les clients intégrant des fichiers de questions, les guides d'annotations, et qui permettrait de faciliter le travail de documentation des évolutions de l'annotation²⁴⁸. Sur caisse de cantine automatique, Yanni, chef

248. "R : (...) Justement, on a proposé une plateforme de communication pour les remontées des agents, comme ça il n'y a pas d'écarts entre les informations communiquées et les propositions. (...) il y a déjà des sous-entendus et des non-dits. Avec la proposition de plateforme de communication, on peut créer des fichiers auxquels tout le monde a accès et ça permet de faire des améliorations à partir des informations communiquées (...) On a déjà utilisé Slack pour les communications entre les clients et les collaborateurs, mais actuellement, on va développer une plateforme indépendante (...)" (Jean - Conseiller RH - Sous-traitant A)

de projet, évoque l'exemple de la reconnaissance d'assiettes qui n'étaient pas décrites dans les consignes. En faisant remonter les éléments non prévus, l'annotateur participe à formaliser de nouvelles problématiques qui vont par la suite être pris en charge par le modèle.

Encadré n°56 : Les remontées d'informations sur le modèle de caisse automatique de cantine

R : « il y a des images (...) qui n'étaient pas dans des tags auparavant. Alors, on envoie des messages au client : « est-ce que vous avez une idée de cette assiette ? ». Et le client va répondre et nous dire de créer le tag par rapport à cette assiette. Donc on crée le nom du tag de la nouvelle assiette. Le client a demandé de le notifier avant de créer un nouveau tag, parce qu'on n'avait plus l'autorisation de créer de tags (...) L'année dernière, on a fait beaucoup d'images qu'on ne savait pas où transférer, parce que c'était comme une sorte de lasagne ou de viande, et on ne savait vraiment pas quelle était ce genre d'image. On avait longuement discuté et on a fait une capture d'image pour l'envoyer au client. Même le client a dit que c'était un peu difficile et de laisser ces images non traitées pour le moment et qu'ils allaient y réfléchir avant de les transférer dans des tags. »

Q : « Et finalement qu'est-ce qui a été décidé ? Ils ont trouvé une solution ? »

R : « Ils ont créé un tag, mais je ne me souviens pas quel était le nom du tag, et on avait transféré l'image dans ce nouveau tag. »

Yanni - Chef de projet - Sous-traitant A

Les deux études de cas illustrent le principe de problématisation des modèles. En corollaire, on a constaté que le travail sur les données implique de nombreux acteurs à différents niveaux de la chaîne, qui contribuent tous à l'IA : employés des entreprises d'IA, les entreprises d'annotation, les clients et les consommateurs. Chaque projet nécessite ainsi un travail d'annotation de la part de plusieurs types d'employés de l'entreprise d'annotation à Madagascar. À une extrémité de la chaîne d'approvisionnement en données, les annotateurs ordinaires utilisent des étiquettes de base (par exemple, "vol", "plat") pour décrire grossièrement des éléments dans des images et des vidéos. Ensuite, des équipes de qualité composées de vérificateurs parfois appelés "agents experts" examinent les annotations. Les annotateurs experts examinent les images difficiles à interpréter et certains se spécialisent dans des types particuliers d'objets ou de situations. Par exemple, certains plats pour les projets alimentaires, certains types de vols à l'étalage pour les projets de surveillance. La communication entre les équipes se fait via une plateforme commerciale en ligne, Slack, qui sert d'espace de travail à distance. Au sein des équipes, la communication entre les annotateurs, les experts et la direction (chefs d'équipe et responsables de production) se fait localement et en personne.

La partie française de la chaîne de valeur formalise également des processus de production d'annotations.

Les deux startups d'IA que nous avons étudiées disposent de leur propre logiciel d'annotation en ligne qui leur permet de contrôler les performances et la qualité des annotateurs. Un "gestionnaire de jeux de données" supervise la qualité des annotations et communique avec les annotateurs au sujet de leur contribution. Deuxième catégorie d'employés français, les "architectes de solutions" sont chargés de concevoir des modèles et de servir d'interface entre le client final, la startup d'IA et la société d'annotation. Ils sont chargés de rédiger des guides d'annotation, en collaboration avec le gestionnaire du jeu de données.

C'est au cours du démarrage des projets que le travail de problématisation est le plus important, notamment lors des phases de test. Au cours de l'annotation d'un premier ensemble de données, des chefs de projet, ou dans certains cas des employés du département de prototypage se rendent compte de ce qu'il manque dans le cahier des charges du client et des ambiguïtés potentielles. Dans plusieurs entreprises, les employés en charge de ces tests indiquent que les clients n'arrivent pas nécessairement avec un cadrage suffisant des projets ²⁴⁹. Dans ces situations, c'est alors aux testeurs de projets (annotateurs et managers) de faire des remontées sur les problèmes potentiels : "Q : donc ils ont pas du tout réfléchi aux catégories qu'il pourrait y avoir ... ? R : ouais, ou aux cas particuliers (...) alors on fait des propositions quand on nous demande oui, quand on voit que les cas particuliers ne sont pas gérés, on met le holà, on leur dit « tel cas on est pas capable de le faire, si vous ne donnez pas de consigne, nous ce qu'on peut faire c'est ... »." (Quentin - Directeur du centre - Sous-traitant D). L'encadré qui suit nous raconte également la nécessité de faire remonter les problèmes liés aux consignes, lors de la phase de test, dans le contexte d'un projet de détection des déchets.

Encadré n°57 : L'annotation du projet de trie des déchets

M : Quand on voit les images, on voit qu'il y a beaucoup de catégories qui ne sont pas dans les specs par exemple. Et là, on les remonte aux clients, et ils ajoutent ça dans leurs documents internes (...). Sur [ce projet] dès le départ on ne pouvait pas vraiment limiter ce qu'il y aura dedans ou ce qu'il n'y aura pas dedans. Il manquait beaucoup de choses en fait - je ne me souviens pas exactement de ce que c'était, mais il manquait beaucoup d'objets. Parce qu'il fallait vraiment séparer les objets transparents, les objets opaques et les objets teintés, les papiers, les sachets en plastique, les sachets étirables, les sachets non étirables ... ! il manquait pas mal de choses quand même (...)

Mety - Chef de projet - Sous-traitant B

Les difficultés peuvent également émerger en cours de projet, lorsque les annotateurs identifient des éléments qu'il faut ou qu'il ne faut plus annoter. Ce fut le cas sur un projet d'annotation pour la détection

249. Martin, responsable d'un plateau de production chez Sous-traitant D : "Ouais ... il y a un client sur 2 qui vient avec un cahier des charges plutôt complet. Souvent de toute façon on fait quand même des retours, il y a quand même des allers-retours, parce qu'il y a toujours certains cas auxquels la personne n'a pas pensé et ça c'est normal." (Quentin - Directeur du centre - Sous-traitant D)

d'éléments liés à l'installation de la fibre optique (encadre n°2).

Encadré n°58 : L'annotation du projet "fibre optique"

Z : vous avez un exemple de quelque chose sur lequel vous avez tous discuté ensemble, et que vous avez fait remonter après à votre team leader qui a fait remonter à *Automatik*, et qui a fait changer le guide ?

H : il y a eu des modifications ... dans notre métier, on cadre quelques photos par exemple. et une fois qu'on a annoté les photos de cadrage, on devait ... Si dans notre annotation notre guide au départ c'était d'annoter toute cette image, une fois le projet en cours, on a reconnu que cette valeur n'était pas si utile, donc on se devait de ne pas le prendre, et de prendre juste ce qui était ... donc on a parlé à nos supérieurs « on ne veut pas prendre celui-ci parce que ceci va erroner notre annotation, donc on doit rectifier tout ce qui avait été déjà ... ».

Z : ensuite le guide a changé ?

H : oui ensuite ce n'était pas le guide qui a changé, mais toutes les consignes de départ

Z : c'était quoi concrètement la photo ? qu'est-ce que c'était l'annotation ?

H : la fibre et tout ça, parce que on était pas tous techniciens [fibre optique], on ne savait même pas ce qui était les jarretières et autres ... donc c'est à eux de nous donner la lumière sur leur projet, et ensuite on doit connaître ce que l'on pouvait faire sur cette image. et puis une fois qu'on a fait ce que l'on devait faire, à nous de vérifier ce qui était important ou pas, et eux de dire « c'est comme ça, vous devez annoter ceci, vous devez annoter cela, arrêtez de prendre cela maintenant parce que ça n'est plus assez utile ».

Hejo - Annotatrice - Sous-traitant A

Enfin, c'est aussi l'environnement qui évolue. Ceci requiert alors une implication des annotateurs pour identifier les données imprévues. Les *data scientists* utilisent le terme de "*data drift*" pour décrire ce phénomène. En français, ce concept désigne la "*dérive conceptuelle*", soit l'éloignement entre le modèle et les données qui remontent du terrain à un instant T, au regard des jeux de données d'entraînement. Il peut en effet se dérouler sur une période plus ou moins longue entre le démarrage d'un projet d'IA et son déploiement en production. Dans l'extrait d'entretien qui suit, Victor décrit justement les changements de l'environnement dans lequel est baigné le modèle. Dans le cas qu'il décrit, il s'agit par exemple de cigarettes électroniques qui se mettent à apparaître sur les plateaux-repas.

Encadré n°59 : Les évolutions de la vérité terrain sur le projet de caisse de cantine automatique

C'était sur une période de temps assez longue et l'environnement évolue. À partir d'un moment, il y a la cigarette électronique qui est arrivée sur le marché et ils se sont retrouvés avec plein d'images de cigarettes électroniques sur le plateau. Et donc ça, ça a généré un biais. Les équipes d'annotateurs ont été capables de dire « Tiens, on voit qu'il y a une récurrence qui se crée sur cet objet c'est-à-dire que ça apparaît de plus en plus souvent et ça ne correspond pas à ce qu'on est censé annoter ; on comprend que ça peut générer un biais parce qu'on a cette notion de contexte et ce qu'est l'IA et le projet, donc on va en parler au client. » Du coup, on montre des exemples au client et le client dit « On ne l'avait pas anticipé » parce que les images de cigarettes électroniques étaient noyées dans la masse de toutes les données, et en fait ça aurait pu générer un biais, donc c'était important pour eux qu'on leur remonte cette situation par exemple. »

Victor - CTO - Sous-traitant B

Au cours de cette section, nous avons vu plusieurs exemples qui témoignent des évolutions des processus de modélisation. Pour nous, la réalité du travail de l'IA est d'identifier cette *"dérive conceptuelle"* qui se situe dans l'environnement et donc dans les données et qui n'est donc pas prévue lors de la phase initiale de *problématisation du modèle*. Les annotateurs sont des travailleurs de l'IA et non uniquement des données, car ils remédient à cette dérive conceptuelle, en contribuant à réduire l'écart entre le modèle et les données du terrain. Pour eux, cela passe par un travail d'identification des éléments imprévus : *"Oui [ça nous arrive de faire remonter] en fait à la base (...) on avait toujours un fichier question par rapport au projet, et dans ce fichier, il y avait les supérieurs, les annotateurs et aussi le client. Et lorsqu'on avait des questions ou qu'on trouvait des nouvelles choses on le mettait dans ce document."* (Soly - Annotateur - Sous-traitant G). L'une des principales conclusions de cette section est que les pratiques de *classification* et de *problématisation* des données s'inscrivent dans plusieurs étapes de micro et macro-décisions qui façonnent en fin de compte le service d'IA et sa concrétisation. Le travail peut être informel (une intuition sur la manière dont certains objets sont identifiés) ou, au contraire, assez formel : compilation et mise à jour de guides d'annotation, détection d'anomalies dans les données, traduction d'instructions dans des langues étrangères, affinage des paramètres, interaction avec les clients, validation des résultats. Cette dimension du travail de l'IA concerne de nombreux travailleurs à différents niveaux de la chaîne de valeur qui définissent et redéfinissent la manière dont les ensembles de données sont structurés et la manière dont les modèles d'apprentissage automatique fonctionnent. Par conséquent, nous considérons qu'il existe des frontières très poreuses entre la modélisation (le *travail d'IA* supposément réservé aux startups en France) et le travail d'annotation (le *travail de données* externalisé à Madagascar).

À des degrés divers, les annotateurs effectuent donc un travail cognitif qui aligne les modèles sur les connaissances spécifiques au domaine. Finalement, au-delà des tâches les plus basiques d'annotation, le rôle de ces travailleurs de l'IA est d'éviter la *"dérive conceptuelle"* des modèles. À l'inverse, les ingénieurs des startups,

les clients et même les consommateurs effectuent un travail sur les données aux côtés des annotateurs. Ce travail distribué va au-delà de l'annotation, c'est-à-dire de l'encerclement d'éléments sur des images. En tant qu'infrastructure numérique, l'IA coordonne et intègre le travail humain dans des processus complexes de classification et de problématisation à chaque étape.

* * *

Les entreprises françaises d'IA utilisent des architectures d'IA pré entraînés sur la base desquelles ils produisent des modèles appliqués. Ce travail de personnalisation rend les tâches liées aux données fondamentales pour le travail de l'IA, et elles vont au-delà de la simple annotation des données. À Madagascar, elles comprennent ainsi la participation à la rédaction d'instructions d'annotation, l'examen du comportement des modèles et la discussion avec les clients finaux. Le client final participe également à fournir un retour à propos de ses besoins ou à apporter des corrections au modèle.

En France, les employés des entreprises d'IA supervisent tout ce travail d'inscription des représentations dans leurs modèles qui passent par tout un ensemble d'acteurs en plus d'eux-mêmes. Les consommateurs qui utilisent ces solutions produisent des données, les clients qui déploient le service participent à affiner les concepts qui s'appliquent à la solution. Enfin, les annotateurs s'occupent de raffiner les données pour les rendre exploitables et font remonter les "outliers". Ainsi tous ces acteurs partagent la charge de travail cognitive qui consiste à examiner les données (*classification*) et à analyser la manière dont elles seront interprétées par le modèle (*problématisation*). Loin d'être le signal d'une disparition du travail, nous considérons l'IA comme le prolongement du numérique comme outil logistique qui facilite et coordonne divers processus de travail liés à ces deux fonctions. Nous contribuons ainsi à la littérature qui tente de comprendre l'activité humaine derrière les infrastructures numériques (Muniesa, 2003). Ce travail s'inscrit également dans la lignée des travaux de Jaton (2017) sur les AI-labs ainsi que de sa définition de la modélisation en tant que processus de problématisation.

Sur la base de la typologie du travail sur les données de Tubaro, Casilli, et Coville (2020a), nous pouvons comprendre comment les travailleurs des données contribuent à l'IA. En décrivant les différentes tâches qui sous-tendent ce travail sur les données et l'IA, nous complétons cette typologie. Les exemples cités ci-dessus illustrent clairement la variété des différentes tâches impliquées dans le travail d'IA et le nombre de travailleurs différents qui y participent. Les tâches combinent souvent la classification des données (que voyons-nous dans les données?) et de la problématisation des modèles (quel problème essayons-nous de résoudre?) qui nécessitent un travail d'interprétation qui suggère un travail plus qualifié que les discours qui accompagnent son développement. Enfin, nous considérons que le développement de ces pratiques au niveau organisationnelle et individuelle suggère aussi une "solidification des pratiques professionnelles".

Conclusion de la partie 3 - La professionnalisation du travail de l'IA : un travail de stabilisation des modèles, une explosion du travail lié aux systèmes d'IA

Dans ce chapitre, nous avons abordé le travail de l'IA en replaçant cet objet dans la controverse sur la déqualification. L'économie et la sociologie du travail questionnent depuis longtemps les impacts de l'automatisation sur le travail. Face à l'hypothèse de la disparition du travail et celle de sa déqualification, la recherche a proposé plusieurs pistes d'analyses plus nuancées. Notre travail sur les services automatisés commerciaux témoigne de plusieurs phénomènes. D'abord, nos résultats suggèrent non pas une disparition, mais bien un déplacement géographique, social et substantiel du travail. La structure de la chaîne de valeur reflète effectivement une polarisation des emplois, mais pas nécessairement des qualifications.

Un travail collectif de constitution de savoirs qui servent à optimiser les modèles d'IA

À cet égard, les résultats de cette partie s'inscrivent dans la continuité des éléments présentés précédemment sur l'intégration des travailleurs des données dans les processus mondiaux de production d'IA. Cette intégration se traduit par la mise en place de processus de coordination. On retrouve ici un élément propre à la critique de [Zarifian \(1983\)](#) de la déqualification du travail industriel dans le contexte de l'existence d'un « savoir collectif » indispensable à la réalisation de la production.

Ce travail relationnel repose sur l'utilisation des nouvelles technologies de communication et nécessite une compréhension de l'ensemble du processus de production. Cet aspect du travail est observé dans les industries flexibles. La nature dynamique et adaptable de l'industrie au cours de cette période découle de sa capacité d'ajustement et d'adaptation rapide aux changements de la production. Ces connaissances ne peuvent pas être transférées aux machines, mais elles sont essentielles au processus de production. Il s'agit, entre autres, des compétences implicites requises pour contrôler la production par le biais de tâches de surveillance et de programmation de machines. Ce sont tous des domaines où l'automatisation pourrait conduire à une requalification des travailleurs (*reconfiguration de la main-d'œuvre des travailleurs*) et, in fine, à une *redistribution* des compétences.

Le rapport aux qualifications met alors en jeu le positionnement des acteurs par rapport à la constitution des savoirs : *"la question d'une recombinaison collective du savoir apparaît selon une modalité qui n'oppose pas tant "savoir abstrait" à "savoir concret" que séparation à coopération par rapport au savoir nouveau qui se constitue."* ([Zarifian, 1983](#)). Notre travail illustre des mécanismes de production de savoirs par la coopération des différentes strates de la chaîne de valeur. Du point de vue de la sociologie de l'innovation, Renou (2006) parle « d'espace de discussion ». Dans le contexte de processus de conception distribuée, il évoque

l'émergence de mécanismes d'échanges permettant de résoudre des problèmes ou d'interpréter les tâches à réaliser. Nous montrons que la chaîne de valeur fonctionne comme "*espace de conventionnement*" qui permet de concevoir l'IA, en tenant compte des éventuelles "dérives conceptuelles" du modèle. Au plus haut degré des compétences requises, les annotateurs s'engagent alors dans un travail d'interprétation, d'autant plus qu'ils sont soumis à de fortes pressions managériales comme on le verra dans notre dernière partie. Ainsi ce travail de l'IA constitue en fait une articulation de plusieurs strates de travail humain, avec des activités parfois très intenses qui impliquent des tâches non automatisables et non dépourvues de qualification. Les entreprises d'IA ne vendent donc pas réellement un modèle, mais la gestion d'un "*espace de conventionnement*" qui permet d'essayer de mitiger la "*dérive conceptuelle*" des algorithmes. Des qualifications supplémentaires, relationnelles et communicationnelles liées à l'externalisation de cette fonction et non seulement de tâches "simples", sont alors nécessaires.

Notre chapitre montre clairement la variété des tâches et des travailleurs impliqués dans le travail de l'IA pour en assurer le bon fonctionnement. Les tâches sont diverses et combinent souvent la classification des données (« Que voyez-vous dans vos données? ») et la résolution de problèmes de modèle (« Quel problème essayez-vous de résoudre? »). L'une des principales conclusions de cette section est que la pratique consistant à classer et à problématiser les données fait partie intégrante de plusieurs étapes au cours desquelles les travailleurs de l'IA sur toute la chaîne prennent des micro- et macro-décisions qui façonnent finalement les services et leurs réalités. C'est précisément d'ailleurs la non-prédictibilité du travail de l'IA qui constitue un facteur de requalification.

De ce point de vue, comme l'expliquent De Terssac et Coriat, concevoir « *les tâches requises comme s'imposant aux ouvriers d'elles-mêmes directement et avec un contenu d'avance et nécessairement fixé, revient à nier l'activité réelle et laisse supposer l'existence d'une stabilité des éléments techniques et de l'activité ouvrière, qui ne résiste ni à l'observation, ni à l'analyse de ce que sont ces systèmes techniques en fonctionnement réel.* » En pratique, l'opérateur doit effectuer diverses tâches qui ne sont pas préprogrammées. La *maintenance* des modèles d'IA nécessite par exemple des compétences et des efforts particuliers tels que la détection d'anomalies, la stabilisation des processus, l'optimisation de la qualité et de la quantité des produits. Les annotateurs se retrouvent à réaliser des tâches qui impliquent des savoirs tacites, par définition difficiles à automatiser : "*the tasks that have proved most vexing to automate are those demanding flexibility, judgment and common sense—skills that we understand only tacitly*" (Autor, 2014 (p. 136))

Une chaîne de coordination du travail cognitif humain de problématisation et de classification

Sous cet angle, notre objet questionne également la substance même des modèles d'IA. Nous observons que ce qui distingue la chaîne production de l'IA, c'est la contribution d'une variété importante de consommateurs-producteurs et de producteurs de la donnée (travailleurs de l'IA). La donnée est littéralement produite par

les annotateurs et à travers l'usage des modèles d'IA qui contribue à en générer de nouvelles. Elle est aussi conceptualisée : data scientist, annotateurs et leurs managers, clients finaux ce qui confirme une fois de plus la véracité de l'adage issu de la sociologie de la quantification : "raw data is an oxymoron" (Desrosières, 2010; Gitelman, 2013; Goeta, 2016) .

Nous assistons donc à un quadruple déplacement du travail : travail gratuit des clients, sur-travail des données de la part des employés du client final, déplacement du contenu des postes chez le client final et déplacement de l'activité productive du service et de l'IA hors des frontières géographiques. L'ensemble de ces données, produites par des travailleurs qui s'ignorent les uns les autres, s'inscrit alors dans le dispositif d'automatisation. La production de services d'IA constitue donc le paroxysme de tendances observables à une autre échelle : l'externalisation et la mondialisation, le travail gratuit des clients, la dataification et son corollaire la travail des données. De ce point de vue, l'IA la répartition du travail, non pas en supprimant des occupations, mais en créant du travail là où il coûte le moins cher.

Dans cette troisième partie, nous avons également montré comment le travail de modélisation et le travail de production et d'annotation de données s'articulent. Ainsi, la production de modélisation ne répond pas nécessairement à des données existantes mais surtout aux possibilités d'optimisation de certaines activités économiques. Ce faisant, la modélisation crée une nécessité de produire et d'annoter des données. Intrinsèquement, le travail de modélisation répond donc, non pas à des perspectives de généralisation, mais plutôt à des cas d'usages très précis qui nécessitent donc autant de données représentant le monde réel.

Cette *vérité de terrain*, la nécessité de coller à la réalité a des implications importantes sur le travail des données. Le schéma ci-dessous décrit ainsi les différents acteurs impliqués dans la construction des modèles d'IA. Ainsi, dans les deux cas observés ici, les utilisateurs finaux produisent eux-mêmes des données réelles et peuvent corriger le résultat des systèmes d'IA. Les travailleurs des entreprises clientes des start-ups d'IA, par exemple les caissiers, vont également faire des retours sur la manière dont fonctionne l'algorithme. Puis, les managers transmettent également leurs à l'entreprise d'IA. Ces modifications vont ensuite avoir un impact sur la force de travail d'annotation. De ce point de vue, l'automatisation constitue une technologie de coordination des forces de travail pour la production, l'annotation et la modification des données. De cette manière, plutôt que d'inventer des systèmes "labourless", les start-ups d'IA proposent en fait des systèmes qui reposent sur la superposition de nombreuses *boucle de rétroaction humaines*.

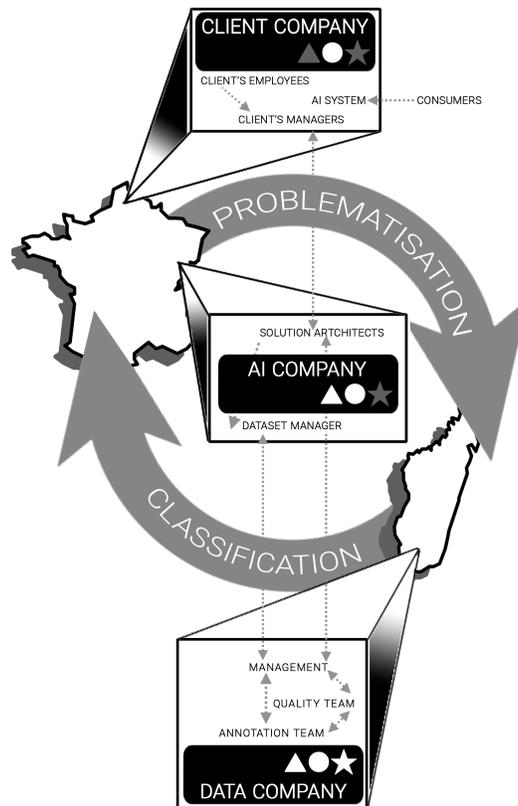


FIGURE 32 – Data work production chain. Source : Authors' elaboration. Source : Le Ludec, Cornet et Casilli, 2023.

Par conséquent, ce qui est finalement vendu et déployé en tant que solution d'IA n'est pas l'algorithme fini, mais le processus de production lui-même. L'IA en tant que processus peut être considéré comme un dispositif sociotechnique qui automatise les tâches du dernier kilomètre. Il faut beaucoup d'efforts pour fournir des services qui utilisent l'IA. Par conséquent, nous soutenons que l'IA commerciale n'automatise pas les opérations de service, mais réplique les opérations pour permettre des changements et étendre la chaîne d'externalisation. De ce point de vue, la production d'IA est principalement motivée par des compromis économiques basés sur la minimisation du coût du travail.

Les débats sur la définition de l'automatisation dans le secteur industriel ont été nombreux dans les années 60. Dans un article, William [Grossin \(1960\)](#) distingue une vision techniciste où l'automation consiste à intégrer des processus de production morcelés par la division du travail, dans une chaîne de production intégrée et automatisée qui peut se corriger elle-même, avec des machines "programmables" disposant de leur propre énergie, d'une vision sociologique analysée à l'aune du degré de recours à la main humaine. De ce point de vue, est automatisé le produit qui parcourt la chaîne de production "sans être touché par la main de l'homme".

Un travail socialisé qui dénote avec l’imaginaire d’une activité déqualifiée

Initialement, la sociologie industrielle a notamment identifié des risques de déqualification pour la plus grande partie des ouvriers, tandis qu’un petit nombre peut progresser dans la hiérarchie salariale (*polarisation des compétences*). Cette évolution de la distribution du travail a également été démontrée dans le contexte de l’externalisation de services numériques. La demande de travail dans le domaine de l’informatique a par exemple été partiellement comblée par des travailleurs de pays du *Global South*. Dans ce cas, ou dans celui du travail féminin étudié par Maruani et Nicole-Drancourt (1989), on a pu constater des mécanismes de minimisation de la réalité du travail effectué (*”mythe du travail facile”*). Finalement, non seulement les annotateurs voient leurs contributions invisibilisées (Irani, 2015a), mais en plus, y compris auprès des donneurs d’ordres finaux, l’intensité de celles-ci peut être minimisée. La réponse, d’Arthur, chargé de communication chez CityVeillance, une entreprise spécialisée dans les caméras intelligentes, illustre bien cette tension.

on en parle pas forcément, car c’est de la sous-traitance, tu mets pas en avant l’imprimeur qui fait les plaquettes pour de la communication. Ils font un travail rébarbatif, de seconde main (...) C’est un taff hyper chiant (...) c’est un métier de l’usine. Bien sûr, sans ça on fait pas d’IA, le constructeur de smartphone [ne] va jamais mettre en avant les composants qui viennent de mines en Afrique, car c’est de la communication, tu n’as pas envie de parler de ces éléments à tes clients. Si tu leur parle de travailleurs malgaches, ça peut le faire fuir. C’est un processus tellement infime dans un projet en IA et le client il s’en fout, il n’a pas besoin de savoir tout ça.

Arthur - Chargé de communication CityVeillance

Sans ces travailleurs, il n’y a donc pas d’IA; pourtant, la répétitivité du travail expliquerait qu’il n’ont pas vocation à être valorisés. Ce processus n’est pourtant pas si ”infime”. On a notamment pu constater la contribution parfois de plus haut niveau des annotateurs tout au long de la production des services basés sur l’IA. Plus le projet d’IA est complexe et appliqué au monde réel, plus cette contribution sera importante et plus les tâches seront réalisées par la main humaine, et non par la machine. Certaines pratiques ”d’élargissement du travail” montrent d’ailleurs qu’il est possible d’organiser différemment le travail, à la fois pour valoriser l’expérience des travailleurs - certaines entreprises permettent tout de même des évolutions hiérarchiques et salariales - et pour mitiger le caractère répétitif du travail. Cela nous interroge sur la différence entre les qualités qui permettent d’accomplir ”réellement” le travail et les qualités supposées, définies par l’employeur. La mesure de la qualification d’un poste de travail reste en effet marquée par une tension forte²⁵⁰, car le processus de standardisation des qualifications est affecté par les hiérarchies sociales du milieu professionnel dans lequel exercent les travailleurs.

250. Largement ravivée durant le Covid19 avec la mise en lumière de certaines activités essentielles mais pourtant sous-évaluées.

Notre étude permet de présenter une facette très différente de la littérature sur les plateformes numériques de travail (PNT), notre étude présente une facette très différente. Les entreprises occidentales ne sous-traitent pas des tâches, mais des processus de production. Cette façon d'organiser la production implique donc un certain nombre de postes de travail plus qualifiés. Les attentes des donneur d'ordres concernent aussi la capacité à communiquer des annotateurs. Nous montrons d'ailleurs des mécanismes d'apprentissages similaires à ceux observés par Margaryan (2019) à propos d'annotateurs présents sur des PNT. Nos résultats suggèrent finalement l'existence de postes de travail plus qualifiés dans l'équipe de direction présente à Madagascar et une requalification des tâches simples de saisie de données (*professionnalisation*).

Du travail de la donnée au travail de l'IA : solidification d'un groupe professionnel aux qualifications encore trop invisibilisées

Pour approfondir cette dimension de notre travail, un détour par le concept de profession et la tension qui existe au regard de ce qu'on peut désigner comme un "groupe professionnel" s'impose. Le terme apparaît aux États-Unis et concerne à l'origine un nombre restreint d'activité et notamment les professions libérales. Cette perspective dépeint alors des groupes professionnels qui se constituent en se démarquant des "occupations" par un statut social plus valorisé. De ce point de vue, la notion de profession a longtemps fait corps avec les dispositions institutionnelles des pays étudiés, c'est le cas de l'acception américaine du terme qui apparaît dans un contexte de poussant des travailleurs de professions libérales à justifier de leur place dans la division du travail (Paraseise, 2003). Bourdoncle (1993) note ainsi que face à cette conception, la notion de profession s'établit en France, soit par une lutte politique pour la valorisation de certains métiers au sein d'une hiérarchie politico-administrative par le jeu des corps d'état, soit par l'agrégation d'individus au sein de communautés de pairs (confréries).

En se concentrant sur le contenu du travail réel, sur les activités qui lient des individus entre eux, la sociologie du travail a opéré un renversement théorique en dépassant un concept jugé trop restrictif par la notion de "groupes professionnels" qui sont *"des ensembles de travailleurs exerçant une activité ayant le même nom, et par conséquent dotés d'une visibilité sociale, bénéficiant d'une identification et d'une reconnaissance, occupant une place différenciée dans la division sociale du travail et caractérisés par une légitimité symbolique"* (Demazière et Gadéa, 2009, p.20). Un enjeu de l'analyse de la professionnalisation réside alors dans notre capacité à saisir les logiques à l'oeuvre au niveau des individus (comment se forment les individus? Comment apprennent-ils?), des activités (comment un groupe de pairs va formaliser les savoirs d'un groupe professionnel?) et des organisations (comment une entreprise va formaliser les compétences présentes au sein de son organisation?).

Nos résultats montrent que la structuration de la formation s'inscrit dans plusieurs logiques de professionnalisation (Wittorski, 2014), telles que la "logique de l'action" qui consiste pour individus confrontés à des

situations nouvelles à ajuster leurs actions ("compétence incorporée"). La "logique de la réflexion et de l'action" s'observe quant à elle dans la mobilisation de ressources documentaires et de ressources sociales dans l'accomplissement d'une tâche. La "logique de la réflexion sur l'action" évoque quant à elle les moments où l'individu prend de la distance avec son activité pour l'analyser et ainsi développer une connaissance sur son activité. La "logique de traduction culturelle par rapport à l'action" renvoie à l'intervention d'un tiers pour l'accompagnement des salariés dans l'accomplissement de leurs missions. Enfin la "logique de l'intégration assimilation" caractérise l'utilisation par les individus de ressources documentaires pour acquérir des savoirs ou des connaissances.

Nos résultats mettent donc en visibilité la continuité des pratiques de travail au sein d'un même secteur et les logiques de "professionnalisation" qui relève d'après Wittorski (2014), "*d'une intention organisationnelle*" et qui passe d'une part par la prescription des "bonnes compétences" à avoir, et d'autre part, par des dispositifs de formation ("identité prescrite"). Toutefois, il est également nécessaire de tenir compte du "développement professionnel" qui constitue un processus de transformation des individus "au fil de leur activité" ("identité agie et vécue"). Au travers, de l'interaction entre ces trois niveaux, se joue la "négociation" entre l'individu et l'organisation dans le processus de reconnaissance de ses compétences, notamment à "*partir de l'évaluation des résultats de l'activité qu'il a déployée*" (identité "reconnue").

Les résultats que nous présentons dans notre dernière et quatrième partie illustre cette tension entre individus et organisations autour de la question des qualifications. Nous y aborderons l'analyse de cette notion à travers l'identité professionnelle des travailleurs. En effet, s'il semble désormais clair que le corpus de connaissances et de savoirs nécessaires démontre un mouvement de *solidification d'un groupe professionnel* à l'intersection de deux secteurs complémentaires qui s'entrecroisent : l'externalisation de services numériques et l'industrie de l'automatisation. Nous y approfondissons la reconnaissance des qualifications des travailleurs des données au regard des questionnements liés à l'impact de l'industrie sur le développement. En effet, il est également essentiel de questionner la chaîne de valeur au regard du partage de la valeur et de sa dimension post-coloniale. Comme le précisent d'ailleurs [Graham et al. \(2017a\)](#) : "*Les recherches ultérieures sur les marchés du travail en ligne devraient se pencher sur des pays et des modèles spécifiques (...) par exemple, [en fonction] des connexions institutionnelles historiques entre colonisateurs et colonisés, et des connaissances tacites persistantes*" Le fait de tenir compte de cette caractéristique introduit d'ailleurs une variable interprétative importante : le contrôle managérial qui s'exerce sur des travailleurs situés dans des pays en voie de développement et qui constitue le facteur limitant de la requalification / professionnalisation des travailleurs malgaches.

Nos résultats montrent que ce qu'on désigne par le terme *automation* correspond en fait à la nécessité des systèmes d'IA d'être basés sur une variété de forces de travail, dont ils sont un dispositif de coordination et qui permettent de faire "vivre" ces derniers, de les concrétiser, de les faire évoluer, voire de les réparer

ou même de s'y substituer. Dès lors, rendre visible l'implication de ces travailleurs, c'est questionner des chaînes de production mondialisées, bien connues dans l'industrie manufacturière, mais qui existent aussi dans le secteur du numérique. Ces travailleurs étant nécessaires à son fonctionnement, ils sont les rouages invisibles de nos vies numériques. C'est aussi rendre visible les conséquences de leur travail sur les modèles. Une partie des biais algorithmiques réside en effet dans le travail des données, pourtant encore largement invisibilisé par les entreprises. Une IA réellement éthique doit donc passer par une éthique du travail de l'IA.

Partie 4 - A l'autre bout de la chaîne, la précarité des travailleurs des données

Nous avons précédemment analysé la diversité des tâches réalisées pour suppléer l'automatisation algorithmique. Au travers du décalage entre le travail prescrit, le travail réel et sa perception par les différents acteurs de la chaîne de valeur, nous avons montré que les *travailleurs du clic* participent en réalité à un travail de l'IA. Un ensemble d'activités se coordonne ainsi autour des données et du modèle dont le résultat est un processus algorithmique sans cesse (re)travaillé à chaque bout de la chaîne, y compris par les consommateurs finaux. Ces résultats correspondent d'après nous à une dynamique d'élargissement du travail et de solidification d'un groupe professionnel : celui des travailleurs de l'IA.

Néanmoins, cette focalisation sur le processus de production présente la limite de ne pas suffisamment tenir compte du vécu des travailleurs. On peut notamment s'interroger sur la reconnaissance de leur apport dans le développement des modèles d'IA d'autant que les entretiens avec des employés de startups soulignent la tension entre l'importance de ce travail et sa répétitivité ce qui justifie alors d'en minimiser l'impact.

Cette dernière partie vient conclure notre travail en mettant en avant les travailleurs ainsi que leurs conditions de travail, et plus généralement l'impact de cette industrie sur les pays qui l'héberge. Ces éléments d'analyses s'inscrivent donc dans une tradition d'études portant sur les effets de la délocalisation d'activités économiques dans les pays en voie de développement²⁵¹. Le travail des données ne fait pas exception, puisque dès l'émergence de cette activité, chercheurs et praticiens des politiques publiques s'interrogent sur ses effets. Sandesh Sharanappa (2013) considère ainsi que le *"micro-travail et l'externalisation des processus d'entreprise (...) peuvent être utilisés comme un outil de développement"*. Porté par les expériences d'*impact sourcing* de Samasource qui associe le travail à la tâche et le développement professionnel de ses travailleurs à travers des formations, ce secteur émerge initialement avec l'image bienveillante d'une solution pertinente pour le développement au même titre que d'autres concepts comme la micro-finance²⁵².

Il convient également replacer l'analyse des effets de l'externalisation de tâches liées à l'IA dans le débat plus général sur les effets de l'intelligence artificielle concernant le travail. Nous proposons de reprendre un argument de (Benanav, 2020), qui considère que la véritable question de l'effet de l'IA sur l'emploi n'est pas la fin du travail, mais plutôt la qualité des activités alors créées. Ces questionnements ne sont pas uni-

251. Tout au long de cette partie, nous parlerons de *Sud global*, le concept de pays en voie de développement étant largement remis en question dans la littérature pour son ancrage dans la pensée occidentale

252. Malgré des critiques plus récentes. Ainsi, Esther Duflo et al. (2013), indiquent par exemple qu'ils constatent *"aucun changement dans les résultats de développement qui sont souvent considérés comme étant affectés par la microfinance, y compris la santé, l'éducation et l'autonomisation des femmes."*

quement corrélés à l'émergence de l'IA, mais s'inscrivent tout autant dans des dynamiques plus générales d'externalisation de la production de biens et services. Partant de nos précédents constats concernant la chaîne de valeur de l'IA, nous pensons qu'il convient également de tenir compte de la dynamique post-coloniale à l'oeuvre derrière cette industrie, c'est-à-dire de la manière dont les relations entre les acteurs se structurent : à travers le lien historique entre la France, ancien colonisateur, et Madagascar, ancienne colonie.

Dans cette partie, nous nous demandons comment le secteur de l'annotation s'est développé à Madagascar et dans quelles mesures ces facteurs de développement contribuent à la dimension extractive du secteur. En plus d'étudier la dimension post-coloniale de l'industrie de l'IA, nous répondons donc à plusieurs questions qui éclairent ce que cette l'IA crée comme activité professionnelle. Qui sont les travailleurs des données de Madagascar ? quelles sont leurs conditions de vie ? Qu'est-ce qui les amène à intégrer ce secteur ? Ensuite, quelles sont leurs conditions de travail ? Dans quelles mesures l'organisation du secteur de l'externalisation et l'industrie de l'IA contribuent à favoriser ou non la précarité des travailleurs et quelles sont alors les différentes facettes de cette précarité ? Enfin, quelles sont les perspectives professionnelles de ces travailleurs des données ?

Dans cette dernière partie, le recours à la littérature critique des chaînes globales de valeurs permet d'introduire la notion de "*paradoxe de l'amélioration*", qui désigne les difficultés des pays du Sud global à remonter la chaîne de valeur et à en capter les bénéfices. Par ailleurs, nous repartons de l'hypothèse d'une industrie "extractiviste" pour questionner à la fois la réalité de cette dynamique évoquée par Crawford (2021) et pour proposer d'autres pistes d'analyses détachées de l'ancrage du concept dans des industries de matières premières. Nous proposons d'analyser l'enjeu du développement à travers le concept de *précarité professionnelle des annotateurs* qui nous permet d'articuler trois dimensions essentielles liées à la qualité du travail : la reconnaissance des qualifications nécessaires à ce travail, la capacité à entrer sur le marché du travail à travers l'industrie de l'annotation de données et la variabilité des conditions de travail selon le degré d'informalité.

Dans notre chapitre huit, nous analysons les mécanismes d'émergence de l'industrie de l'annotation de données en insistant sur le rôle actif des politiques publiques. Nous verrons ainsi que la mise en place de cette organisation du travail est la synthèse de trois logiques : le développement de politiques volontaristes des pays du Sud global pour attirer des IDE (Heeks, 1996 ; Mitter & Pearson, 1992), la libéralisation du marché des services (1994) (Bensidoun & Ünal, 2007), et l'existence de proximités culturelles, linguistiques et économiques entre certains pays occidentaux et leurs anciennes colonies (Huws et al., 2004 ; Saxenian, 2002). L'objectif de ce chapitre est d'articuler la structure post-coloniale du secteur (niveau macro) avec la manière dont les travailleurs vivent et éventuellement subissent ces dynamiques (niveau micro). A travers le concept de "*valorisation coloniale*", nous analyserons le double processus par lequel les entreprises françaises

valorisent des compétences coloniales chez les annotateurs ("*rappel colonial*"), que ces derniers tentent d'activer pour trouver du travail ("*colonial assetization*")

Peu de travaux ont pour le moment étudié les effets de l'industrie des données sur les travailleurs et les pays qui les hébergent. La littérature sur les travailleurs des plateformes a en effet très largement documenté les conditions de travail de ces derniers en se focalisant sur leurs statuts d'indépendants et sur les risques inhérents à une activité inconstante. Alors que ce travail est perçu comme quelque chose que "*tout le monde peut faire*", on peut se demander dans quelle mesure ces emplois sont une voie d'entrée vers des carrières ascendantes dans le marché du travail. A l'intersection entre l'étude des impacts des investissements étrangers dans les pays du Sud global et la sociologie du travail, notre chapitre neuf adresse les effets du micro-travail sur les carrières professionnelles à travers plusieurs interrogations : comment devient-on annotateur ? Dans quelles dynamiques professionnelles s'inscrit la participation à l'industrie de l'intelligence artificielle ? Nous analyserons finalement les impacts sur les travailleurs des pays concernés et concerne la structure de notre population d'étude et sa capacité à tirer de cette activité un salaire assurant une forme d'émancipation économique.

Chapitre 8 - La dynamique post-coloniale de l'extractivisme des données : l'attraction des investisseurs par l'existence de qualifications coloniales

La littérature sur le micro-travail a souligné la distribution de micro-tâches au sein de chaînes de valeurs globales pour les pays anglophones. Sa lacune principale est relative au manque d'études empiriques portant sur les relations post-coloniales entre donneurs d'ordres et travailleurs des données. Pourtant, comme l'explique (Kano et al., 2020), la "*localisation est (...) une des variables clés [des] chaînes de valeur mondiales, mais peu d'études empiriques (...) tentent de mesurer la dispersion géographique des chaînes de valeur étudiées, afin de déterminer si la portée de ces chaînes de valeur est en fait mondiale*". Cette remarque sur les études des chaînes globales de valeur invite selon nous à prendre de la distance sur la dimension mondiale du phénomène d'externalisation des activités numériques en général, et des données en particulier. Dit autrement, parler de mondialisation laisse penser que le "*monde est plat*", au sens où l'externalisation peut se faire aisément de n'importe où. Au contraire, nous pensons que deux variables sont sous-évaluées dans cette littérature : les politiques volontaristes des États en voie de développement et les dynamiques post-coloniales préexistantes à l'externalisation de services liés aux données. Le rôle de l'État a été souligné dans le cas de l'Inde, notamment par la mise en place d'une politique orientée vers l'exportation de services numériques (Mitter et Pearson (1992). Heeks (1996) montre notamment que les politiques d'incitations à l'exportation ont conduit les firmes indiennes de services à chercher des gains à court terme à l'exportation et à délaisser

le marché intérieur les rendant particulièrement vulnérables. Cette politique orientée vers l'exportation s'est traduite par des dévaluations monétaires ainsi que par des crédits impôts pour les sociétés de logiciels dans le cadre de la mise en place de centres d'exportations²⁵³.

Kleibert et Mann (2020) ont mené une étude de cas sur le développement du secteur des technologies de l'information dans trois pays du Sud global (Inde, Philippines et Kenya) afin d'étudier la pertinence des politiques industrielles centrée sur ces technologies pour le développement. Elles critiquent en particulier l'hypothèse selon laquelle tous les pays pourraient reproduire l'exemple positif de l'Inde, en s'appuyant sur trois arguments : la stabilité de la production de services numériques, le fait que leur géographie surtout est déterminée par les coûts de main-d'œuvre et d'infrastructures réseaux et donc que les nouveaux entrants quitteront le marché lorsque les coûts de main-d'œuvre augmenteront. Elles développent une typologie de trois modes d'articulation des entreprises des pays du Sud global avec leurs donneurs d'ordres. Le couplage *structurel* désigne l'externalisation des tâches peu qualifiées par des entreprises étrangères "*principalement motivée par [le] coût (de la main-d'œuvre)*". Le couplage *fonctionnel* est basé sur un "*rapport coût-capacité*", avec "*l'entrée des entreprises étrangères dans les réseaux de production mondiaux*". Enfin, le couplage *indigène* "*offre le plus grand potentiel de capture de valeur, d'accumulation de capacités technologiques et d'autonomie. Dans cette relation, (...) les entreprises nationales se connectent aux marchés mondiaux et deviennent des entreprises de premier plan qui organisent la production mondiale*".

Elles constatent que l'Inde, en plus de bénéficier d'une main-d'œuvre excédentaire anglophone, dispose d'une communauté transnationale très active dans l'entrepreneuriat numérique. L'afflux d'investisseurs étrangers et la diversification des conglomérats indiens existants a par la suite été associé à des politiques de libéralisation brutale de l'économie et plus spécifiquement des télécommunications et de l'éducation. Malgré l'évolution des entreprises indiennes vers un *couplage indigène*, certaines critiques considèrent que ce mode de développement perpétue les inégalités existantes. En ce qui concerne les Philippines, les chercheurs avancent que le besoin d'accents américains neutres a approfondi le recours aux travailleurs de ce pays dans le secteur des centres d'appels. Les principales réponses politiques locales ont consisté à mettre en place des *zones de libre-échange*. Elles qualifient ce couplage de structurel, car "*le manque d'entreprises nationales constitue un obstacle majeur à la diffusion des connaissances et à l'apprentissage, ce qui exclut toute possibilité de couplage indigène dans un avenir proche*." Le Kenya, dernier arrivé sur le secteur, a bénéficié de la hausse des coûts en Inde et d'un "*grand nombre de diplômés du secondaire et de l'université anglophones au chômage*".

253. Des politiques publiques orientées vers l'exportation ne saurait toutefois expliquer entièrement la place de l'Inde dans le commerce de services informatiques. Ainsi Balakrishnan (2011), résume les nombreux facteurs à l'oeuvre dans le développement du secteur informatique en Inde, en plus des facteurs précédemment évoqués : "*the mathematical heritage; an English-language higher-education system; the determination among the Indian elite that having "lost out" in the textile Industrial Revolution they could not afford to lose out in the Computer Revolution; the rise of teachable, standardized programming languages; a process view of work; the decline of the Indian rupee versus the U.S. dollar; and the advent of digitization that made services tradable over long distances is what created the setting for the emergence of the Indian IT industry*".

dans un fuseau horaire favorable”. Le pays se situe donc au bas de la chaîne de valeur, puisque certaines entreprises nationales travaillent parfois pour des intermédiaires indiens (*couplage structurel*)²⁵⁴ Ce travail illustre la variété des effets des politiques de développement basées sur l’installation du secteur des services dans les pays du Sud global et rejoint deux autres enjeux majeurs de la littérature sur les chaînes globales de valeur : la dispersion géographique des chaînes de valeur²⁵⁵ et le développement de recherches plus micro servant à cartographier les chaînes entreprise par entreprise qui permettraient travailler plus directement sur les effets des chaînes de valeur globales, notamment sur les conditions de vie des travailleurs du Sud global.

La question des rapports coloniaux a récemment ressurgi à travers l’étude des mécanismes de distribution de valeur au sein de l’industrie numérique. En 2019, le livre de [Couldry et Mejias \(2019b\)](#) a comparé la production de données des utilisateurs quotidiens des services numériques, à une forme de colonialisme des données (*“data colonialism”*), qu’ils définissent comme étant *“un ordre émergent pour s’appropriier et extraire des ressources sociales à des fins lucratives grâce aux données”*, source de valorisation pour de grands acteurs économiques privés, tels que Facebook ou Google. La comparaison avec le colonialisme s’appuie notamment sur la dimension systémique de ces entreprises qualifiées par certains de *“souveraines”*, car elles empiètent sur les compétences traditionnellement réservées aux États, comme pour la régulation de l’espace numérique en matière de contenus haineux²⁵⁶. Pour [Couldry et Mejias \(2019a\)](#), la *“datafication”* - ou mise en données du monde - contemporaine peut être comparée à la quantification sociale du XIXe siècle, car elles rendent possibles la catégorisation des individus (p. 119). Les géants technologiques actuels prolongent donc les logiques impériales d’hier dans leur entreprise d’appropriation des données. En outre, cette extraction de valeur est facilitée et élargie par les possibilités de coordination du numérique. Ces acteurs maîtrisent l’infrastructure logistique, ce qui leur confère un pouvoir certain sur les autres firmes à travers leurs services de stockage des données et de puissance de calcul qu’ils fournissent à d’autres entreprises plus petites. Ils montrent enfin que ce processus s’appuie sur plusieurs formes de rationalités qualifiées d’extractives, dont la première est économique : il s’agit de présenter les données comme sans valeur, parce qu’elles sont le résultat de nos interactions sociales et non de notre travail. La deuxième forme est qualifiée de légale et met

254. *“Les dirigeants ont souligné que le service à la clientèle sortant et les contrats basés sur les résultats n’étaient pas rentables, et pourtant, des acteurs inexpérimentés ont été la proie d’intermédiaires qui leur ont confié des travaux non rentables. Souvent, ces intermédiaires étaient des entreprises ou des consultants indiens qui cherchaient potentiellement des destinations à faible coût dans le cadre de leurs propres stratégies de chaîne de valeur.”*([Kleibert & Mann, 2020](#))

255. Ainsi, ([Kano et al., 2020](#)) considèrent que, si *“la responsabilité des chercheurs des chaînes de valeur mondiales (est) de mesurer systématiquement l’étendue géographique et la profondeur des activités pertinentes de la chaîne de valeur et de parvenir à une définition précise de ce que représente une chaîne de valeur mondiale”*

256. *“In such an environment, we see the emergence of what Philip Stern called the “company-state”: a corporation with immense powers to regulate not just trade but also law, land, and liberty—in short, to regulate life. At the same time, as Immanuel Wallerstein points out, capitalism (and by extension, colonialism) cannot flourish without the active participation of the state, which controls the relations of production by legalizing (or refus-ing to regulate) certain forms of labor, such as slavery in the past or temporary or gig-based work nowadays.”*([Couldry & Mejias, 2019b](#))

en lumière le fait que les données sont considérées comme en dehors de la propriété des utilisateurs. Enfin, la dernière forme de rationalité consiste à présenter l'utilisation des données comme susceptible d'apporter le progrès (*developmental rationalities*), ce qui passe par le fait de les partager (*cultural rationalities*). Cela se traduit aussi par des discours qui présentent l'appropriation des données comme légitime au regard des buts scientifiques ou économiques (*technical rationalities*).

Que ce soit dans le travail de [Couldry et Mejias \(2019a\)](#) ou dans le reste de la littérature citée dans cette partie, on retrouve pourtant peu d'analyse mentionnant la structure des coûts du capitalisme numérique, et estimant l'apport du travail gratuit. Nous pensons que l'analyse de [Couldry et Mejias \(2019a\)](#) repose sur l'hypothèse implicite d'une spécificité numérique qui faciliterait ce colonialisme des données, hypothèse discutable dans la mesure où les logiques post-coloniales persistent également dans des environnements non numériques. Dès lors, comment savoir si le colonialisme des données désigne simplement l'expansion et l'accélération de la collecte de données ou s'il existe des caractéristiques uniques au colonialisme axé sur les données ?

On peut également s'interroger sur une lecture qui met sur le même plan des travailleurs malgaches très peu payés qui passent 40 heures par semaine à raffiner des données avec des influenceurs ou des utilisateurs d'Instagram qui postent des photos. Peut-on mettre au même niveau l'expérience de travail de ces deux catégories de travailleurs ? Les uns, situés à l'autre bout de la planète exerceraient plutôt pour survivre, tandis que les seconds bénéficieraient de la sécurité économique des pays du Nord. Est-ce que le terme de "colonialisme" s'applique de la même manière ?

Nous pensons qu'il serait peut-être plus pertinent de parler de "*data extractivism*", car la notion d'extractivisme est déjà bien identifiée dans la littérature ([Chagnon et al., 2022](#)) à propos d'autres secteurs comme l'industrie textile ou minière. Il désigne premièrement l'appropriation de valeur dans les pays du Sud par ceux du Nord et l'altération, notamment environnemental, de la source de cette valeur ou encore la centralisation du pouvoir économique et social de la part de ces mêmes acteurs des pays du Nord qui se base sur des inégalités parfois préexistantes. Ces relations impliquent parfois de la violence (en termes de conditions de travail) et de la non-réciprocité (peu ou pas d'investissement dans le pays). Antonio [A. A. Casilli \(2017a\)](#) propose également un cadre pour analyser le travail numérique dans les pays du Sud globale, basé sur un appel à un "*tournant décolonial numérique*" pour rendre visible le travail invisible réalisé par des travailleurs marginalisés. Il affirme que "*le capitalisme de plateforme fonctionne en s'appuyant sur les constructions sociales de la race et du genre*". Nous verrons dans quelles mesures les concepts d'*extractivisme des données*, de "*colonialisme digital*" et de "*colonialité*" prennent beaucoup mieux en compte les mécanismes économiques et institutionnels qui sous-tendent le développement de l'industrie de données entre la France et Madagascar, ainsi que l'agentivité des travailleurs dans leurs trajectoires professionnelles.

Nos travaux s'inscrivent également dans l'analyse de la trajectoire de développement de Madagascar qui suscite de nombreuses questions tant le pays dispose d'avantages importants comparativement à d'autres pays du continent africain, comme l'expliquent Razafindrakoto, Roubaud, et Wachsberger (2018a). Si de nombreux facteurs structurels tels que l'attraction des investisseurs étrangers, la vigueur des institutions et les faibles violences, sont largement en faveur de la Grande Île, les auteurs montrent que la trajectoire de développement reste largement contrariée Razafindrakoto et al. (2018a). Des éléments tels que le clivage entre les zones rurales et les zones urbaines, les inégalités entre les élites et la majorité de la population et le manque de structuration des élites malgaches constituant des facteurs de ralentissement. Selon eux, la *”trajectoire économique malgache se caractérise d'abord par une « énigme », à savoir une tendance récessive qui perdure depuis l'indépendance, il y a plus d'un demi-siècle, et que rien ne semble pouvoir endiguer.”* Notre thèse ne prétend bien sûr pas résoudre cette ”énigme”. Néanmoins, il nous paraît pertinent de repartir du même constat pour tenter de trouver quelques réponses à notre propre paradoxe : comment se fait-il qu'un pays comme Madagascar se retrouve impliqué dans la chaîne de valeur d'une des industries les plus innovantes de notre époque? ²⁵⁷

En effet, d'après, Gil Razafintsalama (2020), le secteur des services numériques et numérisés qui inclue les opérateurs télécoms, les BPO et les ESN pèseraient 417,5 millions de dollars sur un PIB global de 13,72 milliards de dollars. Les entreprises de BPO représenteraient 17% de ce chiffre d'affaires soit presque 71 millions de dollars. En matière d'emploi, il y aurait 23000 travailleurs formels et 85000 travailleurs informels dans des activités telles que le développement web, la rédaction de contenus, la traduction, le design et le graphisme ou le community management selon des estimations évoquées par un rapport non diffusé sur le secteur (TPSD ²⁵⁸, 2020). Les estimations tirées de l'observatoire de l'ARTEC évoquent quant à elle un secteur qui pèserait 42000 dollars en 2016 et plus de 50 millions de dollars en 2019 pour se contracter en 2020 à 19 millions de dollars en 2020, très certainement du fait de la crise liée à la Covid 19. Toutefois, ces données sont susceptibles d'être largement minimisées. En effet, si l'on considère cette fois les données d'exportation de services informatiques de Madagascar de l'OMC, on constate le fort développement du secteur ²⁵⁹.

257. Les deux pays qui se rapprochent le plus de Madagascar de par leur position dans le secteur sont le Kenya et les Philippines. Toutefois, des indicateurs comme l'indice de développement humain montrent une réelle différence de développement, Madagascar étant à la 173ème place sur 191 pays, le Kenya étant 151ème et les Philippines 119ème

258. Trade and Private Sector Development

259. Pour des raisons de lisibilité du graphique, nous avons enlevé le Maroc dont les exportations sont 10 fois plus importantes que Madagascar

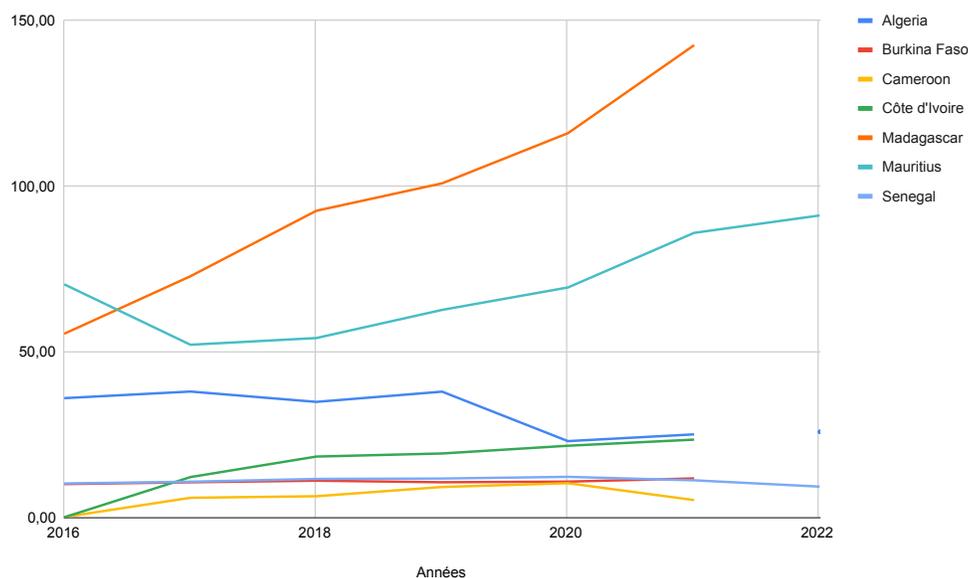


FIGURE 33 – Exportations annuelles en million de dollars des principaux pays exportateurs de services dans les pays d’Afrique francophone (hors Maroc, source : OMC)

L’objectif de ce chapitre est d’analyser les effets d’un secteur aussi dynamique sur le développement. Nous montrons notamment le fait qu’il se développe comme une économie de rente dans laquelle des entrepreneurs français captent la majeure partie de la richesse créée. Pour nous, il convient donc d’analyser la continuité entre la mise en place de ces relations commerciales autour de la donnée et les institutions coloniales, en tenant compte de la manière dont les travailleurs gèrent ces contraintes. Ces derniers, loin d’être passif, envisagent également ces mécanismes et ce secteur comme le vecteur de stratégies individuelles qui visent à se sortir de la pauvreté.

Dans ce chapitre, nous étudions par conséquent l’importance des facteurs coloniaux à travers deux types d’acteurs : les entrepreneurs français et les travailleurs de Madagascar. Une question sous-jacente à la littérature sur les chaînes de valeur se rapporte aux normes locales économiques, juridiques et aussi sociales dans lesquelles ce travail de la donnée est encadré. Nous verrons que le point central du développement de cette industrie à Madagascar est son rattachement à des stratégies plus globales d’attraction des investisseurs mises en place par l’Etat malgache autour du statut de “*zones franches*” : des zones économiques entièrement dédiées à l’exportation que nous détaillons plus bas.

Dans une seconde section, nous nous intéressons à un facteur que nous qualifions de “*compétences coloniales*”. Notre analyse s’appuie pour cela sur deux concepts que nous utilisons de manières complémentaires : le “*rappel colonial*” élaboré par Jan Padios et l’*assetization* coloniale. Selon nous, la chaîne de valeur

se structure d'un côté autour de la valorisation des compétences coloniales par les entreprises d'annotation de données, de l'autre par des stratégies qui consistent, pour les travailleurs, à construire des "*qualifications coloniales*" en vue de les valoriser sur le marché du travail de l'externalisation. Le risque est alors que cette chaîne de valeur modifie les mécanismes d'apprentissage qui existent à Madagascar en favorisant les compétences coloniales renforçant la dimension extractive du secteur.

8.1 Le développement du secteur repose sur un accès facilité aux facteurs de production pour les investisseurs étrangers

Une première dimension qui illustre le caractère extractif du secteur de l'externalisation de services informatiques réside dans les modalités de son installation à Madagascar. Pour analyser cet élément, il convient de revenir aux raisons qui expliquent l'arrivée des investisseurs étrangers sur la Grande-Île. La chaîne de valeur ne s'affranchissant pas du contexte institutionnel dans lequel elle se met en place, nous verrons que le développement du secteur s'est très largement appuyé sur l'instauration d'un régime de zone franche facilitant l'arrivée des investisseurs étrangers à Madagascar, d'abord dans d'autres activités notamment industrielles, puis dans le secteur des services.

Dans une première sous-section, nous analysons des données secondaires provenant de l'agence d'attractivité économique de Madagascar, qui nous permettent de mieux comprendre la provenance géographique des entrepreneurs de l'externalisation. Nous traitons ces résultats à la lumière des entretiens réalisés avec des entrepreneurs du secteur de la sous-traitance en identifiant deux formes de rationalité qui gouvernent leur investissement à Madagascar. Les premiers développent une rationalité *développementale* de par leur attachement identitaire à la Grande Île, tandis que les seconds constituent un groupe d'entrepreneurs qui circulent dans le monde entier à la recherche de pays suffisamment en retard de développement, mais assez développés pour offrir d'une main d'œuvre qualifiée. Par la suite, nous étudions la localisation de la chaîne de valeur sous l'angle de la qualification des travailleurs des données. Nous soulignons une différence fondamentale entre le secteur des services et l'industrie : sa dimension intrinsèquement cognitive et la nécessité de maîtriser des compétences linguistiques. Nous y abordons le niveau de français des annotateurs et leurs modalités d'apprentissage, ainsi que leur niveau de formation en explicitant la tension entre le niveau de qualification des annotateurs et la double dynamique de valorisation - minimisation des entrepreneurs de la sous-traitance et des entreprises d'IA.

8.1.1 Le secteur se développe avant l'arrivée du numérique à Madagascar en s'appuyant sur le statut des entreprises franches

Pour étudier la manière dont le secteur numérique s'est développé à Madagascar, il faut se reporter aux politiques publiques développées depuis l'indépendance du pays 1960. La mise en place d'un code d'investissement centré sur l'attractivité économique du pays fait naître un droit spécifique aux entreprises étrangères. Initialement dédié à des secteurs nécessitant un investissement important en matière d'installation de machines ou d'usines, comme c'est le cas du secteur industriel, agricole ou minier, ce code exclut alors *de facto* le secteur commercial (Rabetafika et Rajaonarison-Razafinavalona, 2010). Il faut attendre la fin des années 80 et la publication du quatrième Code d'investissement ainsi que d'un ensemble de textes sur les zones franches pour voir Madagascar s'inspirer de la politique commerciale de l'île Maurice en matière d'attraction d'investisseurs dans le secteur des services.

L'intérêt principal de la mise en place de ce statut de zone franche a alors été de permettre l'arrivée d'investisseurs dans des industries intensives en travail comme le textile et les services informatiques. Par effet boule de neige, d'autres entreprises se sont ensuite intéressées aux possibilités d'investissements à Madagascar pour finalement s'y installer²⁶⁰. A l'origine, le régime des zones franches vise d'abord à institutionnaliser l'arrivée d'investisseurs étrangers et notamment mauriciens dans le secteur textile à la fin des années 1980 en profitant de certains facteurs d'attraction d'industries nécessitant de nombreux travailleurs. Dimou et Fernand (2008) examinent le développement de trois zones franches textiles de l'océan Indien et identifient les facteurs d'installations des investisseurs étrangers avant la création de ce statut : les "faibles coûts de main-d'oeuvre"²⁶¹, une organisation du travail par "roulement", avec une "organisation flexible" qui facilite le recours à une main-d'oeuvre féminine et l'existence de réseaux privilégiés entre les communautés chinoises et indo-musulmanes de Maurice et la communauté Karane de Madagascar²⁶². In fine, on constate des similitudes entre le secteur des services et du textile : elles sont "labour-intensive", elles bénéficient fortement d'apports de capitaux venus de l'océan indien et d'acteurs occidentaux et du régime des zones franches pour se développer. Comme l'évoque M.N.A, ancien employé malgache de l'ambassade de France à Antananarivo, "Les entreprises d'outsourcing sont surtout installées dans les zones franches, qui disposent d'un statut spécial (...) pour des entreprises d'offshoring pour le textile et de plus en plus les [services informatiques].".

260. La loi 89-027 du 29/12/89 relative au régime de zone franche industrielle, instaure un régime spécial pour les entreprises étrangères industrielles et de services. Un rapport de l'OIT précise que : "Le régime de ZF est un régime d'exception (...) dans le but de promouvoir les investissements liés aux activités tournées exclusivement vers l'exportation."

261. : le salaire moyen malgache est trois fois inférieur au salaire moyen mauricien, lui-même "cinq ou six inférieurs au salaire moyen Hong-Kongais"

262. "Les Karanes ou Karana (...) sont des indo-pakistanaïens musulmans, pour la plupart émigrés à Madagascar dès la fin du XIXe siècle (...) [Certains] ne sont pas citoyens malgaches bien que présents sur l'île depuis deux siècles. Nombreux sont ainsi ceux qui, après l'indépendance de l'île, ont conservé la nationalité française puisqu'ils n'en avaient pas d'autre." (Wikipédia). Dimou et Fernand (2008) indiquent que ces "réseaux informels furent les socles de confiance sur lesquels les transferts de capitaux ont eu lieu dans un premier temps, avant que les régimes de zone franche ne soient institutionnalisés."

Les "entreprises de services" (ES) bénéficient donc de ce régime, sous réserve qu'elles opèrent dans des secteurs bien délimités comme "le traitement des données informatiques" et qu'elles soient quasi entièrement tournées vers l'exportation ²⁶³. Le régime instauré par Madagascar est considéré comme équilibré. D'un côté, il prévoit des exemptions d'impôts pour les entreprises durant les deux premières années puis des impôts faibles (taux fixe de 10% avec des possibilités de réduction). De l'autre, il évite le piège de créer un "ghetto juridique" en ne prévoyant aucune exemption au droit commun en matière de droit du travail. Un rapport de l'OIT précise ainsi que le statut applique "la plupart des huit Conventions fondamentales de l'OIT" et notamment la liberté syndicale, l'abolition du travail forcé, l'égalité et l'élimination du travail des enfants. Ainsi, les entreprises ayant obtenu le statut d'entreprise franche doivent respecter le droit du travail national et notamment :

- elles doivent "s'acquitter des cotisations sociales prévues par le Code du travail et le Code de prévoyance sociale";
- respecter les minimas salariaux (50 euros par mois);
- d'employer des personnes de plus de dix-huit ans;
- "d'affilier ses travailleurs locaux à un organisme de sécurité sociale agréé à Madagascar et à un service médical interentreprises (SMIE)";
- les employés étrangers, hors des cadres, doivent représenter au maximum 5% des effectifs
- "Les CDD peuvent aller de trois mois à trois ans, mais ils sont non renouvelables au-delà de cinq ans et deviennent des CDI";
- "Pour les femmes, comme dans le cas des ZF, le travail de nuit est autorisé mais doit faire l'objet d'une attention particulière (obligation d'assurer le transport)".

Dans un article, Razafindrakoto et Roubaud (1996) examinent les résultats obtenus par ce régime spécial et constatent que ces zones ont accéléré la croissance tout en permettant un meilleur traitement des travailleurs en comparaison avec d'autres secteurs. D'après eux, cette "expérience constitue jusqu'à présent le meilleur exemple de réussite de la transition de Madagascar vers l'économie de marché et l'insertion internationale." Ils notent également que "la montée (...) des activités de "service" (saisie informatique) mérite (...) d'être soulignée". En 1996, sur l'ensemble des entreprises franches, 12,3% sont des entreprises de services et constitue donc le troisième secteur le mieux représenté derrière le textile (48,8%) et l'industrie "diverses" (16%). Sur les 36200 emplois des entreprises franches, le secteur des services représenterait 500 personnes en 1996. Cependant, ce régime est de plus en plus controversé au point que sa suppression a d'ailleurs été entérinée par le projet de loi de finances 2022, notamment à cause d'un manque de contrôle qui s'exerce sur les entreprises ²⁶⁴ et prise en compte des enjeux environnementaux.

263. Ces entreprises doivent réaliser plus de 90% de leur chiffre d'affaires en dehors de Madagascar

264. Ainsi, un rapport constate encore : que "les entreprises franches (...) n'ont jamais été regroupées au sein d'une zone dédiée formant une enclave territoriale, facilement contrôlable. En 2018-2019, le nombre d'entreprises franches s'élève à 449 mais il chute à 249 après une opération de contrôle retirant 200 agréments à des entreprises jugées non conformes au régime de zone franche (notamment

Historiquement porté par le développement du régime de zones franches, le secteur de l’externalisation de services numériques constitue désormais une industrie prépondérante de l’économie malgache ²⁶⁵. Cette dynamique autour de l’externalisation de services et du statut de zone franche se traduit très clairement dans la stratégie de l’*Economic Development Board de Madagascar* (EDBM). Un employé de cette agence publique nous explique en effet constater une recrudescence d’entreprises et d’investisseurs s’installant à Madagascar dans ce secteur : *”Là aussi ces dernières années (...), il y a plein de call [centers] qui sont créés parce que le marché est vraiment très ouvert à Madagascar (...)”* (Fanino - EDBM). Au fil de nos échanges, ce dernier a accepté de nous fournir des chiffres illustrant cette tendance. Tout d’abord si l’on considère les entreprises de technologie de l’information et de la communication, à la fois franches et non franches, l’organe enregistre 587 entreprises entre 2017 et septembre 2022, soit 7,7% des entreprises créées globalement à Madagascar. Depuis 2007, si l’on se focalise sur l’externalisation de services dans son ensemble, 182 entreprises ont été créées. Aujourd’hui, 70 entreprises sur les 587 enregistrées comme entreprises du secteur informatique déclarent faire de l’externalisation de services.



FIGURE 34 – Répartition par année de création des entreprises franches encore actives dans l’externalisation de services

non-respect des obligations d’export).”

265. Un [article](#) indique par exemple : *”Et si la Silicon Valley, c’était dans les Hautes Terres malgaches? C’est le pari que font les entreprises étrangères qui installent ici des centres d’appel ou des services de développement numérique (...). Madagascar devient une destination de plus en plus prisée en la matière.”*

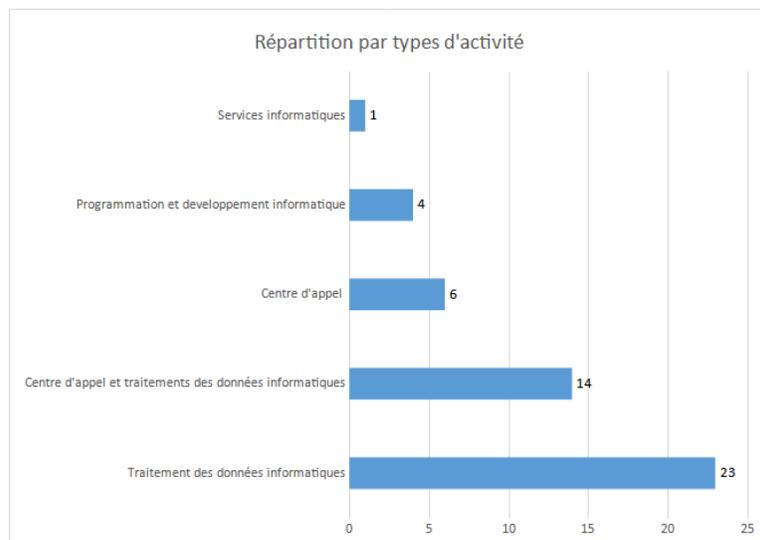


FIGURE 35 – Répartition par types d’activité des entreprises franches encore actives dans l’externalisation de services

Sur les entreprises franches, l’EDBM en enregistre 254 qui déclarent être actives aujourd’hui. Sur ces entreprises-là, 48 sont des entreprises du secteur de l’externalisation de services (17% du total). Parmi elles, 26 sont des entreprises françaises, 7 sont mauriciennes et 4 seulement sont malgaches. Ces statistiques montrent la difficulté des firmes locales à émerger face à la concurrence des entreprises françaises. Bien souvent, ces entreprises sont détenues directement (propriété directe) ou indirectement (actionnariat) par des français. Néanmoins, il s’agit d’une dynamique plus large autour de la propriété du capital à Madagascar : *”il y a toujours eu ce lien entre Madagascar et la France, qui fait que la plupart des investisseurs même dans les autres secteur étrangers [de zones franches] restent des français (...).”* (Fanino - EDBM).

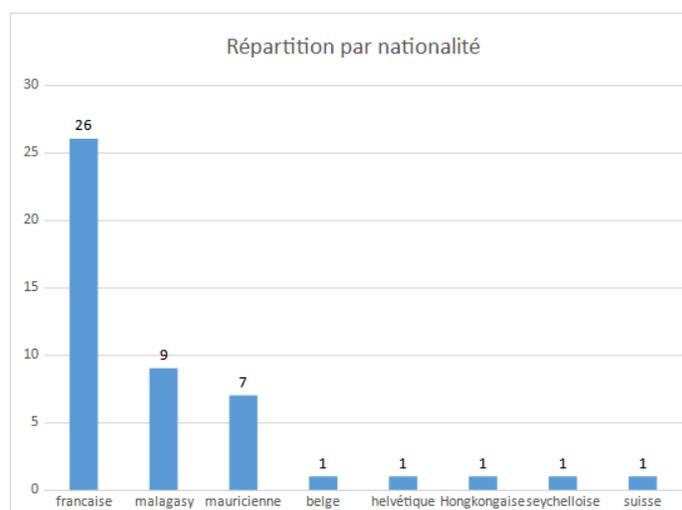


FIGURE 36 – Répartition par nationalité des propriétaires des entreprises franches encore actives dans l’externalisation de services

Cet enjeu de la propriété des entreprises fait écho à débat très important autour de la nationalité des entreprises et des entrepreneurs présents à Madagascar. Selon Fanino de l'EDBM, *"entreprise malgache, ne veut pas dire aussi que ce sont des malgaches (...), parce que c'est peut-être une entreprise malgache mais dont les associés viennent d'autres pays."* (Fanino - EDBM). Il est ainsi fréquent que des entrepreneurs locaux acquièrent la nationalité française, afin de mieux défendre leurs intérêts économiques²⁶⁶. Le discours de Fanino trouve un écho dans un entretien réalisé avec Miky, anciennement DRH d'une grande entreprise d'externalisation, dans lequel elle nous explique que : *"Il y a aussi cet état d'esprit toujours maintenu entre l'homme blanc et l'homme noir. (...) Je pense qu'il y a toujours l'importance de la colonisation. Le fait que les entreprises offshore sont toujours détenues par des vazaha, des étrangers. Il y a aucune crainte pour les vazaha de venir vivre ici."* (Miky - Ancienne RH Sous-traitant C)

La présence des entrepreneurs étrangers peut également être analysée à travers leur parcours professionnel. Dans nos entretiens, on retrouve tout d'abord une minorité de Malgaches ayant émigré en France au moment de leurs études. Cette voie d'accès est celle qui est la plus accessible dans un parcours d'émigration entre Madagascar et la France, mais reste réservée à des étudiants appartenant à l'élite malgache. C'est le cas de Patrick qui part en France faire ses études dans un parcours spécialisé dans le commerce international et l'agroalimentaire²⁶⁷. Il rejoint ensuite l'entreprise familiale située à Madagascar et à l'île Maurice ce qui fait écho aux travaux de [Saxenian, Motoyama, et Quan \(2002\)](#) sur les entrepreneurs indiens des nouvelles technologies, qui montrent une circulation de ces derniers entre les États-Unis, où ils font leurs études et réalisent une progression de carrière, et l'Inde où ils retournent développer des sociétés technologiques.

Encadré n°60 : La circulation des élites franco-malgaches

Je suis franco-malgache d'origine indienne. Je suis né et j'ai grandi à Madagascar dont je suis citoyen aussi. J'ai passé mon bac ici et ensuite je suis parti en France, à Paris, pendant 6 ans. J'ai un parcours assez classique en école de commerce, orienté finance, à l'ESG, entre 2005 et 2010. Ensuite j'ai fait un master spécialisé à l'ESSEC (...) en management international agroalimentaire. Pourquoi (...) ? Parce que l'année suivante je suis rentré à Madagascar rejoindre l'entreprise familiale qui œuvrait dans le domaine de l'agroalimentaire.

Patrick - CEO - Sous-traitant E

On retrouve un type de profil assez similaire dans le cas d'une entrepreneure qui descend de la première

266. Comme l'explique M.N.A, ancien stagiaire de l'ambassade de France à Antananarivo : *"Les fournisseurs sont Total et des entreprises françaises, mais dirigé par des malgaches qui ont acquis la nationalité française. On a une définition très large de ce que c'est une entreprise française. C'est les conséquences de l'histoire coloniale, mais ils prennent la nationalité, car ça permet de protéger ses actifs économiques et commerciaux. Un des arguments qu'on [l'ambassade] faisait valoir, c'était de leur dire qu'on pouvait leur apporter un soutien."* (M.N.A - stagiaire à l'Ambassade de France à Madagascar)

267. Lui-même est issu d'une grande famille d'industriels malgaches

génération d'émigrés malgaches en France. Née en France où son père a émigré depuis Madagascar durant ses études, elle retourne sur la Grande Île pour fonder une entreprise, à la suite de ses études et d'une première expérience professionnelle dans l'humanitaire. Elle nous offre un exemple de la manière dont certains individus naviguent entre leur pays d'origine et des pays occidentaux pour poursuivre leurs objectifs professionnels tout en conservant un lien avec leurs racines personnelles.

Encadré n°61 : La circulation des élites franco-malgaches 2

Je suis née et j'ai vécu presque toute ma vie en France. Mon père est d'origine malgache. Il est né [là-bas] et il est parti en France faire ses études. Moi, j'ai fait mes études en France (...). Au début, j'ai fait des études de bio [à la] fac à Jussieu avec l'idée que je voulais faire les choses un peu différemment. J'avais habité à Madagascar et je voyais le côté difficile de ce que la population vit au quotidien, donc j'ai fait des études en nutrition. Et j'ai enclenché sur un Master de stratégie, si on peut appeler ça comme ça, un Master à l'ESSEC [en] entrepreneuriat social.

Dinasoa - CEO - Sous-traitant E

Un autre type de parcours moins présent concerne cette fois Pierre, un français qui sort d'une grande école d'ingénieur française et qui s'est ensuite spécialisé dans l'intelligence artificielle. Après plusieurs expériences en tant que *data scientist*, il décide finalement de s'orienter vers l'entrepreneuriat dans l'annotation de données à Madagascar. Davantage ancré dans le secteur de l'IA, il justifie lui-même cet ancrage géographique par une connaissance du pays acquise à travers les expériences de deux autres membres de sa famille.

Encadré n°62 : L'exemple d'un parcours ancré dans les entreprises technologiques

J'ai un parcours très technique. Je suis Polytechnicien, j'ai fait une thèse en robotique. Après (...), j'ai fait trois ans dans deux start-ups en tant que data scientist (...) à Paris (...). J'ai une histoire avec Madagascar parce que j'ai deux frères qui ont passé du temps là-bas, dont un qui y est encore. Dans ma dernière expérience en tant que data scientist, je ronronnais un peu. Je voulais devenir l'expert du deep learning, mais en fait je me suis rendu compte que ce n'était pas ce que je voulais faire. J'avais fait pas mal de [hackathon]. Ensuite, via un de mes frères (...), j'ai été mis en relation avec une personne à Madagascar, un français, qui voulait monter une société [d'annotation]. Très vite, ça m'a beaucoup intéressé parce que ça restait dans le domaine que je connaissais bien, mais déjà c'était ma propre société et il y avait un côté beaucoup plus varié et rester dans le domaine sans être moi-même dans les algorithmes, mais rester quand même dans quelque chose de technique. (...) On met en place des outils, il y a un peu de programmation, et ce côté gestion de projet [à] Madagascar.

Pierre - CEO - Sous-traitant G

Un dernier type de parcours est plus majoritaire et se caractérise par sa grande mobilité géographique : il s'agit des *entrepreneurs de l'externalisation*. Ces derniers ont un profil commercial et international avec une spécialisation qui s'est faite au fil de leur carrière et des évolutions du secteur de l'externalisation de services. Dans le cas de Frédéric, cela s'est traduit par un parcours professionnel le conduisant là où se situaient les entreprises d'externalisation : en Thaïlande, aux Philippines et finalement à Madagascar. Après plusieurs projets entrepreneuriaux, il a finalement décidé de créer sa propre structure qui comporte plusieurs centres de production en Roumanie et à Madagascar. Dans d'autres cas similaires, il s'agit plutôt de salariés d'entreprises d'externalisation françaises qui ont circulé entre le Maroc et Madagascar, au fur et à mesure des opportunités de poste et de l'évolution de la chaîne de valeur du secteur des services.

Encadré n°63 : L'internationale des dirigeants d'entreprises d'externalisation de services

J'ai fait des études en école de commerce Bac+5 à Paris. Très rapidement, j'ai créé une première boîte à Paris, pendant 4/5 ans, dans le secteur de l'informatique (...). J'ai revendu cette boîte et je suis parti, à 28 ans (...) en Asie du Sud-Est pendant 4 ans. J'ai fait Hong-kong, Singapour, la Thaïlande, le Laos, le Japon, etc. J'avais créé une agence dans le digital à Singapour, qui était dans le référencement à l'époque. C'était pour référencer des boîtes asiatiques sur le marché européen, et l'inverse (...). J'ai fait ça pendant 2/3 ans. Quand je vous dis « agence digitale », c'était à la fois du référencement et de la production digitale. Je produisais déjà du digital à partir des Philippines, du Vietnam, de l'Inde, etc. Ensuite, je suis reparti. La première boîte que j'avais créée était déjà sur Madagascar. Le pays avait beaucoup de potentiel pour moi dans tout ce qui était sous-traitance. (...) Au départ je travaillais pour quelqu'un, mais au bout de 6 mois j'ai abandonné. J'ai créé une boîte concurrente (...). Ça a duré 7 ans.

Frédéric - CEO - Sous-traitant F

Ces cas illustrent la forte représentation des entrepreneurs et capitaux français dans la création d'entreprises franches de services à Madagascar. Sur la base de ces éléments, nous pouvons déjà retenir quelques points concernant la présence du secteur de l'externalisation à Madagascar. Tout d'abord, le secteur textile et celui de l'externalisation de services informatiques partagent des facteurs communs de développement en profitant à la fois régime des zones franches et de l'accès à une main-d'oeuvre peu onéreuse. Par effet boule de neige, ce système a contribué à faciliter l'installation de plus en plus d'entreprises de l'externalisation de services, principalement détenues par des entrepreneurs extérieurs à Madagascar. ce sont également deux industries comparables, car considérés comme *a-spatiales* et donc fortement mobiles géographiquement, puisque c'est en théorie le coût des facteurs de production et notamment du travail qui sera déterminant.

Une autre facette de cette dynamique post-coloniale porte sur le développement du numérique à Madagas-

car. On a pu en effet constater une dynamique d'allongement des chaînes de valeur du travail numérique qui s'appuie sur une dimension caractéristique du secteur : sa mobilité. L'installation de structures de production nécessite peu d'investissement, car il n'y a pas besoin de machines très coûteuses. Il est finalement assez aisé pour les entreprises françaises de délocaliser leurs chaînes de production vers des pays pour lesquels les bénéfices seront beaucoup plus importants. Dans le cas français, on constate un investissement dans les anciennes colonies françaises à travers ce mécanisme d'allongement des chaînes de valeur. Ce phénomène passe à la fois par des relocalisations d'entreprises initialement implantées dans les pays du Maghreb à Madagascar et dans d'autres cas par de la sous-traitance directe à Madagascar.

Les entretiens réalisés avec les entrepreneurs de l'externalisation confirment une tendance, déjà perçue dans la littérature sur le travail à distance et le micro-travail, au sujet de l'importance de l'accès à Internet dans le développement de ce genre de secteur. Dans la Grande île, la connectivité est très faible pour les citoyens du fait des coûts de connexion et de l'infrastructure, même si l'usage des smartphones à Antananarivo facilite l'accès à Internet ²⁶⁸. D'un point de vue institutionnel, Madagascar dispose de plusieurs acteurs très investis sur les enjeux numériques. Ainsi, après la libéralisation du secteur des télécommunications en 1996, l'Office Malagasy d'Études et de Régulation des Télécommunications (OMERT) a été créé pour réguler le secteur et est remplacé par l'ARTEC en 2014 qui dépend du ministère des Postes, des Télécommunications et des Développement Numérique (MPTDN). Historiquement, c'est depuis 1975 que le ministère des Postes s'est occupé du secteur des télécommunications intégrant petit à petit les possibilités de développement offertes par le numérique, notamment à travers la loi de 17 octobre 2005 qui indique que la politique du pays vise à : *"favoriser la croissance de l'économie et de l'emploi par les télécommunications et TIC"*.

Le développement des entreprises de services repose en effet sur une amélioration des infrastructures réseau, permettant un meilleur accès à un coût réduit. Le développement des activités d'annotation dans le secteur plus large de l'externalisation de services constitue justement une des conséquences du développement progressif des infrastructures d'accès à Internet. Comme l'explique Nirry, qui est aujourd'hui responsable du numérique à l'agence universitaire francophone (AUF) après un parcours universitaire et professionnel dans le numérique depuis le début des années 2000, le développement de la connectivité s'est fait en deux phases : l'arrivée d'Internet d'abord, puis de la fibre à partir du milieu des années 2000. Le secteur des services a suivi ces évolutions en passant de travaux de saisie de données à des activités réalisables à distance. Il fallait initialement envoyer des livres à numériser par avion dans des ateliers situés près de l'aéroport de Antananarivo. Puis, l'arrivée de la fibre a accéléré le développement du secteur au milieu des années 2000 grâce à la possibilité de réaliser des activités complètement à distance.

268. Pourtant, dans un [rapport de 2016](#), l'International Telecommunication Union (ITU) constate que les données sur l'usage des technologies de l'information et de la communication sont très limitées, notamment car l'INSTAT n'a jamais mené d'enquête sur leurs usages par les ménages malgaches. Ce n'est qu'en 2018, que l'Autorité de Régulation des Technologies de Communication (ARTEC) a lancé deux enquêtes sur l'utilisation des TIC par les ménages et par les entreprises.

Encadré n°64 : L'arrivée du numérique à Madagascar en deux phases

”Sur l’informatique (...) l’internet est arrivé en 1995 amené par les français via une filiale de France Telecom qui s’appelait Data Telecom Service(...). Ils sont arrivés grâce à l’aide, de USAID [aide au développement US](...) Le BPO ça commence quand on a [eu] la fibre, on a été connecté en 2007 via un projet de la Banque Mondiale qui a vu le jour avec des câbles sous-marins qui font le tour de l’Afrique. ”

Niry Andriambelo - Responsable du Numérique régionale de l’Agence Universitaire Francophone

Des entreprises initialement situées dans les pays du Maghreb, comme au Maroc ou en Tunisie, se mettent alors à créer des filiales dans la capitale malgache. Louis, CEO du sous-traitant C évoque particulièrement la spécialisation progressive de son entreprise dans des métiers où la connexion Internet est nécessaire comme la modération de contenu, le service après-vente via chat, et finalement l’annotation de données ²⁶⁹. Il convient néanmoins nuancer cette évolution, car le coût d’accès à Internet dans la capitale demeure élevé du fait de l’absence de connexion à la fibre optique ²⁷⁰.

La circulation des investissements et le fait de ”déporter” des activités de service vers Madagascar, pour reprendre les termes de Louis, fait écho aux analyses portant sur les restructurations de chaînes de valeur, Kleibert et Mann (2020) observant un phénomène similaire entre l’Inde et le Kenya. Louis explique également que le paysage de l’annotation de données se structure en deux flux principaux, avec l’Inde pour le monde anglophone et Madagascar pour les pays francophones : ”Sur le data labelling anglophone, ils vont plutôt aller vers l’Inde, pour le francophone, la destination c’est Madagascar (....) j’ai l’impression c’est la destination en vue.” (Louis - CEO - Sous-traitant F). À propos de l’Inde, on a pu constater que certaines entreprises réintermédièrent des contrats d’annotation de données vers des pays moins chers comme le Kenya ou Madagascar (Partie 2). Cette dynamique semble également s’engager du fait de la remontée de l’Inde sur la chaîne de valeur, car comme l’explique Patrick à propos d’un de ses clients américains : ”Les Indiens aujourd’hui sont en train d’avancer énormément dans les technologies et on a des fuites (...) Donc le problème, c’est que vous avez un prestataire qui vous assure qu’il bosse que pour vous et au final, il bosse pour vous, pour

269. ”Où on a pu justement aller dans des nouveaux métiers où la connectivité et la connexion nécessitent justement plus de choses, comme la voix par exemple pour la voix sur IP, puisque sinon ce n’était pas possible et ça s’est vraiment démocratisé ici à partir de fin ... on va dire 2014 2015. Là on a vu WebHelp, TP, en tout cas tous les les gros acteurs qui étaient en France avant, ils ont tous déportés du Maghreb et qui maintenant ont besoin de continuer à se déporter et ils sont arrivés ... pour d’autres raisons ...”(Louis - CEO - Sous-traitant F)

270. ”Parce que comme le pays est pas encore totalement tout fibré, hein, la fibre elle arrive sur les côtes et après c’est par satellite que ça arrive ... ça va être de Tamatave, Majunga, Fort-Dauphin ou autre, ça vient par satellite donc ... C’est pas encore la terre qui est totalement fibrée, mais pour le faire, il faut forcément des coûts et on ne va pas tout de suite descendre”. (Louis - CEO - Sous-traitant F)

vos concurrents et toutes ces équipes sont collées. Donc il y a de très gros risques de fuites et ce n'était plus secure (...)" (Patrick - CEO - Sous-traitant E)

On constate finalement, un allongement des chaînes de valeur du fait de la baisse relative du coût des facteurs de production dans des pays comparativement de moins en moins développés comme Madagascar par rapport à des pays qui se développent et dont les coûts augmentent, comme l'Inde ou les pays du Magreb. Par conséquent, la Grande Île devient beaucoup plus compétitive sur des prestations qui demandent beaucoup de volume comme l'annotation de données. Ces éléments sont confirmés dans les entretiens que nous avons réalisés avec des entrepreneurs de la Grande Île : *"Madagascar offrait, et c'est toujours le cas aujourd'hui, de vrais avantages avec une main-d'œuvre disponible, francophone, compétitive, de l'infrastructure, notamment Internet, de qualité, un décalage horaire avec la France (une heure ou deux) qui est très proche et une culture francophone."* Cette évolution pousse donc des entreprises asiatiques et anglophones à installer à Madagascar leurs activités en matière d'annotation de données : *"depuis deux ans (...) le marché américain et potentiellement asiatique [de l'annotation] se déporte sur Mada. Ils étaient (...) en Inde auparavant. Et comme forcément on cherche tous à optimiser, parce (...) qu'il y a tellement de données (...) qu'au final ça se joue à [rien] (...) ça se joue à trois chiffres ou quatre chiffres après la virgule. L'Inde maintenant commence à perdre son marché. C'est comme ça qu'Amazon s'est déporté. En tout cas, ils sont de plus en plus ici tout en gardant un peu de marché en Inde."* (Louis - CEO - Sous-traitant F).

Il convient maintenant d'évoquer le facteur de production le plus important pour les entreprises de services : la main d'oeuvre. Sur cette question, Louis raconte, à travers son parcours professionnel, l'allongement des chaînes de valeurs. Sa manière de donner du sens à l'émergence de Madagascar comme place forte des services repose sur ce qu'il a pu observer au Maroc où il a commencé sa carrière. Selon lui, c'est l'évolution du modèle marocain en matière de protection du travail et de niveau de salaire qui favorisent cette relocalisation à Madagascar : *"Aujourd'hui, on a des prix qui sont montés. La Roumanie coûte moins cher que le Maroc aujourd'hui. En plus, ils ont installé - je ne dis pas que c'est bien ou mal, mais ils ont installé le modèle français des choses. Donc on se retrouve avec la même problématique que les IRP, donc les instances représentants du personnel. C'est les mêmes problématiques qu'en France, en fait, que maintenant, donc ça manque d'agilité."* (Louis - CEO - Sous-traitant F). Il est également nécessaire de mettre en perspective cette dynamique avec ce que raconte Fanino, chargé de mission de l'agence d'attractivité territoriale de Madagascar à propos de la qualité des travailleuses et des travailleurs malgaches. Pour lui, les jeunes malgaches sont de plus en plus qualifiées en informatique, pour un coût qui demeure peu élevé²⁷¹. Cette question de la formation et de la qualification initiale des travailleurs ainsi que la capacité des entreprises de services numériques à former ces travailleurs est d'autant plus essentielle que la valeur des prestations de services est très largement de

271. "La main-d'oeuvre à Madagascar on va dire est pas très chère, par rapport à d'autres pays, alors qu'il y a quand même beaucoup de jeunes surtout qui sont à la recherche de travail qui éprouvent un intérêt conséquent dans le secteur tic" (Fanilo - Chargé de mission EDBM)

ces facteurs.

Finalement, ces premiers résultats montrent que l'installation des entreprises de sous-traitance s'inscrit dans des politiques volontaristes mises en place par l'État malgache comme stratégie d'attraction d'investisseurs étrangers dans des industries intensives en travail humain. Ces politiques constituent une réponse aux retards de développement identifiés dans les années 1980, ainsi qu'une manière d'appliquer le libéralisme économique promu par la Banque Mondiale en intégrant Madagascar aux circuits d'échanges mondialisés modernes.

Une première conséquence de ces politiques publiques est d'avoir largement encouragé l'installation d'investisseurs français et mauriciens au détriment des investisseurs locaux. Par ailleurs, si le régime de zones franches ne s'est théoriquement pas accompagné d'entorses sévères dans le droit du travail local, ces dernières ne sont pas directement contrôlées par les institutions publiques malgaches et notamment par l'EDBM. Ces résultats suggèrent un tout autre développement analytique pour le concept de "colonialisme des données" (Couldry & Mejias, 2019a). Plutôt que d'évoquer l'extraction de valeurs liée aux données personnelles des utilisateurs par les grandes entreprises numériques, nos résultats illustrent l'inscription du travail des données dans des réseaux de production post-coloniaux qui s'appuient sur le développement inégal des pays.

Par la suite, nous avons identifié deux types d'élites du secteur de la sous-traitance. Grâce à leur capital culturel, lié à leurs études et à leur nationalité (française), ainsi qu'au capital économique qu'ils tirent l'exercice d'activités professionnelles en France, les premiers peuvent aisément créer des entreprises à Madagascar en se présentant comme malgache. Leur position est unique dans le paysage de l'IA puisqu'ils associent deux actifs très difficiles à obtenir simultanément pour les autres acteurs de la chaîne de valeur : la connaissance de Madagascar (coté travailleurs) et la connaissance des pays occidentaux (côté clients) ("*Entrepreneurs de développement*"). Une autre catégorie d'entrepreneurs concerne cette fois des français qui ont des trajectoires internationales. Ces derniers suivent les investissements des entreprises de sous-traitance dans des pays en marge de la mondialisation ("*Entrepreneurs des marges du développement*"). Leurs discours suggèrent deux attitudes vis-à-vis du travail des données. Les premiers affichant clairement un objectif de contribution au développement de Madagascar, tandis que les seconds soulignent davantage la rationalité économique associée à la localisation de leurs activités, passant ainsi d'un pays à l'autre, en fonction de la remontée des pays sur la chaîne de valeur. Pour ces derniers, l'accès aux pays les plus en marge de la mondialisation constitue la source principale de production de valeur.

Dès lors, comment qualifier l'intégration de Madagascar dans la mondialisation à travers le cas du secteur des services ? Dans notre introduction, nous avons abordé le travail de Kleibert et Mann (2020) qui présente différentes formes de couplage : structurel, fonctionnel et indigène. La structure du secteur nous permet

d'identifier l'existence d'une intégration fonctionnelle des entreprises situées à Madagascar. Pour mieux tenir compte de la rationalité économique et *développemental* de ces choix de localisation, nous proposons de qualifier cela d'intégration "*marginale*" pour désigner ce mécanisme d'exploitation par la perception d'un retard de développement.

Nos résultats suggèrent alors que, dans le secteur des services, "*le monde n'est pas plat*", mais plutôt que les entrepreneurs de la sous-traitance délocalisent leurs activités dans des pays à la marge pour garantir leurs marges. Ce rapport entre le coût et la localisation géographique la chaîne, s'inscrit tout d'abord davantage processus de réintermédiation. De plus en plus d'entreprises françaises, américaines ou indiennes délocalisent dans des pays comme Madagascar pour préserver leurs marges.

8.1.2 La qualification des travailleurs des données : des parcours de formation initiaux qui attestent de compétences suffisantes pour intégrer le secteur

Malgré le mode de recrutement et les discours des entrepreneurs de la sous-traitance qui insistent sur le faible niveau nécessaire pour occuper ce poste, nos résultats indiquent que notre échantillon est constitué de travailleurs qui disposent de qualifications importantes au regard de la population générale malgache. Cette deuxième sous-section est l'occasion d'aborder cette piste d'analyse.

Concernant les variables liées à la formation initiale, nos données s'éloignent assez largement de ce qui est observé en population générale (Table 47). Tout d'abord, un rapport de l'Institut National de la Statistique malgache (INSTAT) souligne que l'offre scolaire se caractérise par une "*forte contribution du secteur privé*" : 19,7% des enfants malgaches sont inscrits dans le secteur privé contre 63% dans notre échantillon. Nos résultats concernant la formation initiale des travailleurs marquent une différence importante avec la population générale, notamment sur le plan du recours à la scolarité privée et l'accès aux études supérieures : 75% des annotateurs sont passés par le supérieur contre 10,5% de la population urbaine (Table 48). De manière générale, le niveau d'instruction est très faible avec : 28,4% des malgaches qui n'ont fait aucune étude, 44,4% qui ont fait des études primaires, 17,8% qui ont fait le collège, 6,5% qui ont fait le lycée, 2,8% qui ont fait des études supérieures. Ces résultats suggèrent que les travailleurs des données sont issus d'une population urbaine capable de financer des études, notamment privées.

Comment donner du sens à cette sur-représentation de travailleurs issus de l'université? Nos entretiens suggèrent que les travailleurs pourraient également désigner par cela un éventail de formations post-bac très spécialisées liées aux activités numériques et aux langues étrangères, mais qui ne se déroulent pas nécessairement à l'Université. Il convient également souligner que 19% de notre population est encore étudiante, ce qui correspond à des stratégies délibérées de certaines entreprises de services informatiques qui proposent parfois des horaires aménagés, afin que les annotateurs continuent à étudier. Cette proportion

d'étudiants est également liée à l'accès à l'enseignement supérieur à distance (CNTMAD²⁷²) qui permet à certains travailleurs de poursuivre des formations qu'ils cumulent avec un travail à temps plein ou par le cumul entre des *shift* de nuit et une journée de cours. Il pourrait également s'agir d'une conséquence d'une sur-sélection des annotateurs opérés par les entreprises. Certains sont recrutés à minimum en post-bac, d'autres à bac +2, et d'autres encore à des niveaux encore plus élevés (même bac +3 ou bac +5), le déficit d'emplois dans d'autres secteurs conduisant ces derniers à accepter des emplois "hors" de leur champ initial de qualifications. La nécessité de maîtriser l'informatique ainsi que la langue française et d'autres langues étrangères contribuent à également cette sur-sélection.

Outre la formation initiale, les travailleurs sont nombreux à avoir fait des formations professionnelles (77%), notamment sur les outils informatiques. Les entretiens nous permettent d'expliquer en partie cette tendance à travers deux dynamiques d'apprentissage. Tout d'abord, certaines formations sont faites en complément de l'université pour obtenir des certifications en langues ou en informatique : respectivement 59% et 33% des travailleurs déclarant avoir réalisé des formations professionnelles sur ces sujets (Table 48). Certains travailleurs se forment dans des écoles d'informatique et de langues étrangères comme Ando, annotateur au moment de l'entretien et qui se forme pour "explorer le monde des Call centers". Pour ces derniers, ces formations leur permettent de faire valoir des compétences lors de recrutement dans les entreprises du secteur²⁷³. Dans d'autres cas, cela sert à "pallier" un départ prématuré de l'université, par exemple avant l'obtention d'un diplôme, en développant des compétences mobilisables dans l'industrie des services. Pour d'autres encore, il s'agit également de remédier au manque de moyens pour aller à l'université, comme dans le cas de Félix qui explique avoir choisi une "formation en informatique bureautique" considérant qu'il s'agit du "moyen le plus facile pour trouver un emploi", par rapport à un parcours universitaire qu'il considère comme inaccessible : "je ne pense pas aller à l'université à Madagascar, parce que c'est (...) difficile, (...) parce que j'habite très très loin d'ici, loin de la ville. Et mes parents n'ont pas les moyens de m'envoyer à l'université (...). (Félix - Annotateur - Sous-traitant G). Dans tous les cas, il s'agit d'une stratégie délibérée des jeunes malgaches dans la perspective de postuler dans des entreprises de zones franches, qui sont très connues dans la capitale pour recruter énormément de jeunes travailleurs.

Un autre aspect très important de la qualification des travailleurs repose sur le fait d'acquérir un niveau de français "suffisant". Sur ce point, même si notre indicateur est différent de celui retenu par l'Institut National de la Statistique Malgache (INSTAT), on constate que notre population possède une bonne maîtrise en français : 35% déclarent avoir une maîtrise "avancée" et 38% une maîtrise "intermédiaire" (Table 49). Ce résultat apparaît en apparence éloigné avec la proportion de francophones dans la population générale. Aujourd'hui, 23,6% des malagasy âgés de trois ans ou plus savent parler français. On constate une répartition

272. Centre national de télé-enseignement de Madagascar

273. "après le bac, je maîtrisais déjà là bureautique et j'avais juste besoin d'un certificat, je l'ai passé à l'épsilon, c'est un centre de formation pour avoir le certificat" (Feno - Annotateur - Sous-traitant C)

très inégale de la compétence linguistique en français, notamment au travers de la fracture entre zones urbaines et zones rurales et selon les classes d'âges : 47,2% des urbains parlent français contre 17,8% des ruraux, et à travers la fracture générationnelle, 29,8% des 15-17 ans et 31,1% des 18-25 ans contre 27,5% des 26 ans et plus (p.19). Ce constat est confirmé par le fait que 58,3% des garçons et 56,5% des filles entre 18 et 25 ans habitant en milieu urbain parlent français contre respectivement 24,9% et 22,9% pour les autres classes d'âge de zone urbaine. Cette fracture géographique est confirmée par le taux d'alphabétisation qui est de 76,7% à Madagascar, 77% pour la langue malgache et 37% pour la langue française avec une forte différence entre les urbains et les ruraux : respectivement 93,1% en malagasy et 63,5% en Français contre 72,7% et 29,6%. En termes de classe d'âge, on observe également la même tendance avec un meilleur taux d'alphabétisation des moins de 26 ans en malagasy et en français (p.31), par rapport aux plus de 26 ans, avec une grosse fracture Nord/Sud (p.36). Par conséquent, notre population se situant quasi intégralement dans la capitale ou aux alentours, il n'est pas étonnant de constater une aussi bonne maîtrise du français.

Il paraît également essentiel de rappeler quelques éléments historiques concernant la place du français dans l'enseignement et dans les interactions entre malgaches. Dans le rapport thématique sur les compétences linguistiques et la scolarisation à Madagascar (Instat, 2021), on note plusieurs points concernant l'apprentissage du français. Premièrement, son enseignement a longtemps été la norme, et ce, jusqu'à la malgachisation entamée en 1975²⁷⁴. L'enseignement du français est également resté très présent ensuite, comme l'explique un rapport de l'Observatoire de la Francophonie, 2016 : *"À Madagascar, au niveau primaire, l'enseignement est bilingue français- malgache. Le malgache est employé dès les premières années de scolarisation. Le français sert de langue d'enseignement pour les mathématiques, les sciences physiques, les sciences de la vie et de la terre et la géographie. Pour les classes secondaires et les universités, la langue d'enseignement est le français sauf dans les départements de langues."* Madagascar se caractérise également par la variété de ses groupes linguistiques : dix-huit ethnies sont présentes dans le pays et parlent un dialecte différent²⁷⁵. Ce point a été souligné par Andry, notre guide et interprète. A l'occasion d'entretiens avec des annotateurs de Fianarantsoa²⁷⁶, qui font partie de l'ethnie des Betsima, il nous indique qu'il n'arrive pas à comprendre le dialecte utilisé, lui-même venant d'Antananarivo et étant issu de la tribu des Merinas. De manière similaire, il nous explique à propos des habitants de la côte est qu'il est incapable de les comprendre, et que le français peut être utilisé pour les échanges entre ethnies parmi ces habitants des villes de la côte est.

274. Ainsi, *"L'ère de la première République de Madagascar où l'enseignement était en langue française pour continuer de former les jeunes malagasy au service de l'Etat à devenir des fonctionnaires. C'est en 1975, sous le régime socialiste que la malgachisation de l'enseignement a été adoptée."* (p.7)

275. *"En plus de cet aspect historique, la population malagasy est répartie en dix-huit ethnies parlant chacune son dialecte. Une des particularités de ce pays est que tous les Malagasy, parlent, comprennent et écrivent une seule langue officielle enseignée à l'école : le « Malagasy ofisialy ». Les dialectes ethniques sont non seulement utilisés à l'école pour expliquer les leçons mais aussi employés pour entretenir les relations entre les enseignants et les enfants."* (p.8)

276. Une ville située à 400 kilomètres de la capitale

Ce résultat vient contredire certains discours entendus à Madagascar, notamment celui du responsable opérationnel du sous-traitant G qui nous explique trouver plus facile de recruter des personnes qui parlent français au début des années 2000 par rapport à aujourd'hui. En effet, cette connaissance de la langue française est très contradictoire avec la politique de malgachisation mise en place depuis 1975. Nos entretiens montrent au contraire que les personnes les plus jeunes arrivent à maîtriser le français à travers toute une variété de stratégies qui correspondent à une volonté de solidifier leurs qualifications. Concernant le niveau d'études des annotateurs de données malgaches, nos résultats sont cohérents avec la littérature qui a souligné que les micro-travailleurs sont relativement plus diplômés que la population générale, quelque soit le pays du travailleur (Berg et al., 2018; Posch et al., 2018)

* * *

Cette première sous-section illustre le concept d'*extractivisme des données* en apportant des éléments de compréhension concernant l'installation des investisseurs étrangers à Madagascar. Le travail de la donnée repose en effet sur des relations de production post-coloniales qui s'appuient sur le développement inégal des pays du Sud global. Face à ces retards de développement, les politiques publiques mises en place par le gouvernement malgache ont comme objectifs d'attirer des capitaux, d'attirer des capitaux et des investissements étrangers.

Nous avons constaté l'existence de deux types d'élites dirigeantes au sein du secteur de la sous-traitance. Tout d'abord les entrepreneurs de développement qui ont un lien avec Madagascar (*entrepreneurs du développement*). Les autres sont les entrepreneurs et salariés français des entreprises de sous-traitance. Pour nous, ces derniers se caractérisent par l'exploitation des marges de la mondialisation (*entrepreneurs des marges du développement*). Ces derniers ont des attitudes complètement différentes vis-à-vis de l'accès aux facteurs de production ainsi que vis-à-vis des objectifs de l'installation de leur production à Madagascar. Les premiers évoquent une rationalité de développement tandis que les seconds investissent davantage la rationalité économique liée aux faibles coûts de la main d'œuvre à Madagascar. Nous considérons que ce secteur illustre une *intégration marginale* de Madagascar désignant cette forme de rationalité économique qui guide les entrepreneurs vers certains pays plutôt que d'autres, pour préserver leurs marges basées quasi intégralement sur ce rapport coût-capacité. C'est le développement du pays hôte, sa position subalterne dans la mondialisation et son intégration aux institutions de l'ancien pays colonisateur qui justifient d'externaliser dans ce pays-là en particulier.

Dans cette partie, nous constatons également qu'il s'agit d'une population bien formée qu'il s'agisse de leur parcours universitaire et même de leur formation professionnelle. Ces résultats confirment le travail de Razafindrakoto, Roubaud, et Wachsberger (2018b) sur le développement de zones franches d'exportation

réservées à ceux qui ont un "réseau à l'extérieur du pays"²⁷⁷. Ainsi, malgré leurs qualifications, les malgaches n'ayant pas eu d'expérience d'études ou de travail hors du pays restent en marge de la création d'entreprises de zones franches qui est uniquement réservées aux entrepreneurs français. Ces résultats suggèrent l'inscription de ce secteur dans une forme de capitalisme racial qui s'appuie sur les pays en marge de la mondialisation moderne.

8.2 Un processus d'assetisation coloniale pour transformer des qualifications coloniales en valeur de subsistance

Le choix de Madagascar correspond à un allongement géographique des chaînes de valeur facilité par l'accès aux facteurs de production. Ce phénomène s'explique par des mécanismes institutionnels, avec la mise en place de stratégies d'attraction des investisseurs, par un meilleur développement numérique du pays, ainsi que par un plus faible coût du travail par rapport à des pays comme le Maroc. Néanmoins, cette question du coût est largement mise en balance avec l'accès ou non à des compétences. Au cours de l'enquête, les entrepreneurs de l'externalisation de services ont largement mis en avant la qualification des travailleurs de Madagascar et notamment des plus jeunes. Comment analyser ce discours en apparence contradictoire entre la recherche de faibles coûts de production et la recherche de qualifications chez les travailleurs ?

Il faut pour cela faire un détour par l'histoire des tâches de saisie de données et par un "premier âge de l'information" qui se caractérise par plusieurs développements technologiques et évolutions sociales liés à la maîtrise de l'information. Gardey (2008) propose "un panorama de ces transformations cognitives et matérielles qui ont façonné les sociétés et les économies occidentales au cours du XIXe siècle et ont conduit à notre monde contemporain" (p.). La chercheuse analyse les développements des machines et des activités sur lesquelles repose la société de l'information. Ces dernières soutiennent le développement de la bureaucratiation, en particulier dans les industries naissantes comme les chemins de fer et le secteur commercial. Une nouvelle catégorie d'emploi apparaît alors : les employés de bureau. Machlup (1972) a étudié le développement de l'économie de l'information et montre que l'emploi dans les activités basées sur la connaissance a été multiplié par trois entre 1900 et 1960 représentant ainsi 29% du PIB américain en 1958²⁷⁸. Cette modernisation du fonctionnement des organisations est permise par la capacité à faire circuler écrits et chiffres (les données), mais repose surtout, selon Delphine Gardey, sur un élargissement des compétences. Cette dernière fait le lien entre l'émergence de la société de l'information et le développement de la *literacy* et de

277. "Malgré la taille limitée du marché intérieur, la distribution de produits de grande consommation constitue également un secteur rémunérateur, où les Karana sont très présents. Le développement de la zone franche d'exportation à partir de la fin des années 1990 attire plus particulièrement les investisseurs étrangers. Les nationaux sont en faible nombre dans ce secteur, l'accès au marché international très concurrentiel étant difficile, réservé à ceux qui ont un réseau à l'extérieur." (Razafindrakoto et al., 2018b)

278. Il a également montré comment certains emplois tels que les "employés de bureau" ont fortement augmenté, passant de 3,02% en 1900 à 14,06% en 1959. De plus, la proportion de femmes est de plus en plus dominante dans cette catégorie de travailleurs : 50/50 en 1940 à 3 fois plus de femmes en 1958, alors qu'elle était de 90/10 à la fin du 20ème siècle.

la numeracy²⁷⁹. Ainsi, elle explique que *”D’abord limités à certains milieux (...) le travail du chiffre et le goût pour les représentations numériques tendent à s’étendre à d’autres sphères de la société (...)”* (p. 283). Cette diffusion d’une *”culture du chiffre”* s’observe alors par *”la multiplication des milieux dont l’activité impose la production de données chiffrées, l’augmentation du nombre de ceux qui exercent une activité de calcul à titre professionnel”* (p.284) ainsi que par l’acculturation du public de plus en plus confronté à des chiffres.

Des machines donc, mais également des employés de bureau, et surtout l’élargissement général des compétences de l’écrit et du chiffre sont les piliers de la société de l’information pré-ordinateur²⁸⁰. On peut constater le fil qui relie cette analyse de l’imbrication homme-machine d’avant l’ordinateur avec l’analyse actuelle du développement du *”travail gratuit”*. Finalement, ce qui rend possible l’automatisation, ce n’est pas seulement l’invention et le déploiement des machines, mais surtout l’élargissement général des compétences dans la population générale qui permet un accès général et facilité à des travailleurs plus qualifiés. Le travail réalisé par les clients, qu’il permette d’acheter des billets de train ou qu’il permette d’alimenter des IA par la publication de données (messages, photographies...) sur les réseaux sociaux, repose alors sur la *literacy*, la *numeracy* (capacité à manipuler des nombres) et la *littératie numérique* des citoyens²⁸¹. L’apport du travail de Delphine Gardey concernant le lien entre développement du travail bureaucratique et la montée en compétences générale de la population nous semble également limiter le constat de déqualification. Par ailleurs, le lien entre compétences et développement d’activités réalisées gratuitement par les consommateurs vient aussi apporter une autre limite à cette analyse. Au contraire, ce serait précisément cette augmentation générale des compétences qui permettrait l’émergence de ces nouvelles organisations de production de valeur. Dès lors, si l’on peut parler de secteurs déqualifiés *”relativement”* au reste de la société, il est nécessaire d’aborder ces nouvelles formes de travail au regard des compétences nécessaires.

Les machines ont stimulé la réorganisation des processus de travail, conduisant souvent à la création de nouvelles professions et à l’élimination des anciennes. Dans cette réorganisation du travail, les femmes constituaient la main-d’œuvre idéale pour s’occuper des tâches que les ingénieurs jugeaient trop fastidieuses²⁸². À ce sujet, Mar Hicks (2018) a réalisé la main-d’œuvre informatique féminine au Royaume-Uni

279. On observe au 19ème et 20ème siècle *”l’augmentation des compétences générales (...) des citoyens”* et *”l’augmentation du nombre de personnes dont le niveau de compétence dans le domaine de l’écrit les autorise à y exercer une profession”* (p.281)

280. Cet aspect du travail de Delphine Gardey vient également contredire une vision largement partagée, y compris par certains sociologues du travail comme Michel Freyssenet d’activités liées à l’informatique qui seraient complètement déqualifiées. Il parle notamment du travail des perforatrices de cartes perforées : *”Le travail de perforation est une des activités les plus abrutissantes que l’on connaisse à ce jour (...). Le codage demande parfois du jugement lorsqu’une information qualitative ne se laisse pas aisément réduire à l’un des cas prévus; mais l’analyste fait disparaître ces aléas, en mettant au point progressivement une codification sans faille”* (p.68). Il indique encore que ces tâches sont dévolues à *”de la main-d’oeuvre féminine sans qualification”*

281. *”Le déploiement des activités bancaires, assurantielles, industrielles et étatiques (...) repose sur cette capacité à traiter des données écrites et numériques en grande quantité (...) Il s’appuie sur l’utilisation intensive d’infrastructures techniques dont les capacités sont aussi humaines”* (Gardey, 2008) (p. 285)

282. *”Dès la Seconde Guerre mondiale, l’informatique se féminise dans de nombreux domaines (...). Les nouvelles machines, capables de remplacer des centaines d’ordinateurs humains, nécessitaient une intervention humaine pour établir les problèmes*

de manière historique. Expliquant d'abord les raisons de l'entrée des femmes sur le marché du travail, elle raconte comment cette main-d'œuvre a ensuite été évincée, ce qui a entraîné un retard de la Grande-Bretagne en informatique. Il s'agissait en effet d'un travail mal payé, sans perspective de carrière, et peu qualifié. Par ailleurs, le *turnover* est favorisé par l'interdiction, d'abord légale puis culturelle, de laisser les femmes travailler après leur mariage. Un autre tournant intervient dans les années 1970, au moment où l'informatisation des administrations et des entreprises devient un impératif stratégique. Ces fonctions devenant importantes, elles ont été progressivement occupées par des hommes. Le travail de Hicks (2017) témoigne donc des rapports de pouvoir sous-jacents à la valorisation des emplois sur le plan des qualifications. Lorsque les femmes étaient engagées de manière massive dans ces activités, ces dernières voyaient leur emploi mal jugé socialement. L'arrivée des hommes dans les métiers de l'informatique a entraîné à une valorisation sociale et donc salariale.

Comme le montrent (Maruani & Nicole-Drancourt, 1989), la numérisation de la presse facilite l'arrivée de femmes recrutées à des postes similaires aux employés du livre, plutôt des hommes, pour des salaires bien moins élevés. L'univers de la presse et de l'édition constitue alors un des secteurs les plus en prise avec les évolutions technologiques. Elles y décrivent les résistances et les oppositions entre les anciens typographes et les nouvelles *OS du tertiaires* au sein de l'imprimerie d'un quotidien français. L'informatisation progressive de l'imprimerie permet en effet à des femmes spécialisées en bureautique d'intégrer des emplois de saisie de textes sur ordinateurs. Ces transformations sociales et technologiques se traduisent pas des conflits importants qui révèlent une tension importante : une éventuelle déqualification des métiers existants et en réponse à cela, l'émergence de nouveaux métiers informatisés captant parfois de nouvelles populations jusqu'alors éloignées de l'emploi. Ainsi, les "typographes" sont des ouvriers du livre qui font valoir un haut degré d'expertise du fait de leur formation. Ce sont eux qui manipulent les caractères et qui composent les documents imprimés sur des chaînes de production au plomb. L'arrivée de l'informatique permet à la direction du journal de recruter des "clavistes", uniquement des femmes qui sont reléguées au rez-de-chaussée de l'imprimerie. Invisibles aux yeux des anciens typographes devenus depuis des "hommes de l'ordinateur", les femmes saisissent au kilomètre, tandis que les hommes anciens typographes deviennent les correcteurs des manuscrits entérinant une prétendue supériorité de leur métier. Cette distinction se traduit par une convention collective moins avantageuse pour les femmes, considérées comme des "employées de bureau" alors que les hommes sont des "ouvriers des métiers du livre".

Cette recomposition technologique, loin d'aller de soi, révèle en fait une organisation du travail basée sur les caractéristiques supposées des populations de travailleurs, en l'occurrence de travailleuses. Ainsi, "la part des déterminismes technologiques apparaît bien faible. Il n'y a pas d'évolution linéaire dans la définition

mathématiques. Sans précédent sexué, le métier d'opérateur d'ordinateur, tout comme les emplois nouvellement créés de "sténographe dactylographe" et de "scanneuse", devient un travail de femme." (Light, 1999)

de la place et des postes occupés par les femmes. Leur dépréciation constante est le produit d'une construction sociale : le refus de traduire leur savoir-faire en qualification." (Maruani & Nicole-Drancourt, 1989) (p.56) Évoquant cette fois l'introduction de l'informatique dans les métiers du secrétariat, les autrices soulignent l'existence d'un "mythe du travail facile". L'introduction des machines bureautiques permet de soulager les employées de certaines tâches, opérant alors un transfert de savoir-faire : les machines font pendant que "l'employée qui l'utilise est reléguée au rang d'exécutante si transparente qu'on en parle même pas" (p. 61). De ce point de vue, le travail des deux autrices rejoint les conclusions de (Cockburn & Ploux, 2004) qui montre également que les hommes ouvriers de l'imprimerie réussissent à préserver leur statut "d'artisan". En se présentant avec des qualités distinctes des femmes qui rejoignent alors cette industrie, ils définissent ce qu'est un travail qualifié de manière à ce qu'il corresponde à leurs qualités.

Les exemples que nous avons donnés précédemment nuancent la thèse du lien entre développement technologique et déqualification. On constate surtout que la technologie peut être l'occasion de mettre en place une division sexuée du travail permettant aux organisations de réduire leurs coûts en dévaluant le travail réalisé par une population alors minorisée sur le marché du travail. Ce "mythe du travail facile" synthétise bien l'enjeu de cette réorganisation du travail : puisqu'il est exercé par des femmes, c'est qu'il ne doit pas être compliqué. Il n'est donc pas nécessaire de lui attribuer une importante valeur sociale et économique. Cette tendance va ensuite se poursuivre avec le recours aux travailleurs de pays en voie de développement. L'externalisation de tâches des pays du Nord vers les pays du Sud global soulève des questions similaires. La première vague d'informatisation des entreprises des années 1970 et 1980 et l'apparition des premiers réseaux d'interconnexions ont permis aux entreprises d'externaliser des activités liées aux données, soulevant déjà des débats sur l'externalisation. L'image ci-dessous illustre cette tendance à l'externalisation du travail des données. Cette iconographie s'inscrit dans une vision "occidentale" valorisant le prix et les performances de ces petites mains féminines.



FIGURE 37 – L'externalisation de la saisie de données aux États-Unis dans les années 70 (source : Computerworld the Newsweekly for the Computer Community 1974-07-17 Vol 8 Iss 29)

Loin du déterminisme technologique affiché dans les conclusions de certains articles scientifiques et médiatiques promettant la fin du travail, on constate que la littérature en sociologie du travail a mis en avant plusieurs arguments pour relativiser cette "destinée manifeste" vers la fin et la déqualification du travail. Dans cette section, nous abordons cette question sous l'angle de la construction et de la valorisation ou de l'absence de valorisation des qualifications par les entreprises de sous-traitance et les donneurs d'ordre finaux. Nous proposons de traiter la question sous l'angle de deux concepts que nous développons à l'aide de notre terrain d'études. Il s'agit d'abord de la "valorisation coloniale" qui fait référence au travail de Padios (2018) à propos des centres d'appels aux Philippines et du processus de valorisation de la proximité culturelle et linguistique par les entreprises de sous-traitance. Nous proposons d'étudier, "l'assétisation coloniale" ("colonial assetization") en faisant référence aux travaux de Jarrett (2022) qui analyse l'assetization dans le travail des plateformes sous l'angle de la construction "d'actifs" donnant plus d'agentivité à certaines catégories de travailleurs des plateformes comme les influenceurs²⁸³, en comparaison avec des chauffeurs Uber et des travailleurs des données opérant sur la plateforme Appen. Pour nous, ces deux concepts se répondent et sont complémentaires, car ils illustrent une même dynamique de qualification des travailleurs. La valorisation coloniale renvoie plutôt à la dimension de la qualification sous l'angle du jugement social des qualifications par les entrepreneurs de la sous-traitance alors que "l'assétisation coloniale" permet d'étudier la "colonialité" à travers les stratégies individuelles d'apprentissages des travailleurs susceptibles de favoriser l'émergence du secteur tout en facilitant l'accès de ces derniers au marché du travail. Pour ces derniers, la présence et le développement de ce secteur les pousse à développer des stratégies d'acquisition de compétences coloniales ("actifs coloniaux") qu'ils vont pouvoir *in fine* transformer en valeur de subsistance, soit pour travailler dans le secteur de l'externalisation, soit pour émigrer. L'objectif de cette section est alors d'articuler la tension entre, d'un côté, l'agentivité des travailleurs qui vont développer ces compétences dans leur parcours et, de l'autre la reconnaissance ou non des qualifications coloniales.

8.2.1 Le jugement des entrepreneurs de la sous-traitance sur les qualifications des travailleurs. Analyser la tension entre captation et valorisation des qualifications et minimisation de la contribution des travailleurs

Il convient tout d'abord de souligner que les travailleurs malgaches font en effet face à des freins importants dans l'entrée sur le marché du travail, notamment pour les jeunes qui ont le bac ou qui sont diplômés et qui habitent la capitale malgache (INSTAT, 2022). Ainsi, les citoyens ayant suivi un enseignement secondaire

²⁸³. Elle considère que "Il s'agit d'une position avec un plus grand sens de l'action que celle attribuée au travailleur aliéné par la marchandisation de sa force de travail. Ce pouvoir n'est peut-être pas celui associé au travailleur non aliéné ou au collectif de classe auquel aspire la critique marxiste orthodoxe, mais il reste une forme d'intégration inaliénable et une source de résistance. Cela signifie que si le travail numérique peut (...) impliquer l'exploitation des subjectivités des travailleurs, il est possible de dépasser les estimations entièrement négatives de cette pratique tout en continuant à critiquer son rôle économique. Comprendre l'incorporation de la subjectivité dans le capitalisme en tant qu'"assetization", plutôt que marchandisation, permet à la vie désaliénée et significative de se manifester (...)"

représentent 59,9% des chômeurs urbains, contre 4,4% pour ceux qui n'ont aucun enseignement et 13,9% pour ceux qui ont suivi un enseignement supérieur. Comme l'explique Fanino de l'EDBM, il y a une forte restriction pour les jeunes dans l'accès au marché du travail ce qui incite les jeunes malgaches qualifiés à rejoindre ces entreprises de sous-traitance ²⁸⁴.

Les entreprises de sous-traitance disposent d'une force de travail variée essentiel à sa pérennité, car comme l'explique Miky, ancienne directrice des ressources humaines dans une grande entreprise, il n'y a pas de profil-type ²⁸⁵. La variété de ces besoins pousse donc les entreprises d'externalisation à mettre en place des stratégies dédiées d'acquisition de compétences ou de qualifications présentes chez les travailleurs. C'est le cas dans une entreprise étudiée, où le service RH a une stratégie concrète qui consiste à entretenir un vivier de travailleurs. La nécessité de diversifier les profils s'accompagne alors de stratégies de recrutement spécifiques selon les besoins de certains projets lorsqu'ils requièrent des qualifications expertes et spécifiques.

Encadré n°65 : Trouver des qualifications variées au sein d'un vivier de travailleurs

”Les « people » correspond au fait d'être capable de trouver des personnes qui ont les compétences pour travailler sur le projet et le niveau de compétence requis [qui] dépend du niveau de complexité du projet (...). Pour faciliter ça, parce qu'on a une variété de projets et de besoins, on entretient ce qu'on appelle un vivier de compétences, c'est-à-dire qu'on sait qu'on a plus de projets dans tel domaine ou tel domaine, ou alors on est en train de travailler sur une opportunité dans tel domaine, le service RH va commencer à voir qui pourrait être intéressé et qui a de l'expertise (...)”

Victor - CTO - Sous-traitant B

La spécificité de certaines thématiques conduit d'ailleurs les entreprises à recruter des profils beaucoup plus experts ²⁸⁶. Il devient ainsi plus clair à mesure que l'on analyse les entretiens réalisés avec les dirigeants d'entreprises, que l'annotation de données et l'externalisation constituent en fait un *business des qualifications*. La réussite de ces firmes dépend largement de leur capacité à recruter les personnes les plus qualifiées, aux coûts les plus faibles et à les former.

284. : ”Pourquoi ça explose, parce que déjà y a beaucoup de jeunes et comme je te l'ai dit après le bac, bah le monde du travail autre que les entreprises franches est quand même assez restreint et du coup les investisseurs profitent de cette situation que y a beaucoup de jeunes, actuellement qui cherchent du travail et qui ont quand même les capacités de travailler dans ces entreprises franches.” (Fanino - EDBM)

285. ”Tu n'auras jamais de profil type, pourquoi, parce que c'est des métiers qui nécessitent des compétences très diversifiées. Le challenge, c'était de concentrer sur un plateau toutes les compétences.” (Miky - RH - Sous-traitant C)

286. ”(...) il y a des cas spécifiques (...). Par exemple dans le domaine médical, (...), mais l'accès aux compétences médicales est compliqué, ce ne sont pas des ressources qui courent les rues ; deuxième point, il faut comprendre les sujets médicaux, moi je ne suis pas médecin (...). On a un centre d'expertise avec des radiologues, etc. qui nous conseillent, qui nous aident à comprendre les sujets et à structurer et à les valider. En fait, ce sont eux qui disent « ce que vous allez faire, c'est effectivement conforme à la bonne pratique et à la bonne manière dont ça devrait être fait” (Victor - CTO - Sous-traitant B)

Encadré n°66 : L'annotation, un business des qualifications

”Forcément, si on doit travailler sur des sujets très techniques, on va essayer de trouver des personnes qui ont une formation plutôt technique/ingénieur. Si c’est un sujet dans l’agronomie, on va prendre des gens qui ont des connaissances là-dedans. Ça, c’est la formation initiale, donc c’est un premier critère (...). Il y a ensuite une formation donnée par [notre entreprise] sur l’IA. (...) Et puis après, troisième niveau de formation, ce sera la formation spécifique au projet. On donne beaucoup d’informations sur le contexte du projet (...) ce qui leur permet de devenir autonomes sur le projet (...)”

Victor - CTO - Sous-traitant B

Nos résultats suggèrent qu’à travers ce point en particulier, la question de la valorisation des qualifications des travailleurs est en tension. Même si les entreprises se défendent d’avoir besoin de recruter des travailleurs qualifiés, leur modèle d’affaires repose entièrement sur ce facteur. Au sein d’une entreprise de taille plus modeste (sous-traitant G), nous avons également pu constater que le dirigeant français a une conscience aiguë des compétences de ses travailleurs. Sa manière d’attribuer les projets témoigne d’ailleurs d’une vraie réflexion concernant les qualifications des travailleurs et la manière de les faire évoluer professionnellement.

Encadré n°67 : Une conscience très claire des qualifications des travailleurs

”(...)On a les profils qui sont très bien formés, qui parlent très bien français et qui se débrouillent vraiment en anglais. Les chefs d’équipe tiennent des conversations avec des clients anglophones aussi, et ils sont capables d’écrire des mails, de bien comprendre rapidement les problématiques et d’avoir des fichiers de suivi. Aujourd’hui (...) j’ai quatre profils comme ça (...). J’ai quelques profils intermédiaires qui sont reviewers, qui ont parfaitement compris la tâche et [sont] capables de voir ce que font les autres, même s’ils ne sont pas toujours très bons en communication. (...) Et ensuite on a les annotateurs de base, qui forment une grande partie des effectifs (...) à Tana où les profils sont variés. Beaucoup parlent plutôt bien français, donc ils sont potentiellement surqualifiés pour le job. ”

Pierre - CEO - Sous-traitant G

Dans l’encadré qui suit, le discours de Louis sur la qualification des travailleurs illustre bien cette dichotomie entre qualification et minimisation des compétences. D’un côté, il considère que les travailleurs malgaches s’adaptent bien à ces activités d’annotation, puisqu’*”il n’y a pas besoin de réfléchir”*. Par ailleurs, il naturalise

les travailleurs de Madagascar en considérant qu'il s'adaptent particulièrement bien à des processus de travail standardisé, ce qu'il justifie par l'histoire du pays. Pour lui, il s'agit d'une conséquence des oppressions successives subies par le peuple malgache ²⁸⁷.

Encadré n°68 : Une évolution des qualifications de la main d'oeuvre des entreprises de services

Et [c'est ce qu']ils viennent chercher (...) C'est con ce que je vais dire, mais pour ces métiers-là (...) j'aime pas dire ça (...) mais il faut pas réfléchir au final (...). C'est des process qui doivent être appliqués par les humains. (...) Et c'est ce qui coûte le plus cher. Après c'est la formation des gens à pouvoir le faire. Mais une fois que c'est fait, les gens, ils le font aussi bien que n'importe où et c'est ça qui est extrêmement bien(...) c'est que tous les Malgaches sont des éponges à comprendre (...) [et très bon] dans l'application! Et ça, c'est malheureux à dire (...), mais c'est un pays qui a été comme vachement opprimé, que ce soit par [les] colons mais aussi par les rois et reines (...) ! Et donc ils ont été plus formé à faire en sorte de bien respecter les choses, de bien faire son travail. Peut-être moins (...) de sortir du cadre, ce que maintenant les clients attendent aussi de Madagascar (...)

Louis - Vice-président - Sous-traitant C

A l'inverse, il explique également rencontrer des travailleurs particulièrement compétents pour maîtriser les règles des projets. Il constate également une évolution récente au sein de la population des travailleurs de l'externalisation qui sont de plus en plus qualifiés ²⁸⁸. Nos données confirment cette évolution, puisque les travailleurs les plus récemment entrés sur le marché du travail sont aussi les plus diplômés ^{Table 47}, ce qui influence la stratégie de son entreprise : *"Les principaux axes sur lesquelles on tourne, c'est la formation, former les jeunes, même les anciens pour (...) qu'on puisse répondre aux besoins grandissant du marché."* (Louis - CEO - Sous-traitant F) On comprend de cet extrait que la piste principale de croissance de ces entreprises réside dans les qualifications des travailleurs afin qu'ils réalisent les activités actuellement présentes dans ces entreprises, mais également pour qu'ils développent des qualifications permettant à l'entreprise de vendre des activités à plus forte valeur ajoutée (par exemple : le développement web).

287. Notons que ce supposé "trait culturel" constitue une des catégories d'analyse de Razafindrakoto et al. (2018a) lorsqu'elle évoque une population marquée par un "tabou de la violence". Ainsi, la "stabilité sociale de long terme de la société malgache repose sur une théologie politique faisant du recours à la violence un tabou et du respect des dirigeants et de l'ordre hiérarchique une donnée essentielle"

288. "Mais on le voit que la nouvelle génération, c'est complètement différent. Ce sont des gens qui ont accès à Internet, c'est des gens qui ont une culture, des formations (...). Et puis l'école, elle est coûteuse quand même. Ce qu'il y a après le bac, avant, il y a encore dix ans en arrière, il n'y avait pas d'université très ... on va dire très reconnues, et les gens ils partaient souvent (...) pour aller faire leurs études [ailleurs]." (Louis - CEO - Sous-traitant F)

	[1,4]	[4,7]	[7,25]	Total ¹
Pas passé dans le supérieur	20.8	24.5	54.7	100
Passé dans le supérieur	34.9	37.1	28.0	100
Ensemble	31.6	34.2	34.2	100

¹Réponses en pourcentage, Fisher.test = 0.002144

TABLE 11 – Niveau d'études selon l'ancienneté sur le marché du travail

On retrouve également cette tension à travers l'entretien réalisé avec Yannick, qui dirige les opérations d'une grande entreprise d'annotation. D'un côté, il évoque l'entreprise comme relevant d'un projet social, ce qui passe par l'embauche de travailleurs faiblement qualifiés, précisant que le niveau d'études requis par le poste est d'avoir le bac ²⁸⁹. Pourtant, l'analyse de la page de recrutement de l'entreprise sur Facebook montre que les qualifications requises dépassent ce niveau de diplôme. Outre la maîtrise du français, des postes d'agents experts requièrent a minima un Bac +2, tandis que d'autres emplois qui portent sur le contrôle qualité nécessitent des Bac +5. Cependant, ce simple critère constitue un niveau de qualification très discriminant à Madagascar, étant donné que seule 6,5% de la population malgache a été au lycée. Le fait que l'entreprise ait par ailleurs recours à des partenariats avec des universités nuance le caractère "déqualifié" du travail demandé ²⁹⁰.

Ces entreprises profitent donc d'une disponibilité importante de la main d'œuvre qualifiée dans la capitale malgache. Ce résultat est très contradictoire avec ce qui a été observé par Padios (2018) à propos du développement des qualifications aux Philippines : "Les organisations sectorielles tentent de remédier à cette pénurie de talents de différentes manières, notamment en intensifiant les efforts visant à générer du biocapital, c'est-à-dire la force de travail et donc les bénéfiques potentiels des capacités cognitives et émotionnelles utilisées dans les centres d'appel." (p.78). Dans le cas de Madagascar, la forte présence de jeunes travailleurs qualifiés ayant du mal à s'insérer sur le marché de l'emploi facilite le recrutement de travailleurs qualifiés. Il faut néanmoins nuancer ce résultat au regard des efforts mis en place par les entreprises pour développer

289. "Depuis le début, l'objectif est d'être une société à impact social. [Le CEO] a toujours été convaincu que ce travail d'annotation pouvait être fait par n'importe qui et qu'on n'avait pas besoin de personnes avec Bac+2 ou +3 (...). Donc on a recruté des personnes qui venaient de milieux défavorisés. Le seuil de recrutement était d'avoir le Bac (...). On fait bien sûr passer des tests à tous les candidats (...) on a toujours essayé de recruter des personnes et de les former en annotation de base, computer vision, dessiner un cadre autour d'une voiture ou d'une personne. Ce sont des tâches qui ne nécessitent pas de connaissances approfondies. Il ne faut même pas maîtriser l'outil informatique, il suffit de savoir manipuler une souris et un clavier. On a donc ouvert pas mal de portes à des personnes qui autrement auraient eu des difficultés à trouver du travail dans le domaine de l'informatique." (Loik - COO - Sous-traitant E)

290. "On a signé pas mal de partenariats avec des facs ici. Par exemple, on a signé un partenariat avec des gens qui travaillent à l'université d'Antananarivo, surtout pendant le COVID, parce que tous les étudiants cherchaient du boulot. Donc on a noué pas mal de liens avec des universités parce qu'il y avait des étudiants qui venaient bosser chez nous à mi-temps ou à temps plein. Ces personnes sont ensuite restées, donc c'est un peu comme ça qu'on a fait." (Dinaso - CEO - Sous-traitant E)

des viviers de travailleurs qualifiés et des partenariats avec des institutions académiques et de formation. Conscientes des compétences nécessaires et en même temps promptes à minimiser leurs besoins, les entreprises de sous-traitance ont une position très ambiguë vis-à-vis des travailleurs - entre reconnaissance ou non des qualifications des travailleurs - qui suscite des interrogations quant à ses effets sur les stratégies des travailleurs.

Si certaines compétences semblent sous-valorisées, les qualifications reposant sur le rapport à langue française que nous appelons "coloniales" le semblent encore davantage et constituent même un facteur d'attractivité publiquement abordé par l'EDBM dans ses plaquettes commerciales destinées à attirer les investisseurs étrangers. Nous défendons l'hypothèse qu'un processus de valorisation des qualifications coloniales se met en place dans le secteur de la production de données. Nous nous appuyons pour cela sur le concept de "colonial recall" (Padios, 2018), un processus que l'auteur définit comme la valorisation économique des connaissances et des qualifications coloniales des travailleurs des Philippines par les entreprises américaines. Pour nous, cet aspect se traduit par trois éléments : les stratégies d'implantation des entreprises, la valorisation des compétences linguistiques et la valorisation des institutions d'apprentissage présentes à Madagascar.

Tout d'abord, les entreprises d'externalisation tiennent compte de ces qualifications jusque dans leur stratégie d'implantation. Nos entretiens font état de plusieurs cas d'entreprises qui développent des stratégies délibérées d'accès à certains pays du Sud global en fonction des langues recherchées par leurs clients. Par exemple, le vice-président du sous-traitant C explique que de nombreuses entreprises sont actuellement en train d'essayer de s'implanter en Angola, car la langue portugaise est de plus en plus demandée. Cette stratégie de diversification des qualifications linguistiques se traduit également au sein du sous-traitant B dont les stratégies d'implantations servent à accéder à différentes langues et cultures.

Encadré n°69 : Des projets d'implantations basés sur les compétences linguistiques des travailleurs

(...) ce qui va justifier le choix d'un pays (...), c'est aussi l'accès aux compétences. Nous on essaie [d']accéder à des cultures différentes, à des langues différentes, à des gens qui ont des formations différentes pour (...) élargir notre panel de solutions. On sait que c'est intéressant d'être aux Philippines parce qu'on peut accéder plus facilement à des gens qui ont une expertise dentaire tandis qu'à Madagascar c'est plus compliqué. Donc ça nous permet d'accéder à des projets d'imagerie médicale dans le dentaire (...). A contrario, il y a des projets qui vont être plus en Thaïlande sur du NLP dans des langues asiatiques parce que (...) c'est un peu un hub au centre de l'Asie du Sud-Est. Donc il y a cette notion d'accès à des nouveaux types de compétences.

Pour lui, l'intérêt de Madagascar réside dans la capacité du pays à former des travailleurs doués dans plusieurs langues. Il évoque une facilité à trouver des travailleurs qui parlent l'allemand, grâce à l'implantation d'espaces d'apprentissage de l'allemand comme l'Institut Goethe, qui témoigne d'ailleurs de stratégies de *soft power* des pays occidentaux²⁹¹. Par ailleurs, l'existence de plusieurs diaspora à Madagascar, notamment chinoise, japonaise et indienne s'accompagne de l'existence d'instituts de formations dans des langues moins évidentes à trouver : *"Il y a pas mal d'écoles de langue japonaise. Les Malgaches sont très intéressés par la culture japonaise. On n'a pas eu de mal à trouver des personnes qui avaient un très bon niveau de japonais. Même chose pour le mandarin. Il y a une forte présence chinoise à Madagascar, donc il y a pas mal d'instituts et d'écoles qui donnent des cours de mandarin."* (Loik - COO - Sous-traitant E)

Le rapport à la langue française est également un aspect central de l'externalisation d'activités liées au numérique et aux données. Que ce soit en marchant à Antananarivo ou bien lors de nos déplacements entre notre logement et les locaux des entreprises étudiées, nous avons constaté la forte présence de la langue et de la culture française de par l'existence de nombreux centres de formation professionnels à la langue française, à la langue anglaise et à l'informatique. Cette stratégie est encore davantage poussée à l'extrême lorsqu'on considère que le secteur de l'externalisation a explosé à Madagascar avec l'installation de nombreux centres de relations clients. L'arrivée de ces entreprises a été justifiée à de nombreuses reprises lors de nos entretiens par un accent malgache moins présent dans les interactions en français que dans les pays du Maghreb ce qui témoigne d'un *"parler français"* vue positivement par les entrepreneurs de la sous-traitance qui exprime également la relation coloniale entre la France et Madagascar²⁹².

En plus du coût du travail, l'industrie de l'externalisation profite donc de travailleurs qui parlent couramment le français. Ils l'ont appris à l'école, par Internet et à travers le réseau des Alliances françaises. Comme l'explique l'ancien directeur de l'Agence Universitaire de la Francophonie : *"Il y a un gros réseau d'alliance française (29), un des plus gros en Afrique. Sur les principales villes de Madagascar, il y a des AF plutôt actives avec des gens qui viennent apprendre le français et le malagasy."* Or, il convient de préciser que cette institution d'apprentissage du français a été initialement créée en 1883 afin de renforcer la colonisation par l'extension de l'utilisation du français par les populations colonisées (Horne, 2017)²⁹³. Ces perspectives de formation peuvent être considérées comme des substrats de la période coloniale française. L'Alliance Française et les

291. *"la relation entre l'Allemagne et Madagascar qui font qu'il y a pas mal de gens qui ont des compétences linguistiques en allemand, on passe par l'Institut Goethe (...), on rentre dans [leur] réseau et on essaie de voir si des gens peuvent être intéressés et qui ont un niveau d'allemand correct."*(Victor - CTO - Sous-traitant B)

292. *"Notre avantage, c'est qu'on a moins d'accent, de l'aveu du PDG de Webhelp, les malgaches ont une manière de parler moins agressive."* (M.N.R - Adjoint au service économique à l'Ambassade de France à Madagascar)

293. *"Il est clair qu'une partie de l'histoire des réseaux culturels et humanitaires précoces de la France est enracinée dans ses convictions colonialistes et ses visées impériales."*

autres centres de formation continuent à assurer un "service après-vente" de l'appartenance au monde francophone, ce qui illustre de manière particulièrement saillante la relation post-coloniale. L'apprentissage du français via une institution coloniale devenant le levier d'adaptation d'une main-d'oeuvre étrangère aux besoins des entreprises françaises.

Or, ces qualifications linguistiques sont largement valorisées par les entrepreneurs de la sous-traitance, puisqu'ils considèrent que c'est qui leur permet de vendre des services liés à la l'écrit et à la dimension conversationnelle²⁹⁴ comme c'est le cas dans l'annotation de données. Pour les dirigeants de ModeAI le français constitue par exemple un terrain commun de discussion pour éviter les ambiguïtés qui sont liées à la description des concepts à annoter²⁹⁵. C'est donc bien la capacité des travailleurs de Madagascar à maîtriser la langue française et à comprendre les enjeux culturels français avec un niveau de français suffisant pour être capable de discuter des ambiguïtés de l'annotation de données qui constitue un facteur décisif de la présence du secteur dans le pays.

Comment analyser cette contradiction apparente entre la reconnaissance et la minimisation des qualifications? Tout d'abord, il est important de souligner que la reconnaissance de *compétences coloniales* telles que la "neutralité de l'accent" par les entreprises de sous-traitance procède d'un processus plus large de désenclavement des travailleurs vis-à-vis de leur identité personnelle. Une directrice de production française de l'entreprise C évoque particulièrement bien cette facette des "qualifications post-coloniales" en indiquant que les travailleurs qu'elle recrute oublie²⁹⁶ de signifier sur leur CV qu'il parle malgache (Aneesh, 2015). Ainsi, le "rappel colonial" ne fonctionne pas uniquement comme un moyen de valorisation des qualifications des travailleurs, mais aussi comme un *rappel de leur passé colonial*. De par leur position d'employeur, les entreprises contribuent alors à façonner les identités des travailleurs et également à structurer leurs formations. Padios (2018) parle également d'une "rhétorique de la fiabilité" à propos des discours portant sur la valorisation des qualifications coloniales Pour la chercheuse, "Il est possible de lire la rhétorique de la fiabilité comme un autre exemple dans lequel les courtiers en main-d'œuvre et les élites, qui cherchent désespérément des investissements dans le pays (...), essentialisent l'identité philippine pour s'assurer un avantage comparatif sur le marché mondial. Ainsi, le rappel colonial implique une évocation instrumentale du passé colonial qui

294. "on a quand même la chance pour le marché français et le marché mondial je dirais : à Mada, on maîtrise la langue française qui est pas simple, mais on maîtrise de plus en plus [d'autres langues], il y a beaucoup de gens qui parlent l'anglais à Madagascar et ça, ça rapporte. Parce qu'il y a d'autres pays, il y a souvent une ou deux personnes dans l'entreprise qui y arrivent ... aujourd'hui on peut se retrouver en production avec des gens qui parlent anglais (...)" (Louis - CEO - Sous-traitant F)

295. "J : Nos modèles de données à la base ils étaient en anglais, ça parlait en anglais. Maintenant c'est vrai que en français c'est mieux. A : on est tellement sur des trucs des fois difficile à expliquer, même si on parle anglais c'est pas un souci, mais nous notre niveau d'anglais dans cette expertise de la fashion on ne l'a peut-être pas, parce que on n'est pas non plus des spécialistes ... je ne suis pas sûr que les prestataires aient le niveau d'anglais nécessaire pour répondre. Donc à un moment il y a trop d'ambiguïtés en fait." (Jean et Matthias - CEO - ModeAI)

296. "ils oublie^t qu'ils parlent malagasy, c'est pas une compétence, les recruteurs oublie^t ça, je vous dis ça car ça m'avait choqué quand j'ai fait des entretiens, ils mettent pas en avant le fait qu'ils parlent malagasy" (Sous-traitant C)

l'aseptise pour les besoins du capital mondial et du projet national néolibéral." ((Padios, 2018) p.83). Dans le cas de Madagascar, cette "rhétorique de la fiabilité" se traduit selon nous par une "*rhétorique des qualifications coloniales*" par laquelle les entreprises de sous-traitance se retrouvent parfois à "*mettre l'accent sur la complexité du travail et la variété des compétences (cachées) nécessaires dans les activités de service interactives*" (Beerepoot & Hendriks, 2013), tout en essentialisant l'identité des travailleurs malgaches à travers ce qu'ils peuvent fournir comme qualifications nécessaires à la production de services d'annotations.

Nous proposons donc de plutôt de parler d'un "*paravent des qualifications coloniales*" pour désigner cette tension dans la valorisation des travailleurs et qui illustre selon nous le "mythe du travail facile" identifié par Maruani et Nicole-Drancourt (1989). En effet, au-delà de la "fiabilité" dans la délivrance de services informatiques qui témoigne de qualifications importantes dans l'absolu, les travailleurs possèdent également d'autres qualifications dans d'autres domaines qui sont, tantôt ignorés, tantôt valorisés par les entrepreneurs de la sous-traitance et de l'IA lorsque des clients ou des projets requièrent ces compétences. Dans notre prochaine sous-section, nous verrons d'ailleurs que derrière ces qualifications spécifiques se cache en fait un niveau parfois important de formation initiale pour les jeunes de la capitale malgache. Pour nous, cela éclaire autrement l'idée d'un processus de déqualification perçu lié aux emplois induits par l'IA, puisque l'on est en présence de travailleurs dont les compétences autres que coloniales sont niées alors même qu'elles sont le signal de leur qualification à travailler dans une activité qui nécessite pourtant des compétences. Nous verrons d'ailleurs que le développement des qualifications coloniales s'inscrit dans des stratégies d'émigration liées à la situation de Madagascar.

8.2.2 Les stratégies d'apprentissages des travailleurs de l'IA malgaches : une volonté d'échapper à leurs conditions par le (*virtual*) migrant labor

Si les entreprises de sous-traitance profitent des qualifications disponibles à Madagascar, à quoi cela correspond en matière de développement professionnel des travailleurs? Pourquoi ces derniers mettent-ils en place des stratégies de construction de ces qualifications coloniales? Dans cette sous-partie, nous abordons les stratégies d'apprentissages mises en place par les travailleurs afin d'améliorer leur vie personnelle et professionnelle. Notre enquête montre que les travailleurs investissent une part substantielle de leur formation dans la maîtrise des compétences linguistiques. Or, ces stratégies s'inscrivent dans le cadre de stratégies individuelles de mobilité ascendante, souvent contrastées, soit dans le secteur de l'externalisation soit dans la perspective d'une émigration en Europe.

Dafy illustre particulièrement bien les deux facettes de cette tendance. Ce dernier raconte qu'il a essayé d'obtenir une certification en langue française à l'Alliance Française. Son objectif est alors de déménager en France pour faire ses études. Faisant face à des difficultés financières, il se retrouve ensuite à travailler dans le secteur de l'externalisation très fortement demandeur de ce type de qualifications incarnant ce que

(Aneesh, 2006) appelle la migration virtuelle c'est-à-dire "un nouveau régime de travail dans lequel le travail se déplace et migre sans le corps du travailleur." (p.68)

Encadré n°70 : L'apprentissage du français comme stratégie de migration

(...)X : comment est-ce que tu t'es formé en français ?

D : c'était il y a beaucoup d'années (...), à l'école. C'était parce que je communiquais le plus souvent avec des enfants (...) je côtoyais des gens qui parlaient la langue. Et aussi après le bac j'ai pris mon DELF B2 (...) c'est un certificat de langue

X : comment on fait pour l'avoir ?

D : on passe l'examen à l'alliance française

X : pourquoi tu as voulu la passer ... ?

D : c'était parce que je voulais travailler en France, j'ai voulu continuer mes études en France. Mais ça n'a pas abouti pour des raisons financières

Dafy - Annotateur - Sous-traitant G

C'est également le cas d'Eve, qui nous explique qu'elle souhaitait initialement "continuer [ses] études en France", mais qu'elle a été obligée d'arrêter "à cause du budget familial". Elle nous raconte avoir appris le français "à l'école" où, "les profs enseignent tout en français.", car "c'est le règlement ici à Madagascar". Ce qui est particulièrement remarquable dans la narration de son apprentissage du français, c'est qu'elle a pleinement intégré la nécessité d'apprendre cette langue "parce qu'il y a beaucoup d'élèves qui vont partir à l'étranger tout comme ça pour étudier ou travailler." (Eve - Annotatrice - Sous-traitant G). Fanny, une annotatrice opérant sur la plateforme AIEthic nous raconte son parcours pour partir étudier en Allemagne depuis le Cameroun et qui constitue alors son projet professionnel principal, qu'elle finance à l'aide de son travail sur la plateforme²⁹⁷. C'est également le cas pour Hejo rencontrée dans une entreprise malgache d'annotation qui explique qu'elle a "étudié l'anglais et l'allemand" afin de "chercher du travail" en Allemagne, ce que la crise sanitaire l'a empêchée de faire. Coralie, une autre annotatrice malgache continue quant à elle à apprendre l'allemand afin d'"avoir un certificat international et de devenir jeune fille "au pair en Allemagne". Le cas de Mélodie, également annotatrice au Cameroun, illustre la volonté de mobilité à long terme qui n'est pas uniquement centrée sur le fait de faire des études à l'étranger, mais bien d'y travailler²⁹⁸.

297. "En fait actuellement, j'ai fait des cours de langues allemand, car je veux faire dans la médecine, mais je voulais les faire en Allemagne, car là-bas on nous offre des bourses partielles, il me faut des cours de langue et il me faut payer la procédure ici. Actuellement, j'ai fait les cours de langue jusqu'au B2, j'ai réuni un peu d'argent, c'est pas encore assez" (Fanny - Annotatrice AIEthic)

298. "Moi, (...) ... mon rêve c'était de faire ... la médecine, de travailler dans la santé. Dans le projet que j'avais proposé, c'était pour faire les études en Allemagne, Voilà pourquoi je suis des cours d'allemand, parce qu'il y a des écoles en Allemagne qui offrent des bourses pour les infirmiers." (Mélodie - Annotatrice AIEthic)

Ces stratégies d'émigrations alternatives à la France se vérifient encore davantage à travers l'unique cas d'émigration réussi rencontré durant notre terrain. Sandrine, est annotatrice au moment de notre entretien et constitue alors notre informatrice privilégiée pour interagir avec d'autres annotatrices camerounaises. Nous échangeons ainsi à plusieurs reprises via LinkedIn, ce qui me permet d'en apprendre plus sur son évolution personnelle et professionnelle postérieure à notre première discussion. Au moment de l'entretien, Sandrine est une étudiante, qui réalise des tâches sur AIEthic, en parallèle de ses études. Elle souhaite déjà partir de son pays pour étudier à l'étranger, notamment en France, mais c'est finalement en Belgique qu'elle réussit à s'expatrier²⁹⁹.

Pour les autres annotateurs, ces années d'apprentissages se heurtent au durcissement des politiques migratoires occidentales (notamment en France). Dès lors, comment réinvestir ces qualifications en restant dans leur pays d'origine ? Que faire de ces années de pratique ? Nos résultats suggèrent que cette stratégie d'émigration et de développement professionnel contrariée leur permet alors de constituer des "actifs coloniaux" qu'ils peuvent mobiliser pour rejoindre le secteur de l'externalisation très demandeur de ces qualifications. Néanmoins, cette constitution d'"actifs" ne s'intègre pas uniquement à des stratégies d'émigration. Certains travailleurs étudient des domaines comme l'informatique et les langues avec comme unique perspective de rejoindre directement le secteur de l'externalisation. C'est le cas pour Yanni qui occupe un poste de manager dans l'entreprise A au moment de l'entretien. S'il commence son apprentissage de manière autodidacte, il décide finalement de réaliser une formation à l'Alliance Française afin de pouvoir postuler à des emplois dans le secteur de la sous-traitance.

Encadré n°71 : L'apprentissage du français comme stratégie de *virtual migration*

Q : « Est-ce que tu peux nous raconter comment tu as appris le français ? »

R : « (...) je n'avais pas beaucoup étudié la langue française, mais j'ai regardé des documentaires et des séries françaises. Et en 2015, j'ai fait une formation en français à l'Alliance française de Tananarive (...) Il y avait beaucoup de modules (...), mais faute de moyens, je n'ai pu accomplir que l'un des modules. Ça consiste à s'exprimer en français et éviter d'être timide quand on parle avec quelqu'un (...). Et aussi avoir un diplôme si on veut travailler dans une boîte qui parle en français. Avec une formation en français il y a beaucoup d'avenir, parce qu'à Madagascar il y a beaucoup de boîtes qui font de l'appel en externalisation. C'est pour ça que j'ai effectué ma formation en langue française. »

Yanni - Chef de projet - Sous-traitant A

299. "Salut Clément, désolée de te déranger j'espère que tu vas bien et que ta thèse s'est très bien terminée. Si je t'écris c'est parce que je vis maintenant en Belgique mais je suis actuellement à la recherche d'une école ou université et d'une entreprise en France où je pourrais faire un Master en développement d'applications en alternance en contrat d'apprentissage donc si tu as des pistes s'il te plaît passes les moi sinon pas grave. Bonne soirée à toi!" (Sandrine - Annotatrice AIEthic)

C'est aussi le cas pour Jean, qui, au moment où nous le rencontrons travaille comme annotateur dans le secteur informel, après avoir occupé plusieurs emplois dans le secteur de l'externalisation. S'il n'a pas suivi de formation universitaire, il a néanmoins fréquenté plusieurs formations au sein du CNTMAD : "Après le bac j'ai suivi des cours de langues étrangères, français, anglais, chinois aussi, j'ai un peu des notions. J'ai commencé à étudier les langues étrangères, au CNTMAD." En plus de cela, il a également "suivi des cours en informatique (...) dans un établissement de formation". Il justifie sa stratégie d'études en expliquant que vu "la situation actuelle", il a considéré qu'il "serait mieux de maîtriser les langues étrangères" afin de "trouver un boulot facilement" et notamment dans tout "ce qui est basé sur la langue étrangère, comme la rédaction [de contenus Web]." Au-delà de son inscription dans des stratégies individuelles, nous avons constaté l'importance de la formation professionnelle à la langue française pour travailler dans l'externalisation, nos données quantitatives montrant un lien entre le niveau de langue et la participation à des formations professionnelles.

	Non	Oui ¹	Total
Avancé	56.8	43.2	100
Pas avancé	71.6	28.4	100
Ensemble	66.5	33.5	100

¹Réponses en pourcentage, Fisher.test = 0.03353

TABLE 12 – Niveau de français selon la formation professionnelle aux langues. Lecture : pour des raisons de taille de l'échantillon, nous avons regroupé les "intermédiaire" et les "débutant" dans une modalité "pas avancé"

Nos entretiens montrent bien que cela correspond à des stratégies conscientes de la part des travailleurs. Nous avons déjà évoqué le cas de l'Alliance française, mais il existe également d'autres centres à Madagascar, comme le centre culturel Albert Camus qui permet à des travailleurs de bénéficier de ressources pour l'apprentissage du français ³⁰⁰ tels que des livres, à un cinéma, ainsi qu'à des expositions. Mevo y développe sa maîtrise du français à travers le contact avec la culture française (livres, presse) et explique également avoir amélioré ses qualifications à travers une activité professionnelle de rédacteur au sein d'une grande entreprise d'annotation.

Encadré n°72 : L'apprentissage du français via un centre culturel et le travail en entreprise

X : sur le français, tu as dit que tu avais appris en autodidacte. Mais ça veut dire quoi ?

M : (...) on utilisait le français depuis les petites classes, ici à Mada. Alors, autodidacte, j'étais membre du centre de culture française, CCF(...)c'est le centre culturel Albert-Camus, mais ça, c'est changé

300. Il s'agit désormais de l'Institut Français de Madagascar qui "a été créé le 1er janvier 2011 par regroupement du Centre culturel Albert-Camus et d'une partie du service de coopération et d'action culturelle de l'ambassade. (Wikipédia, consulté le 28/08/2023)

en centre culturel français. J'aimais beaucoup lire, j'aimais beaucoup aussi écrire

X : par exemple ... c'est quoi que tu aimes comme roman ?

M : Non, c'est pas trop les romans. J'aime lire surtout les journaux et magazines (...) Tout ce qui est sport, l'Equipe, tout ce qui est magazine sportif (...) mais je n'étais pas ... je n'étais pas vraiment ce gars qui aime le Français, qui va lire les romans.

X : ouais, tu lisais la presse, enfin ouais, le sport ... ok

M : le Français pratique (...) dès qu'on a testé en tant rédacteur, à écrire, ils ont vu le potentiel « il écrit bien le gars »

X : oui, tu avais eu l'habitude de tout le temps lire aussi donc ...

M : Oui, et j'étais bien organisé, j'étais clair et cetera ... j'ai eu la chance aussi d'avoir un natif qui était chez Sous-traitant C, qui a lancé le plateau rédaction, qui m'a soutenu ... elle m'a inspiré en fait, que c'était ça, si on utilisait le français sous cet angle-là, c'était ça. Ça s'est fait tout seul

Mevo - Responsable - Sous-traitant C

Néanmoins, au-delà d'un accès à langue française de par l'environnement culturel de la capitale (présence de publicités en français, utilisation à la radio et centres culturels en langue française), l'accès à Internet constitue un élément décisif du fait d'une sur-représentation de la culture française sur l'Internet Malagasy³⁰¹. Cet élément a été souligné par plusieurs travailleurs, comme Sylvain, actuellement annotateur qui explique suivre "des séries *en français(...)* des films d'actions(...)" en français, ainsi que "le théâtre radiophonique en Français" principalement via Youtube.

	Non	Oui ¹	Total
Avancé	11.6	88.4	100
Intermédiaire	26.3	73.7	100
Débutant	36.5	63.5	100
Ensemble	23.9	76.1	100

¹Réponses en pourcentage, Fisher.test = 0.0005241

TABLE 13 – Niveau de français selon le fait de posséder un smartphone

Enfin, la présence de nombreux établissements d'enseignements primaires, secondaires et universitaires dont le français est la langue d'enseignement - certains travailleurs passés par des écoles privées expliquent

301. Sur ce point, nous reprenons l'argument de [Graham, Hogan, Straumann, et Medhat \(2014\)](#) qui parle d'une "pauvreté informationnelle" à propos des inégalités géographiques en matière d'accès à Internet. En utilisant des variables liées à la publication de contenus par les utilisateurs, il montre la surreprésentation d'informations concernant certains occidentaux qui concentrent également le plus d'internautes.

d'ailleurs qu'il leur était interdit de parler malgache, y compris en dehors des cours - constitue également un facteur d'apprentissage. S'il semble désormais clair que l'apprentissage du français fait partie des stratégies de "développement professionnel" des travailleurs, de quoi ce niveau de français est-il le synonyme? Nos données permettent de constater que ce niveau de français est atteignable pour les travailleurs qui sont passés à l'Université, car le passage par les études supérieures favorise significativement le fait de parler français.

	Supérieur Non	Supérieur Oui ¹	Total
Avancé	7.0	93.0	100
Intermédiaire	35.5	64.5	100
Débutant	31.6	68.4	100
Ensemble	24.4	75.6	100

¹Réponses en pourcentage, Fisher.test = 0.0000004277

TABLE 14 – Niveau de français selon le niveau d'études

Une autre explication nous est suggérée par le livre de Padios (2018) qui considère que : "Même si l'enseignement de l'anglais est dispensé à tous les Philippins dans le cadre de l'éducation publique, seuls ceux dont la position sociale et les ressources matérielles leur permettent d'entendre et de pratiquer la langue ont de bonnes chances de pouvoir la maîtriser au niveau utilisé dans les centres d'appel. Ce capital culturel découle de l'éducation formelle, mais il est également hérité des parents ou d'autres membres de la famille dont l'emploi et/ou l'éducation nécessitent des compétences linguistiques de haut niveau en anglais" (p.73). Sur ce point, nous constatons que le niveau d'étude des membres de la famille proche des annotateurs est très élevé. Ainsi, 64% des conjoints d'annotateurs sont passés par l'université. C'est 29% pour leurs pères et 21% pour les mères, soit un niveau d'éducation bien plus important que dans la population malgache ³⁰² (Table 50, Table 51)). Pour tester ces variables, nous avons réalisé une régression ordinale afin de tester le niveau de français selon plusieurs variables. Ces résultats suggèrent que toute chose égal par ailleurs, le fait d'être étudiant et de posséder un *smartphone* sont corrélés avec le niveau de français.

L'apprentissage du français révèle qu'il s'agit de stratégies délibérées des travailleurs, qui profitent simultanément aux entreprises de sous-traitance qui valorisent ces qualifications. Nous parlons alors d'un processus "d'assetisation coloniale", car la constitution de ces actifs coloniaux permet aux travailleurs de faciliter leur entrée sur le marché du travail en faisant valoir des formations et des qualifications linguistiques parfois très spécifiques, mobilisables pour obtenir des salaires plus importants. Les plus doués en français et en anglais ont davantage de chances d'occuper des positions managériales nécessitant des interactions avec les

302. Pour rappel, 10,5% des malgaches urbains et 3,1% des malgaches au total sont passés par l'université

	Estimate	Std. Error	Z value	Pr(>z)
Revenus foyer (500-750Kar)	0.23676	0.39906	0.593	0.55299
Revenus foyer (750-1325Kar)	0.04063	0.40152	0.101	0.91939
Revenus foyer (1325-1000000Kar)	-0.13776	0.43540	-0.316	0.75170
Salaire (400-439Kar)	0.46446	0.40232	1.154	0.24832
Salaire (439-540Kar)	0.41722	0.41078	1.016	0.30978
Salaire (540-2200Kar)	-0.51731	0.41055	-1.260	0.20766
Possède un smartphone	-0.88172	0.32878	-2.682	0.00732 ¹
Est étudiant	-1.00736	0.36204	-2.782	0.00540 ¹
Âge 25-34 ans	-0.11399	0.30048	-0.379	0.70443
Âge +35 ans	-0.90936	0.47870	-1.900	0.05748

¹Signif : 0.01

TABLE 15 – Modèle de régression ordinale pour tester le niveau de français selon plusieurs variables

clients finaux. Néanmoins, ce processus reflète également les difficultés des travailleurs à se projeter dans un métier et dans un pays qu'ils perçoivent comme *"sans avenir"*, autre que de partir. Partir où? Comment? Une partie de la réponse réside dans des stratégies d'apprentissage qui valorisent l'extérieur, l'intégration à la mondialisation et *in fine* la connexion aux pays occidentaux. Nous considérons que la valorisation de ces actifs est alors fortement limitée par la difficulté d'accéder à l'émigration en France et en Europe. Le secteur de l'externalisation constitue alors une des seules portes de sortie pour valoriser ces actifs. Les travailleurs finissent par *"migrer virtuellement"* pour reprendre les termes de (Aneesh, 2006) travaillant effectivement pour l'extérieur, mais depuis une situation géographique et professionnelle qu'ils considèrent comme sans avenir.

Cette dimension s'exprime également dans l'identité professionnelle des travailleurs qui est exprimée par ces derniers en fonction des caractéristiques du secteur : tourné vers l'étranger et portant sur des sujets liés à l'informatique. Symboliquement certains travailleurs valorisent le fait d'appartenir à un secteur perçu par certains comme moderne. Toutefois, il s'agit d'une *"identité contestée"*, car elle s'accompagne d'un *"flou professionnel"* vis-à-vis de la finalité du travail effectué et des perspectives d'avenir.

* * *

Dans ce chapitre, nous avons d'abord adopté une perspective *macro* sur la structure du secteur de l'externalisation de données et les modalités de son émergence. Nous partons d'une critique du concept de *data colonialism* (Couldry & Mejias, 2019a), comme étant une conception maladroite des mécanismes d'extraction de valeur qui se déroulent au sein des grandes plateformes numériques et plus particulièrement dans le secteur de l'intelligence artificielle. A la suite de (A. A. Casilli, 2017a) qui appelle à un *"tournant décolonial du numérique"*, nous pensons nécessaire de renverser ce concept pour analyser le secteur de l'IA dans

ses dimensions extractives et dans la *colonialité numérique* des travailleurs malgaches. En effet, l'extractivisme permet de rendre compte des mécanismes d'appropriation de valeurs économiques qui se mettent en place entre les pays du nord et les pays du sud tandis que la colonialité articule les mécanismes économiques et institutionnels qui sous-tendent le développement de l'industrie des données tout en tenant compte l'*agentivité* des travailleurs.

En analysant les politiques gouvernementales d'attraction des investisseurs, nous avons montré qu'elles ont largement favorisé l'établissement d'investisseurs français et mauriciens au détriment des investisseurs locaux ce qui illustre une première dimension extractive du secteur. Par ailleurs, bien que le régime de zones franches n'ait théoriquement pas entraîné de violations graves des lois locales du travail, mais ces aspects ne sont directement supervisés par les institutions publiques malgaches, notamment l'EDBM. Plutôt que de se concentrer sur l'extraction de valeur liée aux données personnelles des utilisateurs par les grandes entreprises numériques, nos résultats mettent en évidence l'intégration du travail de données au sein de réseaux de production post-coloniaux qui profitent du développement inégal entre les pays.

Nous avons également analysé les trajectoires des entrepreneurs de la sous-traitance et identifié deux catégories d'élites du secteur. Les premiers sont les *entrepreneurs du développement* et ont des liens avec Madagascar. Par conséquent, leur position est exceptionnelle dans le paysage de l'IA, car ils combinent deux atouts difficiles à obtenir simultanément : une connaissance approfondie de Madagascar du côté des travailleurs et une connaissance des pays occidentaux du côté des clients. Les "*entrepreneurs des marges du développement*" se trouvent dans les entreprises de sous-traitance et ont comme caractéristique principale d'opérer dans les zones en périphérie de la mondialisation. Ces deux groupes d'entrepreneurs affichent des attitudes très différentes à l'égard de l'accès aux facteurs de production et des objectifs liés à l'implantation de leurs activités à Madagascar. Les "*entrepreneurs de développement*" mettent davantage en avant une rationalité liée au développement de leurs pays, tandis que les "*entrepreneurs des marges du développement*" sont davantage motivés par la rationalité économique liée aux coûts de main-d'œuvre avantageux à Madagascar passant d'un pays à l'autre en fonction des opportunités économiques des pays les plus en marge de la mondialisation qui constituent la principale source de création de valeur.

Ces résultats suggèrent une forme spécifique d'intégration des entreprises de Madagascar dans la mondialisation. Dans notre introduction, nous avons fait référence au travail de [Kleibert et Mann \(2020\)](#), qui présente diverses formes de couplage. Le couplage structurel désigne l'externalisation des tâches peu qualifiées par des entreprises étrangères, principalement motivée par la considération des coûts de main-d'œuvre. Le couplage fonctionnel repose sur un rapport coût-capacité, avec l'entrée des entreprises étrangères dans les réseaux de production mondiaux. Enfin, le couplage indigène implique que les entreprises nationales se connectent aux marchés mondiaux et deviennent des acteurs majeurs qui organisent la production à l'échelle mondiale. En se basant sur la structure du secteur, nous pouvons identifier une forme d'intégration

fonctionnelle des entreprises situées à Madagascar. Nous proposons plutôt de qualifier cette intégration de "marginale" pour décrire un mécanisme qui se situe à la convergence de considérations économiques et *développementales* et qui semble être déterminant dans les stratégies d'externalisation. Au-delà de l'analyse de l'intégration des entreprises nationales dans la mondialisation, nos résultats suggèrent que, dans le secteur des services, *"le monde n'est pas plat"*. Bien au contraire, les entrepreneurs de la sous-traitance délocalisent leurs activités dans des pays en marge pour préserver leurs bénéfices. Outre la relation "coût/capacité", la localisation géographique de cette chaîne de production s'inscrit dans un processus de réintermédiation qui voit les pays les plus développés diversifier leurs destinations de sous-traitant en fonction de cette perception du développement chez les pays hôtes de ces investissements.

Analyser la position géographique de la chaîne, constitue un autre enjeu de ce chapitre, puisqu'elle détermine presque exclusivement, ou presque, la structure de la captation de valeur au sein de la chaîne. C'est finalement parce que ces travailleurs malgaches sont malgaches qu'ils toucheront un salaire moins élevé. Même si ces derniers possèdent des qualifications, ceux qui n'ont pas eu l'opportunité de poursuivre leurs études en France ou de travailler à l'étranger demeurent exclus de la création d'entreprises dans les zones franches, qui semble être réservée aux entrepreneurs français. Ces constats suggèrent une forme de capitalisme qui s'appuie sur des distinctions raciales et qui favorise les pays situés en périphérie de la mondialisation contemporaine.

Nous avons également analysé la tension entre la reconnaissance et la dévalorisation des compétences par les entrepreneurs de la sous-traitance. Nos résultats révèlent que les entreprises de sous-traitance jouent un rôle déterminant dans la configuration des compétences des travailleurs, les orientant parfois vers des compétences que nous qualifierions de *"coloniales"*. Les travailleurs sont incités à développer des stratégies d'acquisition des compétences requises pour l'industrie de l'externalisation des données. Nous avons constaté que les stratégies des travailleurs pour apprendre le français ont comme finalité principale d'améliorer leurs conditions de vie à travers deux perspectives : l'émigration et l'entrée dans des secteurs tournés vers l'extérieur comme l'industrie de l'externalisation. Nous décrivons ce processus comme relevant de *"assetisation coloniale"*. La constitution de ces atouts coloniaux facilite l'insertion des travailleurs sur le marché du travail qui mettent en avant leurs formations et qualifications linguistiques, ce qui améliore leurs perspectives professionnelles.

Néanmoins, le contexte actuel de cette chaîne de valeur, mélangeant les vestiges de l'ancienne colonisation avec les nouveaux acteurs français du secteur des services, modèlent l'acquisition de compétences et l'accès au marché du travail pour les jeunes travailleurs malgaches à Antananarivo. Malgré les discours de certains entrepreneurs de la sous-traitance ou de l'IA, il est clair qu'ils accordent de la valeur aux compétences linguistiques et culturelles héritées de la colonisation. La capacité des travailleurs malgaches à comprendre les subtilités de la langue française, ainsi que les aspects culturels français, avec un niveau suffisant pour

discuter des éventuelles ambiguïtés dans les tâches d'annotation, joue un rôle crucial dans la pérennité du secteur dans le pays. Ces entreprises exercent donc une influence sur les identités professionnelles des travailleurs et leur perception de la dimension coloniale de leur travail. Malgré les difficultés liées à l'acquisition de ces compétences qualifiées de "coloniales" et en dépit du niveau d'éducation attesté par ces qualifications, l'appréciation sociale du travail de traitement des données par les entrepreneurs oscille entre le mythe du travail facile (Maruani & Nicole-Drancourt, 1989), d'une part, et la reconnaissance de leurs contributions, d'autre part, en fonction des interlocuteurs avec lesquels les entrepreneurs discutent de ces compétences.

Comment expliquer cette apparente contradiction? Tout d'abord, il est primordial de noter que la reconnaissance de l'importance de la "neutralité de l'accent" par les entreprises de sous-traitance s'inscrit dans un processus plus vaste de "rappel colonial". Padios (2018) évoque la notion de "rhétorique de la fiabilité" en ce qui concerne la mise en avant des compétences héritées de la période coloniale. Cette rhétorique de la fiabilité constitue une stratégie par laquelle les entreprises de main-d'œuvre essentialisent l'identité philippine pour obtenir un avantage concurrentiel sur le marché mondial, dans le contexte d'une quête d'investissements nationaux et d'emplois décents. Cela signifie que le rappel de la colonialité est utilisé de manière instrumentale pour les besoins d'entreprises internationales. Cette "rhétorique des qualifications coloniales" implique que les entreprises de sous-traitance mettent l'accent sur la complexité et la variété des compétences requises dans les services interactifs, tout en essentialisant l'identité des travailleurs malgaches en fonction de leurs compétences coloniales pour les tâches d'annotation. Nous proposons de considérer cela comme un "paravent des qualifications coloniales". En plus de la fiabilité dans la prestation de services informatiques, qui implique des compétences considérables en soi, les travailleurs possèdent d'autres qualifications dans divers domaines qui sont néanmoins négligées ou mises en avant par les entrepreneurs de la sous-traitance et de l'IA en fonction des besoins spécifiques des clients ou des projets.

Ces processus d'assetisation coloniale et de rappel des qualifications coloniales révèlent donc les obstacles auxquels les travailleurs sont confrontés lorsqu'ils envisagent un avenir professionnel dans un domaine et un pays qu'ils perçoivent comme ne présentant que peu de perspectives, sauf à partir. Mais partir vers où et comment? Une partie de la solution réside dans les stratégies d'apprentissages qui mettent l'accent sur l'extérieur, l'intégration à la mondialisation et les liens avec les pays occidentaux. Toutefois, la valorisation de ces "actifs coloniaux" est fortement limitée par le contexte colonial dans lequel ils sont acquis. Les difficultés d'émigrer vers la France et l'Europe, ainsi que le secteur de l'externalisation, qui apparaît comme la seule option pour valoriser ces compétences, restreignent considérablement les perspectives des travailleurs. Ainsi, pour reprendre les mots de (Handley, 2015), les facteurs de cet *extractivisme des données* et de cette "colonialité numérique" sont multiples "le racisme et le mercantilisme de la politique économique coloniale", des réglementations gouvernementales qui "favorisaient systématiquement les entreprises (...) de la métropole, limitant ainsi les possibilités offertes aux entrepreneurs indigènes"

Chapitre 9 - Vers le développement d'une *global precarity chain* : l'industrie de l'IA vectrice d'une *intégration laborieuse* de la jeunesse urbaine malgache dans le marché du travail

La littérature sur le micro-travail s'est principalement saisie de la question du développement à travers deux enjeux principaux : les effets de cette activité sur les travailleurs et les facteurs d'attractivité des pays pour ce type d'activité. Au sein des travaux académiques et de politiques publiques qui se sont développés à propos de l'impact des plateformes dans les pays du Sud global, on peut identifier deux catégories de travaux : les études qui considèrent cette activité comme une opportunité et qui analysent les freins à son émergence dans les pays du Sud global et les travaux qui analysent l'impact sur les travailleurs. Cette littérature souffre toutefois de plusieurs limites. Tout d'abord, les études qui portent sur l'attractivité des pays en développement abordent peu la dimension post-coloniale. Par ailleurs, ces travaux souffrent en partie de leur portée normative. En effet, ils répondent à un enjeu pratique - de proposer une solution pour le développement - mais ne prennent pas assez en compte la subjectivité des travailleurs. Du côté de la littérature plus critique, les travaux de (Graham et al., 2017a) ou de Berg et al. (2018) présentent l'intérêt d'analyser les conditions de vie et de travail des travailleurs des plateformes avec une focalisation sur le rapport des travailleurs à leur activité.

Nous avons déjà abordé la *solidification de ce groupe professionnel* dans notre précédente partie. Une question demeure toutefois quant à la valorisation de ces compétences par les individus, les organisations et les institutions. La formation d'un groupe professionnel permet aux individus d'exister sur le marché de l'emploi de à travers les compétences précédemment acquises, valorisables dans d'autres emplois. Une question sous-jacente réside dans la capacité des acteurs à évoluer professionnellement : à quelles perspectives d'évolutions cette expérience professionnelle permet-elle d'accéder ?

Du point de vue des carrières professionnelles, on peut donc se demander si cette activité perçue comme répétitive et transitoire est réellement une source de développement professionnelle pour les travailleurs ? Ayomikun Idowu et Amany Elbanna (Idowu et Elbanna (2019b)) explorent la manière dont les travailleurs du clic Nigériens vivent leur travail sur les plateformes afin de déterminer s'ils peuvent "*développer leurs compétences et organiser leur avenir*". Ils identifient une typologie des de ces travailleurs en fonction de leur statut d'emploi précédent et de leur "*familiarité avec le travail des plateformes*". Les "*switchers*" sont des travailleurs qui étaient auparavant dans un emploi formel et qui cherchent un revenu complémentaire. Les "*awakened crowdworkers*" sont des individus auparavant engagés dans un emploi plus ou moins formel à temps plein et qui essayent de fonder leur propre affaire via les plateformes. Les "*early birds*" sont des travailleurs qui ont commencé à avoir une activité sur une plateforme lorsqu'ils étaient encore à l'université.

Les "green starters" étaient auparavant au chômage et se sont lancés dans cette activité pour pallier leur manque de revenus. Ils identifient quatre étapes de carrière dans le travail des plateformes qui se distinguent par les activités et les compétences : les "travailleurs non-fiables" essaient de construire une réputation et un réseau professionnel, les "travailleurs de confiance" tentent d'améliorer leurs compétences et de diversifier leurs activités, les "vétérans" ont un niveau suffisant de compétences et de diversification pour organiser et gérer une équipe de travailleurs³⁰³. Enfin, les "travailleurs qui quittent ce travail", que ce soit pour "suivre leur passion de toujours" après avoir gagné suffisamment d'argent ("dream chasers") ou pour développer une activité entrepreneuriale déjà en cours pendant leur période d'activité sur les plateformes.

Dans ce chapitre, nous analysons l'articulation entre les contraintes du secteur et l'inscription du travail d'annotateur dans des stratégies de carrière professionnelle. Nous pensons que cette dimension du travail des données peut s'analyser sous l'angle d'une tension entre "stabilisation professionnelle" et "précarité professionnelle". Nous pensons que cette tension est d'autant plus déterminante pour le développement professionnel des travailleurs qu'elle s'inscrit dans un contexte local perturbé par le manque de perspectives économiques et professionnelles et par la captation de la valeur produite par les autres acteurs de la chaîne. Un autre élément de notre étude porte sur la notion de précarité définie dans la littérature comme "le fait d'être vulnérable à l'exploitation en raison d'un manque de sécurité. La précarité suggère le potentiel d'exploitation et d'abus, mais pas sa présence certaine." Banki (2013). De nombreuses études suggèrent l'émergence d'un précaire comme une forme "hégémonique de la relation salariale moderne" (Castel, 2007) qui diffère dans la forme selon l'"intégration" des travailleurs (Paugam et Vendramin, 2022)³⁰⁴. Dans son examen de l'émergence de nouvelles formes de travail virtuel, Huws (2014) réutilise le concept de précarité en considérant que le concept "lie (...) trois choses : le besoin de subsistance des travailleurs, leur travail et la plus-value expropriée des résultats de ce travail, sans laquelle le capital ne peut être accumulé ni le capitalisme perpétué." L'intérêt de cette conceptualisation est de s'inscrire plus directement dans un questionnement sur les stratégies utilisées par les travailleurs des données pour assurer leur subsistance et non uniquement sur le potentiel d'abus et d'exploitation des emplois précaires. D'après la chercheuse, cette analyse pose des questions fondamentales sur ce qu'est-ce que la subsistance ? En d'autres termes, qu'est-ce que les travailleurs tirent de cette activité ?

303. "Certains crowdworkers à ce stade ont révélé qu'ils travaillent moins sur les tâches en tirant parti de leur grande réputation sur la plateforme pour obtenir autant de travaux que possible et ensuite servir d'intermédiaire en les externalisant à des travailleurs sur et hors de la plateforme. Ils sont davantage engagés dans des tâches administratives qui impliquent le recrutement, la gestion, l'organisation et le suivi de plusieurs tâches, travailleurs et employeurs en même temps" Idowu et Elbanna (2019b)

304. Serge Paugam et Patricia Vendramin (2022) analysent la question de la précarité sous l'angle de l'instabilité de l'emploi (déficit de protection) et de l'insatisfaction au travail (déni de reconnaissance). Ils vérifient ainsi empiriquement quatre formes d'intégration professionnelle : "l'intégration assurée" (stabilité de l'emploi et satisfaction au travail), "l'intégration incertaine" (instabilité de l'emploi et satisfaction au travail), "l'intégration laborieuse" (stabilité de l'emploi et insatisfaction au travail) et, enfin, "l'intégration disqualifiante" (instabilité de l'emploi et insatisfaction au travail). Ainsi, la précarité de l'emploi peut se superposer à la précarité du travail, mais cette superposition n'est pas systématique.

Ce point nous permet d'aborder à nouveau une critique fondamentale du travail de [Couldry et Mejias \(2019a\)](#), qui selon nous oblitère la structure du marché et la relation contractuelle qui s'établit entre d'un côté les travailleurs - consommateurs qui utilisent Internet et de l'autre les travailleurs qui sont employés à Madagascar pour produire des données. Au-delà de la question de la subsistance, il est nécessaire de tenir compte du caractère "extractif" d'un secteur au sein duquel les travailleurs de Madagascar sont dans une position de "migrant virtuel" de par leur appartenance à un secteur tourné vers l'extérieur ([Aneesh, 2015](#)). Il faut pour cela analyser l'individu dans son rapport au travail dans ce qui constitue un prolongement et une focalisation sur un aspect précis de la socialisation secondaire : la carrière. Dans notre cas, la notion de carrière renvoie à tout ce qui touche à l'univers professionnel et notamment à la transition étude-emploi, et encore davantage dans le contexte de notre population qui est particulièrement jeune. Comment étudier la place de cette activité dans la vie et le développement professionnel des annotateurs ? Qu'est-ce que cette analyse explique des effets à long terme sur le développement de Madagascar ?

En lien avec cet enjeu, ([Kano et al., 2020](#)) identifient la question de la distribution de la valeur au sein de la chaîne de production comme étant une autre lacune de la littérature ³⁰⁵. Sur ces questions ils opposent deux types de littératures scientifiques, le premier type portant uniquement sur les firmes focales (*business studies*) et le second analysant les relations entre les firmes focales et leurs sous-traitants sous l'angle des asymétries de pouvoir (sociologie). Les auteurs indiquent toutefois que peu de données et études empiriques permettent d'identifier de manière claire les mécanismes de distribution de valeur. Ils soulignent d'ailleurs la difficulté qu'ont certaines grandes entreprises transnationales à savoir exactement comment la valeur créée est distribuée au sein de leur réseau de sous-traitant ³⁰⁶ ([Seppala et al., 2014](#)). C'est également le sens de la proposition théorique de Benjamin [Selwyn \(2015, 2019\)](#) qui propose d'inverser le concept des chaînes globales de valeur en partant plutôt du constat des effets de la chaîne sur les habitants des pays du Sud global, considérant que le fait de placer la focale sur la "valeur" obscurcit les acteurs les plus impactés par les chaînes de valeur. Il parle ainsi de "global poverty chain" pour désigner une forme d'intensification de l'exploitation opérée par les grandes entreprises multinationales sur les travailleurs. Selon lui, l'existence de chaînes de valeur plus ou moins formelles rend également cette tâche plus difficile à accomplir et ce d'autant plus lorsqu'un sous-traitant d'une grande entreprise sous-traite lui-même le travail confié par le donneur d'ordres.

Nous avons identifié les stratégies des travailleurs les menant à rejoindre le secteur de l'externalisation. Dès lors, comment analyser ce premier résultat au regard des enjeux de développement ? Comment concilier

305. Ainsi "Bien que la question de la valeur dans une chaîne de valeur mondiale ait fait l'objet d'une attention particulière de la part des chercheurs, la question de savoir comment les entreprises chefs de file devraient coordonner la création, la capture et la distribution de la valeur n'a pas encore été résolue."

306. On peut rapprocher cet argument des recherches portant sur l'IA comme boîte noire et notamment le constat de [Burrell \(2016\)](#) sur le processus de développement de l'IA lui-même fragmenté entre différents services au sein des entreprises du numérique ce qui impact la compréhension des mécanismes d'accaparement de valeur

stratégies individuelles et effets structureaux du secteur en matière de développement ? Nous pensons qu'il est nécessaire de repartir d'une analyse de la composition sociale des travailleurs des données afin d'analyser la dimension "extractiviste" du secteur des données. En effet, le concept "d'extractivisme des données" se distingue du "colonialisme des données" de par le principe d'altération de la source de valeur. Or, si la valeur réside dans les qualifications des travailleurs des données, comment et dans quelles mesures le secteur altère-t-il ces qualifications ? Nous proposons d'étudier ce point en faisant l'hypothèse d'un secteur qui permet une "stabilisation" relative des travailleurs, stabilisation qui cache cependant un développement professionnel contrarié par le manque de perspectives professionnelles. *In fine*, nous considérons que cette industrie est "extractive" de par sa captation d'une jeunesse issue de la classe moyenne et diplômée dont les compétences profitent au secteur de l'externalisation et non à Madagascar ou aux travailleurs eux-mêmes. Il convient pour cela de se demander qui annote des données à Madagascar et dans quelles trajectoires professionnelles s'inscrit la participation à ce secteur ? L'enjeu de l'apport d'un secteur au développement doit également se saisir des effets sur les travailleurs du secteur. Ainsi, dans la continuité de la proposition théorique de Benjamin Selwyn (2015, 2019), nous proposons d'inverser le concept des chaînes globales de valeur en partant plutôt de l'analyse des effets de la chaîne sur les habitants des pays en voie de développement. Finalement, dans quelles mesures faudrait-il parler d'une "chaîne globale de précarité" plutôt que d'une chaîne globale de valeur ?

Si la question de la valeur demeure pertinente comme axe d'analyse de la chaîne, elle n'en est qu'un des aspects. Surtout son analyse doit s'inscrire à la fois dans une meilleure compréhension du partage de la valeur, et plus encore dans une contextualisation de ce niveau de partage au regard du "coût de la reproduction sociale" des travailleurs (Selwyn, 2019). Il nous faut par conséquent analyser les conditions de vie des travailleurs au regard de ce qu'ils tirent de leur activité professionnelle en se demandant : à quoi correspondent les revenus de cette activité pour les travailleurs ? Par ailleurs, nous pensons qu'il est nécessaire de contextualiser ce niveau de revenu à l'échelle de la production de valeur au sein de la chaîne en se demandant : comment est partagée la valeur au sein de l'industrie de l'IA ?

9.1 Le secteur de l'externalisation comme unique piste d'entrée sur le marché du travail pour la jeunesse diplômée d'Antananarivo ?

Lors d'une visite dans Antananarivo, nous rencontrons une cheffe cuisinière malgache très connue comme présentatrice d'émissions télévisées portant sur la cuisine *malagasy*. Au détour de la conversation, elle nous raconte qu'elle s'est retrouvée à réaliser des tâches de rédaction de contenus sur Internet pour gagner de l'argent alors que son restaurant était fermé à cause de la crise sanitaire liée au Covid19. La cheffe est pourtant très bien installée dans sa vie professionnelle, puisqu'elle gère depuis trente ans un restaurant italien sur les hauteurs de la capitale. Cet exemple pourrait suggérer que beaucoup de personnes différentes sont

concernées par le secteur de l'externalisation et donc de l'annotation de données. Cela soulève également la question des trajectoires professionnelles dans lesquelles s'inscrit le travail dans le secteur.

Dans cette section, nous analysons deux questions : qui sont les malgaches qui rejoignent ce secteur ? Quelles sont les perspectives du secteur en matière de développement professionnel ? Sous cet angle, nous proposons de compléter la littérature sur le micro-travail qui, à de nombreuses reprises, a tenté de documenter la composition de la force de travail. Notre analyse s'inscrit quant à elle dans des questionnements liés au cadre théorique de "l'extractivisme de données" en considérant les travailleurs et leurs qualifications comme une source de valeur altérée par le secteur de l'IA. L'analyse des trajectoires professionnelles s'inscrit de son côté dans une étude davantage centrée sur les individus et sur ce que représente l'annotation dans leur trajectoire professionnelle.

Notre travail contribue également à explorer d'autres approches pour la compréhension de l'émergence de classes moyennes dans les pays du Sud global et notamment sur le continent africain. Il s'agit ici repartir d'une tension mise en lumière par la littérature à propos du rôle des classes moyennes dans le développement (Banerjee & Duflo, 2008). Ainsi, des recherches soulignent que la littérature "n'a pas pu identifier de preuves suffisantes que les classes moyennes africaines favorisent et contribuent "automatiquement" à la croissance économique, comme les modèles normatifs tendent à l'affirmer" (Kroecker, O'Kane, & Scharrer, 2018; Melber, 2022). L'importation du concept de classe sociale hors du contexte occidental est également critiqué, car source de flou conceptuel au regard "réductionnisme économique" dont font preuve les recherches sur les classes moyennes basées sur le salaire. Nos recherches permettent alors d'explorer un autre contexte de formation de la classe moyenne, hors d'une conception occidentale qui implique traditionnellement le passage par une économie industrielle³⁰⁷. Notre première sous-section porte donc sur la composition sociale, économique et culturelle de la force de travail employée par les entreprises d'annotation de données. Nous nous inspirons pour cela du travail de Savage et al. (2013), qui proposent une analyse de la population britannique par "classes sociales" en se basant sur plusieurs indicateurs économiques, sociaux et culturels. Puis, nous nous basons sur les approches développées en sociologie autour de l'analyse de séquences professionnelles (Robette & Thibault, 2008) afin d'étudier ce que représente le fait d'être annotateur dans le parcours professionnel de ces derniers.

9.1.1 Une industrie qui repose largement sur la jeunesse diplômée de la classe moyenne

Pour commencer, revenons sur quelques caractéristiques socio-démographiques des annotateurs de données, que nous comparons avec des résultats d'enquête en population générale. Ainsi, sur les 25 674 196

307. "second point is that African countries are unlikely to develop a strong manufacturing sector. This means that they might go directly from an economic situation that relies on agriculture and natural resources to a post-industrial economic order marked by the predominance of services, trade, and the knowledge industry in the national economy". (Kroecker et al., 2018) (p.16)

millions d'habitants que compte Madagascar, il y a 50,7% de femmes, contre 33,3% dans notre échantillon, ce qui pourrait s'expliquer par de fortes difficultés de la population féminine à rentrer dans le marché du travail ³⁰⁸. Concernant le genre, notre population d'annotateurs est donc principalement composée d'hommes (66,7%), ce qui est cohérent avec ce qui est observé dans la littérature étudiant les micro-travailleurs de pays en voie de développement.

Au sujet de l'âge, la population malgache est âgée de 22,4 ans en moyenne avec une médiane à 18 ans. Ainsi, comme dans d'autres pays du continent africain, la population est très jeune : 41,3% de la population a moins de 15 ans et 17,1% de personnes ont entre 18 et 24 ans et 13,81% entre 25 et 34 ans (p.79). Concernant l'âge des annotateurs, il convient de préciser que la population que nous étudions est très majoritairement présente dans des entreprises formelles dans lesquelles le travail des enfants est interdit, ce qui explique au moins partiellement que notre population soit composée principalement de personnes d'au moins 18 ans. Les annotateurs concernés par notre étude ont ainsi majoritairement entre 18-24 ans (26,9%) et entre 25-34 ans (59,9%). Par ailleurs, l'INSTAT (2020) explique que *"la population au chômage est dominée par les jeunes puisqu'en effet 68,5% des chômeurs ont entre 15-24 ans et ils sont au nombre de 295 229 jeunes. La proportion atteint 84,6% si l'on considère les jeunes entre 15-29 ans."*

La situation matrimoniale des annotateurs de données présente des caractéristiques similaires à la population générale avec 45,2% de personnes mariées (et 12,9% en couple) et 36,7% de célibataires. Dans notre échantillon, les célibataires se trouvent principalement parmi les plus jeunes, parmi ceux qui sont encore étudiants et parmi ceux qui habitent avec leurs parents. Ainsi, en population générale, Madagascar compte 38,4% de célibataires et 53,2% de personnes mariées (p.124). Par ailleurs, 80,7% des personnes vivent en milieu rural. La région de la capitale (Analamanga) dans laquelle presque tous les annotateurs interrogés habitent et travaillent, est la plus urbanisée (37,8%).

Si l'on s'attarde sur le travail des conjoints, un point assez remarquable est que 27,1% des conjoints travaillent dans des activités informatiques (Table 44). Nos entretiens confirment d'ailleurs que certains travailleurs se retrouvent régulièrement à exercer dans les mêmes entreprises que leurs partenaires. C'est par exemple le cas de Hejo, annotatrice au sein du sous-traitant A qui explique que son *"mari travaille encore chez [Sous-traitant E]"*, ou de Fiary qui raconte avoir été recruté chez le sous-traitant B après son *"fiancé"*, *qui était recruté en premier, en juillet, en CDI, et moi en Novembre en CDI, dans le même projet"*. Au sujet de la taille du foyer, nous obtenons une médiane et une moyenne à 4 personnes, très similaire à ce qui observé sur l'ensemble de la population urbaine (3,9). Ce sont principalement les conjoints (45%), les parents (39%)

308. D'après l'enquête de l'INSTAT (2020), les résultats *"mettent en lumière également la contribution relativement faible ou l'accès relativement difficile des femmes dans les activités salariales, notamment en milieu urbain. Si plus de 38,4% des hommes sont des salariés, la proportion n'est que de 31,8% chez les femmes. Plusieurs facteurs pourraient être à l'origine de cette situation : contraintes sur le temps disponible chez les femmes pour exercer des activités en dehors du foyer, insécurité, partage des tâches au sein du ménage, différentes sortes de discrimination à l'embauche sur le marché du travail."*

Variable	N = 287
Taille du foyer	4.00, 4.00 (1 - 29)
Conjoint	
Non	154 (55,4%)
Oui	124 (44,6%)
Parents	
Non	169 (60,8%)
Oui	109 (39,2%)
Parents du conjoint	
Non	262 (94,2%)
Oui	16 (5,8%)
Enfants	
Non	164 (59%)
Oui	114 (41%)
Autres familles	
Non	220 (79,1%)
Oui	58 (20,9%)

¹ Mean, Median (IQR); n (%)

TABLE 16 – Composition du foyer

et les enfants (41%) (Table 16) qui habitent avec les annotateurs. Il n'est pas étonnant qu'une proportion non négligeable de répondants habitent avec leurs parents, notamment parmi les plus jeunes travailleurs qui sont souvent célibataires. En effet, comme dans d'autres pays en voie de développement, Madagascar se caractérise par un départ "tardif" du foyer qui intervient la plupart du temps après le mariage ou la mise en couple. Avant de revenir plus en détail sur la construction de notre typologie, continuons à analyser les résultats des variables qui n'ont pas encore été étudiées dans cette sous-section jusqu'ici : le salaire, les revenus du foyer et la position sociale. Concernant le salaire, il s'agit de souligner la présence de plusieurs individus considérés comme des "outliers" (en rouge sur le graphique) aux revenus est beaucoup plus importants que ceux des autres annotateurs. Ces résultats correspondent à la présence de quelques travailleurs qui occupent des postes de management. Comme on le verra plus tard dans ce chapitre, les revenus liés au salaire diffèrent selon l'entreprise, ainsi que selon la position hiérarchique dans l'entreprise. Ces deux facteurs expliquent les différences entre les 25% d'annotateurs les mieux payés (entre 108 et 454 euros) et les 25% les moins bien payés (25 et 82 euros par mois). Ces fortes différences sont également présentes au niveau des revenus globaux des foyers des annotateurs et s'expliquent cette fois par la composition du foyer (Table 16).

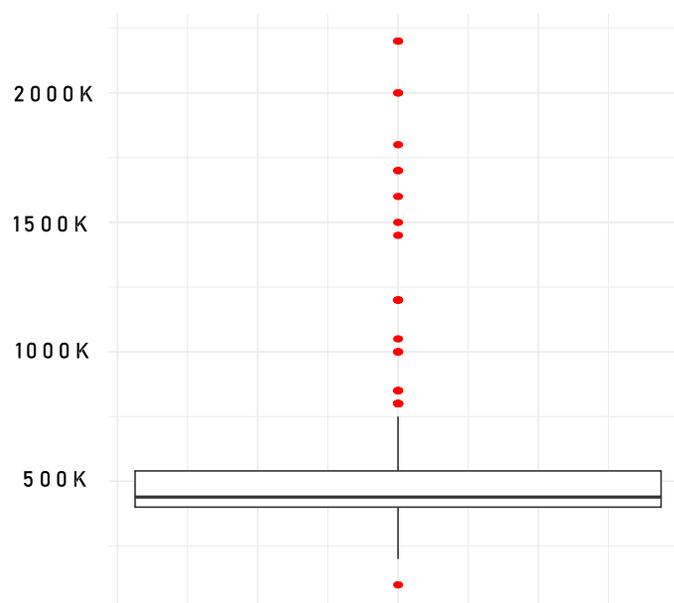


FIGURE 38 – Distribution des salaires des annotateurs de données. Lecture : les valeurs sont exprimées en ariary (1 euros = 5000 ariary) et les points rouges sont des "outliers" du fait de leur niveau de salaire.

En ce qui concerne la position sociale, nous attribuons des scores aux professions sélectionnées par les annotateurs dans le générateur de position. Il est alors possible d'extraire plusieurs métriques qui servent à caractériser la position sociale de notre échantillon (Li & Verhaeghe, 2015). Le graphique ci-dessous illustre la distribution des scores moyens de position sociale des annotateurs. On constate que cette distribution suit quasiment une loi normale. On observe une forte concentration des annotateurs autour des scores "moyens". Ainsi, la moyenne des scores se situe à 53 avec un écart-type à 7, ce qui indique que les annotateurs se situent pour la majorité d'entre eux autour de la moyenne : 50% des annotateurs ont un score entre 49,5 et 56,1. La littérature encore très parcellaire sur l'émergence de la classe moyenne à Madagascar retient le chiffre de 7,8% d'individus à 9% en population générale (Ncube et al., 2011; Razafindrakoto et al., 2015), dont 62% d'entre eux qui correspondent à une "classe flottante" d'individus gagnant entre 2 et 4 dollars par jour, ainsi qu'à des individus dont le revenu oscille autour de 2 dollars par jour et qui forment ce que la littérature sur les pays en voie de développement à dominante agricole appelle la "petite prospérité"³⁰⁹. Ce concept analytique sur la classe moyenne décrit des individus qui sortent de la précarité grâce à des activités informelles liées au monde agricole et "dont les moyens d'existence leur permettent d'enclencher une dynamique d'accumulation et donc d'être parmi les acteurs principaux des changements en cours en milieu rural". Si l'on retient l'analyse par le salaire, 92,5% de notre échantillon appartiendrait à cette classe moyenne gagnant au moins deux dollars

309. Ce "Issu de la notion confucéenne xiaokang, décrivant la situation des paysans moyens dans la Chine antique, ce concept a été proposé par Darbon et Toulabor (2011) pour saisir l'intuition derrière l'idée de classe moyenne en Afrique subsaharienne et désigne la sortie de la précarité (Darbon, 2012)" (Andrianampiarivo, 2018)

par jour (55,8 euros mensuel). Nos résultats suggèrent une très forte présence de la classe moyenne quelque soit la variable utilisée (économique ou sociale). Comment dès lors adapter le concept de "petite prospérité" à cette population? Faut-il y voir une dynamique d'émergence d'une classe moyenne urbaine (ou d'une "petite prospérité urbaine"), encore peu étudiée par ailleurs dans la littérature sur Madagascar (Bonnefond & Andrianampiarivo, 2020)?

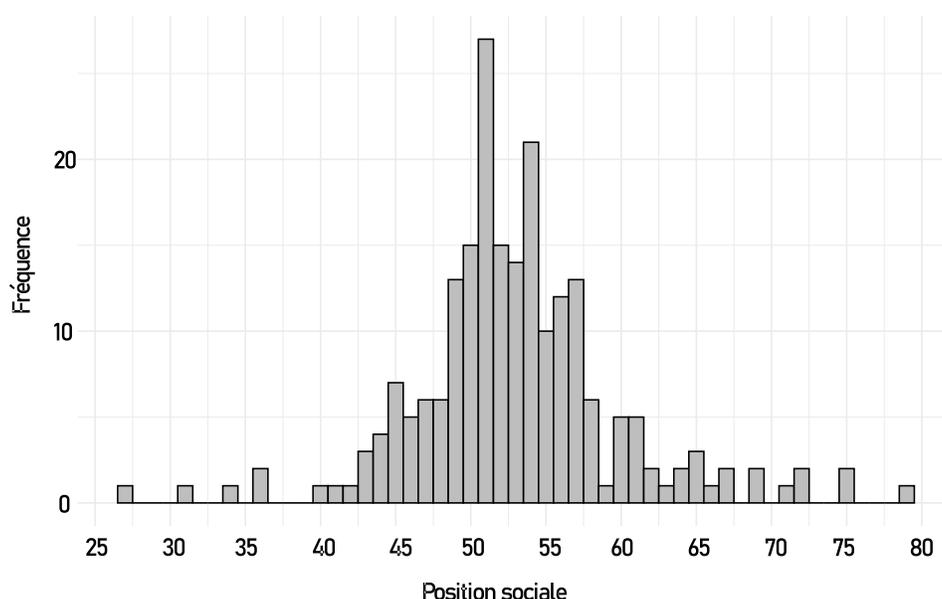


FIGURE 39 – Distribution de la position sociale des annotateurs de données. Lecture : les valeurs correspondent à la moyenne des scores associés aux professions indiquées dans le générateur de position sociale (voir la méthodologie)

Le travail de (Bonnefond & Andrianampiarivo, 2020) nous donne des indications quant à la segmentation de cette petite prospérité dans le monde urbain. Ainsi, les autrices et auteurs considèrent qu'il est possible de segmenter cette classe en deux catégories : "la classe moyenne urbaine malgache appartenant à la strate haute de la catégorie aux cadres et aux quelques employés qualifiés des secteurs public et privé formel ainsi qu'à quelques indépendants entrepreneurs à revenus élevés (...)" tandis que "Les petites prospérités, constituant la strate basse, la plus conséquente de la classe moyenne urbaine malgache, peuvent être associées aux employés non qualifiés et aux quelques ouvriers des secteurs public et privé formel ainsi qu'à une multitude de petits entrepreneurs et aux quelques employés à revenus réguliers du secteur informel."

Notre travail offre une autre perspective sur l'émergence de cette classe moyenne par l'utilisation de la position sociale ³¹⁰. De ce point de vue, nos résultats illustrent une forte hétérogénéité des ressources sociales

310. Pour rappel, cet indicateur est utilisé comme métrique d'accès à des ressources professionnelles susceptibles de permettre

dans notre échantillon : de nombreux annotateurs identifient à la fois des positions sociales dont les scores sont faibles et des professions dont les scores sont très élevés. Il est néanmoins possible d'analyser ce résultat comme le témoin d'un développement d'une "classe moyenne en transition" composée d'une "classe moyenne urbaine" et d'une "petite prospérité urbaine". Les premiers sont les travailleurs les mieux dotés socialement et économiquement qui travaillent et vivent à Antananarivo. Les autres se caractérisent par un score de position moins élevé et qui demeurent à cheval entre un réseau lié au monde rural (profession liée à l'agriculture) et au monde urbain (profession liée au secteur des services), entre le réseau familial, parfois resté dans les autres régions de Madagascar, et le réseau lié à leurs études et à leur métier dans le secteur de l'externalisation.

Variable	Moyenne ¹	Écart-type	Intervalle
Volume des ressources sociales	9,54	5,6	0-22
Prestige professionnel moyen des ressources du réseau	53	7,01	0-78,57
Prestige professionnel le plus élevé des ressources du réseau	78,6	8,8	26,7 - 78,6
Prestige professionnel le moins élevé des ressources du réseau	31,8	9,7	26,7 - 78,6
Différence de prestige professionnel des ressources du réseau	44,9	10,35	8,38 - 51,9

¹N = 215

TABLE 17 – Métriques du générateur de position

Pour compléter cette première description de notre population, nous nous basons sur le travail de Savage et al. (2013) afin de vérifier s'il est possible de distinguer plusieurs catégories de travailleurs. L'enjeu de ce travail est de classer les travailleurs en fonction de leurs caractéristiques afin de proposer une typologie. Les questions subséquentes de cette première interrogation consistent à se demander les "propriétés sociales" qui expliquent que les travailleurs de Madagascar finissent par travailler dans le secteur de l'annotation de données tout autant que les caractéristiques qui facilitent l'embauche dans cette industrie qui reste minoritaire dans le pays : 20,1% des actifs sont dans le secteur tertiaire dans un pays toujours largement agricole (74,7% des actifs). Si l'on se focalise sur la population urbaine, le rapport est toutefois inversé puisque "moins d'un quart (24,2%) des travailleurs se trouvent dans le Secteur Primaire et près des trois quarts (65,1%) de la main d'œuvre est concentrée dans les activités du commerce ou service" (INSTAT, 2021).

Notre travail se base sur les mêmes méthodes qu'un article déjà publié (Cornet et al., 2022) et reprend une approche développée par Savage et al. (2013) qui s'inspire de l'analyse bourdieusienne des classes sociales, par le positionnement des individus dans le champ social. Nous utilisons une analyse en correspondance multiple, classique en sociologie et qui s'est peu à peu popularisée du fait de sa puissance en matière de

à des travailleurs de mieux résister aux aléas de la vie en facilitant leur recherche d'emploi

”synthèse de données”, permettant si elle est bien maniée d’identifier certains phénomènes [Benzécri \(1980\)](#). Ceci étant dit, la littérature a également mis en avant certaines faiblesses à son emploi et dans ses capacités interprétatives. Ainsi, [Blanchard et Patou \(2003\)](#) soulignent le risque que des *”aspects décisifs de la problématique peuvent être absents des premiers axes, masqués par des dimensions plus massives mais triviales ou déjà mises en évidence par ailleurs”*. Ils identifient plusieurs méthodes statistiques complémentaires et notamment la classification hiérarchique que nous mobilisons également dans notre analyse et qui permet de ”projeter” des classes *”sous formes de nuages sur les cartes factorielles”*. Nous proposons une autre piste qui consiste ultérieurement à s’appuyer sur le terrain qualitatif pour caractériser et expliquer les ”propriétés sociales” des travailleurs des données de Madagascar. Par ailleurs, sur la base de la revue de littérature de [Blanchard et Patou \(2003\)](#), nous proposons plusieurs éléments méthodologiques pour limiter les risques liées à l’analyse factorielle.

En ce qui concerne le paramétrage de l’analyse factorielle, les données manquantes ont été imputées à l’aide du package R MissMDA. Comme c’est toujours le cas avec l’imputation, cela repose sur l’hypothèse que la distribution des données manquantes est aléatoire (qu’aucune variable observée ou non observée ne peut expliquer le fait que la valeur soit manquante). Nous avons réalisé l’ACM à l’aide du package R FactoMineR [Husson, Josse, et Lê \(2008\)](#). Pour dériver des catégories structurelles discrètes à partir de notre axe indépendant, nous avons donc appliqué des méthodes de regroupement géométrique sur les coordonnées des travailleurs de Madagascar. Le nombre d’axes pertinents pour notre analyse, ainsi que le nombre de clusters pour la partition ont été choisis en utilisant la méthode du point d’inflexion, basée respectivement sur le pourcentage de variance expliquée pour l’ACM et le *screeplot* pour le regroupement des individus (*clustering*). Pour les deux projets, le scree plot des k-means a suggéré une coupe entre trois et quatre clusters. Les décisions de partition ont finalement été prises en fonction de la pertinence analytique des clusters. Un autre point d’attention réside dans le choix des variables et des modalités retenues qui doivent être de nature à permettre une explicabilité suffisante des axes et clusters retenus tout en préservant les objectifs initiaux de l’ACM. Dans ce premier cas : comprendre qui sont les travailleurs et tenter d’expliquer comment sont-ils devenus annotateurs ?

La [table 1](#) résume les variables retenues pour répondre à ces questions. Nous avons sélectionné des variables socio-démographiques (âge, genre, éducation), des variables économiques (revenu du foyer, salaire comme annotateur), le score du générateur de position qui sert de *proxy* pour le milieu social d’origine des annotateurs. Plusieurs combinaisons de variables ont été testées de manière à conserver le maximum de variance et donc de puissance explicative de l’analyse factorielle malgré la faiblesse de l’échantillon. Il a ainsi été nécessaire de recoder les modalités de manière à ce que chaque modalité possède des effectifs suffisants. Par exemple, les trois variables liées aux études ont été ”dichotomisées” selon que les individus aient fait des études supérieures ou non. Concernant la variable ”Situation matrimoniale”, nous avons également re-

groupé les personnes "veuves" et "divorcées" dans la variable "célibataire" tenant compte du fait que si ces personnes avaient été avec une autre personne, elles auraient plutôt choisies "en couple". Par ailleurs, nous avons découpé les variables continues utilisées en classes de taille égale.

Variable	Types	Niveaux	Catégories
Âge	Facteur	18-24, 25-34 et 35+ ans	Socio-démographique
Situation	Facteur	En couple, Marié, Célibataire	Socio-démographique
Foyer parents	Facteur	Oui, Non	Socio-démographique
Foyer enfants	Facteur	Oui, Non	Socio-démographique
Foyer parents	Facteur	Oui, Non	Socio-démographique
Foyer conjoint	Facteur	Oui, Non	Socio-démographique
Étudiant	Facteur	Oui, Non	Culturel
Études supérieures	Facteur	Lycée ou avant, Supérieur	Culturel
Salaire ¹	Facteur	X	Économique
Revenus du foyer ¹	Facteur	X	Économique
Position sociale ¹	Facteur	X	Social

¹Ces variables continues ont été découpés en classes de taille égale

TABLE 18 – Descriptifs des variables utilisées pour l'ACP

Analysons maintenant plus en profondeur la composition sociale, économique et culturelle de notre échantillon. Dans notre ACM, plusieurs variables contribuent proportionnellement plus que les autres au premier axe : la composition du foyer avec la présence ou non d'un enfant, des parents ou du conjoint dans le foyer, le fait d'être marié ou célibataire et l'âge. Concernant le deuxième axe, le salaire, le niveau d'éducation, les revenus du foyer et le niveau d'études sont les variables qui participent le plus à sa construction. Les deux graphiques qui suivent présentent la contribution des modalités aux axes.

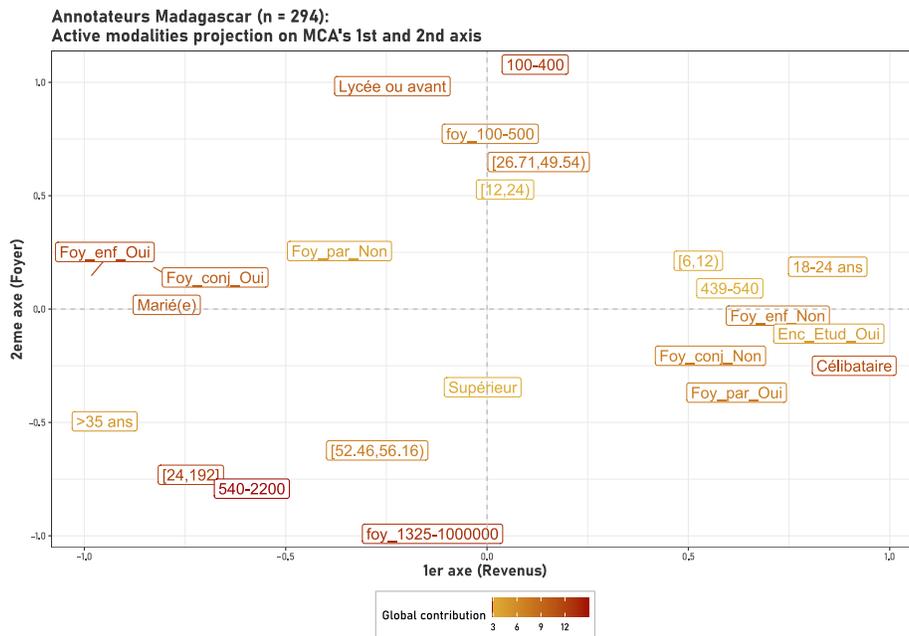


FIGURE 40 – Contribution des variables de notre questionnaire à l'ACM. Lecture : la position des variables sur les axes déterminent indiquent les variables déterminantes dans la composition des axes

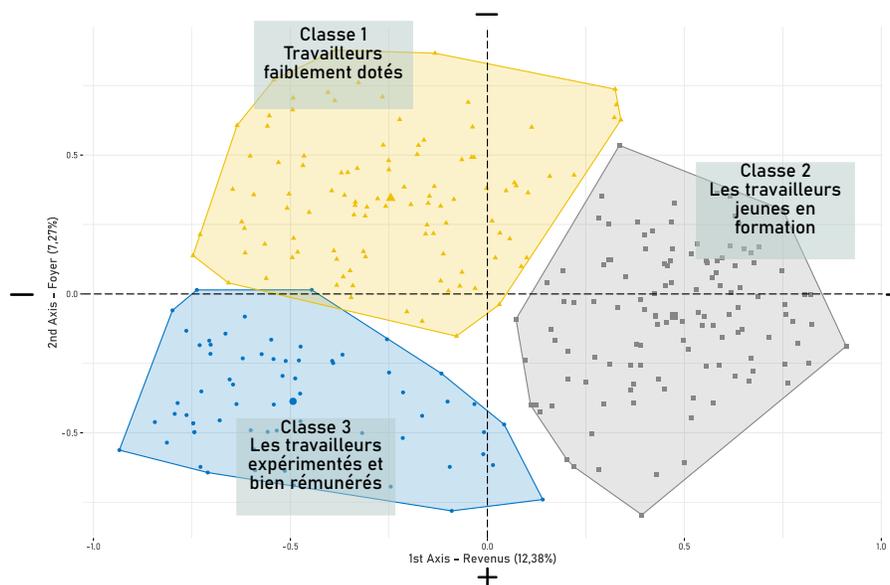


FIGURE 41 – Classification des annotateurs en trois groupes homogènes. Lecture : la position des individus est déterminée par les variables qui composent les axes.

On distingue alors trois catégories de travailleurs. Les deux premières classes sont très ressemblantes sur le plan du genre avec majoritairement des hommes alors que la troisième classe se distingue par une sur-représentation des femmes (Table 52). Il s'agit également de la classe qui comporte le plus d'annotateurs de plus de 35 ans. La classe 2 se distingue au contraire par la sur-représentation d'annotations plus jeunes (18-24 ans) (Table 53), ainsi que par la part importante d'annotateurs célibataires, habitant chez leurs parents et pour la totalité d'entre eux sans enfants (Table 54, Table 55). Au contraire, on retrouve, au sein de la classe 1 et 3, davantage de personnes mariées habitant avec leurs enfants. Concernant notre variable culturelle, on retrouve une proportion moins importante d'annotateurs diplômés du supérieur dans la classe 1 et davantage d'annotateurs encore étudiants dans la classe 2 (Table 57, Table 58). Enfin, le capital économique et social est distribué entre les classes de la manière suivante : la classe 1 a un faible capital économique et une faible position sociale moyenne, la classe 2 a un capital économique et social équilibré tandis que la classe 3 se distingue par un meilleur capital économique et social (Table 59, Table 60). Si l'on suit la littérature sur l'utilisation des générateurs de position (Lin & Erickson, 2008), notons d'ores et déjà que la position sociale relative de nos classes illustre une forte distinction entre des annotateurs qui ont des ressources sociales importantes pour leur vie professionnelle (classe 2 et 3) et des annotateurs qui ont des ressources beaucoup plus faibles.

Comment qualifier ces classes ? La première classe se caractérise par des individus plus jeunes, qui sont plutôt en couple ou mariés et qui habitent avec leurs conjoints et leurs enfants. Ils sont moins bien dotés en capital économique et social que les autres classes. On appellera cette classe les *"travailleurs faiblement dotés en capital"* (35,3% de l'échantillon). Si l'on reprend la terminologie utilisée dans la littérature sur le développement de la classe moyenne à Madagascar, on peut également qualifier cette classe de *"petite précarité"*. La deuxième classe se caractérise par la présence de travailleurs jeunes qui sont célibataires et habitent avec leurs parents (42,7% de l'échantillon). Ils sont plutôt bien formés et certains sont d'ailleurs encore étudiants et viennent d'entrer dans l'activité d'annotateur. Il s'agit des *"jeunes travailleurs encore en formation"* que l'on peut également considérer comme étant *"la petite prospérité urbaine"* en référence au processus de construction des qualifications à l'oeuvre au sein de cette classe et qui peut constituer une forme *"d'asset"*. La troisième classe se caractérise par un plus haut niveau de revenus du foyer, un salaire plus important, ainsi que par un niveau en français et d'éducation des parents plus élevés. Ils sont moyennement âgés, plutôt mariés et dans le secteur depuis beaucoup plus longtemps (22% de l'échantillon). Il s'agit des *"annotateurs installés et plus privilégiés"* qui correspondent alors à une véritable *"classe moyenne urbaine"*.

Pour comprendre encore un peu mieux le sens de ces classes sociales, il est utile de revenir à nos données qualitatives. Ainsi, deux cas de figure illustrent selon nous la première classe. Un exemple de cette classe est assez spécifique au sous-traitant G qui possède un centre à Fianarantsoa dans lequel les annotateurs ont un plus faible niveau d'études et parlent bien moins français. Concernant notre classe 2, nos résultats

suggèrent que le travail d’annotateur peut être perçu comme un travail d’appoint, notamment pour des étudiants habitants encore chez leurs parents, mais également comme un moyen *facile* de commencer à travailler, d’acquérir de l’expérience et de financer leurs études pour les jeunes malgaches de la capitale. Dans cette classe, on retrouve certains annotateurs qui vivent une double voire une triple journée, du fait de leur engagement dans une activité principale rémunératrice, de leurs études, parfois d’une deuxième activité professionnelle (voir ci-dessous). On y retrouve un résultat identifié à propos des micro-travailleuses françaises sur la triple journée des femmes (Tubaro et al., 2022).

Encadré n°73 : La double voire la triple journée des annotateurs encore étudiants

Z : Présentez nous, un peu votre parcours au lycée, à la fac ... ?

A : Comme parcours/études ... en ce moment je travaille encore à l’université de Madagascar, cinquième année en faculté de sciences dans un parcours de pharmacologie et cosmétologie. Je pense que je vais finir mes études cette année. A part ça, (...) je travaille ici en tant qu’annotateur (...) Je travaille ici et j’étudie encore aussi. Et j’ai encore un autre travail, je suis enseignant de sciences de la vie et de la terre dans un collège français. (...) je suis un agent, dans un shift de nuit, je travaille à partir de minuit, jusqu’à neuf heures du matin, et après je prends quelques heures de sieste et après je continue les autres postes. C’est un peu compliqué, mais je gère(...) Z : et pourquoi tu gardes ce double boulot ?

A : Parce que je suis orphelin depuis 2019 et aussi c’est moi qui m’occupe de ma petite sœur et elle est encore à l’université aussi alors pour la location du foyer, pour l’alimentation etc., il faut beaucoup d’argent, alors je sacrifie un peu, parce que pour avoir quelque chose de bon, il faut sacrifier un peu. (...)

Bodo - Annotateur - Sous-traitant A

Concernant les travailleurs de la classe 3 (*“les travailleurs expérimentés et bien rémunérés”*), ils illustrent l’ancienneté de l’industrie de l’externalisation à Madagascar. Cet élément explique la présence dans notre échantillon de travailleurs plus vieux qui ont parfois commencé à travailler dans cette industrie il y a déjà plus de 15 ans. Certains annotateurs sont ainsi entrés dans le secteur au début des années 2000, souvent avec au maximum le baccalauréat et ont continué à y travailler sans forcément systématiquement bénéficier d’une progression salariale. C’est notamment le cas de plusieurs femmes interrogées par le biais de l’enquête Facebook, qui expliquent avoir vu leurs revenus passer seulement de 200000 ariary à 250000 ariary entre le début des années 2000 et aujourd’hui. L’évolution des activités externalisées à Madagascar les ont parfois conduits à travailler comme annotatrices. Pour ces dernières, la relative stabilité de ces emplois et des salaires par rapport aux activités informelles peut expliquer qu’elles restent dans ce secteur. Dans cette classe plus haut niveau de revenu correspond au fait que leur conjoint travaille aussi.

C'est également le cas Serge qui a eu son baccalauréat en 2006, avant de poursuivre par un Brevet de Technicien Supérieur en commerce international. Les effets de la crise politique de 2009 qui a entraîné de forts ralentissements économiques, notamment de la part des investisseurs étrangers, le pousse à monter sa propre entreprise de communication et d'événementiel qui selon ses mots a été un "échec". Il travaille une première fois dans une entreprise d'externalisation à partir de 2012, mais doit quitter l'entreprise à cause d'un "accident" : "après 3 ans (...), quelqu'un m'a poignardé dans le dos (...) en rentrant de mon travail". Après un autre projet entrepreneurial qui échoue, car il avait "des difficultés à trouver des clients", il rejoint à nouveau le secteur de l'externalisation et devient annotateur pour le sous-traitant B, ce qu'il considère comme un travail stable, puisque lorsque nous lui demandons ce qu'il pense de son activité, il répond que "j'encourage les jeunes à y rentrer s'ils cherchent un emploi stable". Pour certains profils mieux formés et d'une origine sociale plus élevée, ce secteur a également pu être l'occasion d'avoir des opportunités de progressions professionnelles et salariales en devenant manager. C'est le cas pour Mevo, qui, après un bac en 2011 et quelques années universitaires qui ne lui permettent pas d'obtenir un diplôme, rejoint une première entreprise du secteur de l'externalisation en faisant de la modération de fiches produits sur des sites web. Après deux ans dans cette entreprise, il rejoint une importante firme d'externalisation dans laquelle il réalise des tâches d'annotations de rédaction de contenus pour le web. Au travers de cette activité, il finit par monter les échelons pour devenir formateur, puis "chargé de formation et de qualité" et enfin "responsable de plateau". Dans tous les cas, pour les individus des classes 1 et 3, le travail d'annotateur peut constituer un travail stabilisant les revenus du foyer lors de l'arrivée d'un ou plusieurs enfants et surtout lors du mariage, traditionnellement synonyme de départ du foyer familial, mais aussi de frais plus importants en termes de loyer et de nourriture.

Comment analyser ces premiers résultats concernant les classes sociales impliquées dans le travail des données à Madagascar? De manière générale, on constate que notre population s'éloigne de ce qui est observé en population générale de par son niveau d'études. Par l'analyse de notre générateur de position, nous constatons une distribution qui révèle la présence d'une forte proportion d'annotateurs appartenant à une "prospérité urbaine", avec une forte concentration des individus autour des scores moyens du générateur de position. Le résultat principal de notre analyse par les classes sociales, à savoir que le fait que cette industrie repose sur l'émergence d'une classe moyenne urbaine, paraît toutefois compliqué à analyser tant il est difficile de trouver des éléments de comparaison dans la littérature (Bonnefond & Andrianampiarivo, 2020). Nous pensons toutefois que nos résultats suggèrent l'émergence d'une classe moyenne urbaine en transition, elle-même composée de deux sous-classes, une classe moyenne urbaine bien plus prospère économiquement et dotée socialement, et une petite prospérité urbaine qui tente d'assurer ses moyens de subsistance, mais qui reste soumise aux aléas des crises politiques, des crises environnementales ou de la crise sanitaire du Covid-19.

Au travers de notre analyse en composantes principales, nous avons utilisé différentes variables socio-démographiques afin de qualifier encore davantage les groupes de travailleurs impliqués dans le travail des données. Une première classe est constituée de travailleurs faiblement dotés économiquement et socialement qui sont majoritairement en couple. Notre deuxième classe regroupe des jeunes travailleurs en formation ou à peine sorties des études - mais avec pour certains de ces derniers l'ambition de les continuer en parallèle de leur travail d'annotateur ou après avoir économisé un peu d'argent - et qui habitent chez leurs parents. La dernière classe concerne enfin des travailleurs qui sont plus expérimentés qui sont mariés et ont des revenus plus élevés.

Comment peut-on donner du sens à ces résultats au regard d'une analyse par la classe sociale? Nous pensons que notre première classe peut être qualifiée de "*petite précarité urbaine*", une classe qui se caractérise par de faibles ressources sociales pour trouver du travail et par de faibles moyens économiques, malgré la présence de familles et/ou de couples dans cette classe. Si l'on se réfère maintenant à notre deuxième classe, on peut considérer qu'elle peut être classée comme étant une "*petite prospérité urbaine*", dans le sens où il s'agit encore de jeunes qui sont en formation et qui construisent leurs qualifications, qui pourront éventuellement constituer des "*assets*" pour gagner davantage d'argent. Bien sûr, nous détournons ici le concept de petite prospérité, qui désigne plutôt des personnes qui maîtrisent leurs moyens de production et qui font référence à l'agriculture. Nous considérons toutefois que les travailleurs de cette classe (des jeunes qui habitent chez leurs parents et qui se forment), peuvent constituer une classe à part entière dont la prospérité repose sur la capacité à évoluer dans le secteur de l'externalisation. Enfin, notre dernière classe peut être catégorisée comme étant une "*classe moyenne urbaine*", dont les composantes ont certes des difficultés économiques liées à la taille de leurs foyers, mais qui reste relativement stable d'un point de vue économique et surtout qui a davantage de ressources sociales pour trouver du travail.

D'un point de vue plus général, nos résultats suggèrent également que le secteur de l'externalisation s'appuie très largement sur les membres d'une classe moyenne, profitant de leur formation et des besoins économiques liés à leur trajectoire de vie, notamment pour les classes 1 et 3. Le secteur profite également du fonctionnement traditionnel de la société malgache, où les jeunes ont tendance à rester habiter chez leurs parents tant qu'ils n'ont pas fondé une famille ce qui donne davantage de possibilités pour les entreprises de recruter ces jeunes prêts à accepter des salaires moins élevés (classe 1). D'un point de vue empirique, notre travail permet d'illustrer d'autres approches pour la compréhension des classes moyennes dans les pays du Sud global, qu'une analyse par le salaire et les habitudes de consommation. À la lumière de la controverse scientifique sur le rôle des classes moyennes dans le développement des pays du Sud global, nous illustrons l'émergence d'une classe moyenne urbaine liée au secteur formel de l'externalisation de services et qui dispose de ressources sociales non négligeables dans leur développement professionnel.

9.1.2 Un secteur qui permet une *stabilisation professionnelle et financière* des travailleurs malgaches

La présence de cette classe moyenne au sein des entreprises de l'externalisation interroge quant à la structure extractive du secteur ce qui nous amène à nous poser la question suivante : le travail d'annotation est-il un travail de qualité permettant une sécurisation financière^{BT}. De ce point de vue, nous rejoignons des questionnements propres aux économistes du développement et notamment de Banerjee et Duflo (2008) qui considèrent que si : "*la classe moyenne est importante pour la croissance*", car elle se caractérise par une stabilité professionnelle qui permet d'être attentif à sa carrière et à ses enfants (Banerjee and Duflo 2008, 26). Ces caractéristiques soulèvent selon nous la question de la trajectoire professionnelle des annotateurs. Autrement dit, dans quel type de trajectoire cette activité s'inscrit-elle? Quelle est la carrière d'un annotateur de données?

La notion de carrière est ancrée dans la position qui consiste à pondérer la capacité d'action des individus avec les dispositions qui orientent leurs conduites de vie. Le travail de Peter Berger et Thomas Luckmann (1966) permet d'opérer un dépassement de l'explication des conduites, par l'individu d'un côté ou par le collectif de l'autre, en considérant la force des interactions et de leurs effets sur l'individu dans le cadre de ses activités sociales. Dans quelles mesures la place de l'individu dans une organisation va contribuer à orienter une conduite future? Un enjeu consiste également à tenir compte des accidents et des imprévus qui surviennent au cours d'une vie et qui modifient parfois les trajectoires des individus (Leclerc-Olive, 1997; Grosseti, 2006). On parle aussi de "*chocs biographiques*" pour rendre compte de ces changements structurels qui dépassent l'inertie des trajectoires censée amortir des changements. L'analyse de séquences associée à nos entretiens nous permet d'articuler les effets structurels et individuels à l'origine des bifurcations professionnelles des annotateurs.

Ces dimensions liées aux trajectoires professionnelles sont rarement explorées dans les études sur le travail numérique, mais sont essentielles si l'on veut comprendre le rapport entre cette activité et l'évolution professionnelle des travailleurs : s'agit-il d'une précarisation du travailleur ou non au regard de ses activités précédentes? Quels sont les facteurs qui déterminent cette trajectoire et quelles sont les conséquences de ces trajectoires pour les travailleurs? Pour répondre à ces questions, nous nous appuyons sur deux sources principales : les variables de notre questionnaire relatives à la trajectoire professionnelle et nos entretiens qui retracent le parcours des annotateurs de données. Pour rappel, notre questionnaire comporte un module "*trajectoire professionnelle*" dont l'objectif est de collecter des informations relatives aux différentes expériences de notre population : les travailleurs devaient ajouter leurs expériences précisant l'intitulé de poste, le salaire, le type de contrat de travail et l'année de début de leur contrat. Sur la base de ces variables,

311. "A good job is a steady, well-paid job—a job that allows one the mental space needed to do all those things the middle class does well" (Banerjee & Duflo, 2008)

nous avons procédé à un recodage manuel afin de réaliser une analyse de séquence. Avant de revenir sur ce processus, présentons brièvement en quoi consiste cette méthode d'analyse.

L'analyse de séquence est une méthode statistique dont l'objectif "consiste principalement à décrire et explorer les parcours de vie, à en identifier les régularités ou les différences." (Robette, 2011) (p.11). Elle s'inscrit dans la perspective d'études centrées sur les trajectoires biographiques plus facilement réalisables grâce à la disponibilité accrue de données longitudinales. D'un point de vue théorique, l'analyse de séquence se positionne en réponse aux analyses centrées sur les "transitions" qui étudient davantage les événements marquants (les "turning point") qui déterminent les choix relatifs à la carrière³¹². L'analyse de séquence permet d'appréhender les événements "dans leur enchaînement", un parcours de vie "comme un tout", afin de dégager les facteurs explicatifs de son déroulement³¹³. Nous appliquons cette méthode afin de mieux comprendre deux aspects des carrières d'annotateurs : les perspectives ou non de stabilisation professionnelle à travers le passage entre les contrats informels et formels et la circulation des travailleurs entre les activités de l'externalisation de services numériques. Pour ce faire, nous avons procédé à un recodage manuel des intitulés de postes remplis par les annotateurs dans le module *trajectoire professionnelle*. Nous obtenons alors 8 catégories qui sont autant d'états potentiels.

Nom de la variable	Exemples
Contrat exter annot	CDD et "Annotateur"
Contrat non exter	CDI et "Vendeur"
Contrat exter annot cadre	"Superviseur"
Contrat exter autre	"Agent centre d'appel"
Sans contrat exter annot	"Annotateur"
Sans contrat exter autre	"Agent centre d'appel"
Sans contrat non exter	"Magasinier"
Contrat exter autre cadre	"Responsable de production"

TABLE 19 – Recodage des variables du module parcours professionnel

Le principal problème de notre échantillon étant les durées de carrières différentes, nous avons sélectionné une durée de six ans. Puis, les données manquantes ont été imputées à l'aide du package R MissMDA. La durée utilisée correspond à l'optimum constaté, car au-delà, l'imputation est trop aléatoire. Avant d'analyser les classes de trajectoires obtenues, revenons sur quelques éléments généraux concernant notre échantillon.

312. "Dans la pratique, une grande partie des travaux empiriques sur les parcours de vie en sciences sociales sont fondés sur des méthodes centrées sur les transitions. Il est cependant possible d'adopter une autre approche, offrant la possibilité de les étudier en tant qu'unité d'analyse, comme un tout." (Robette, 2021)(p.12)

313. L'analyse des trajectoires d'emplois constitue une application largement utilisée de cette méthode (Robette, 2021)

La durée moyenne d'ancienneté sur le marché du travail est de 6,4 ans. Cette durée d'ancienneté masque néanmoins le faible nombre d'emplois moyens occupés par les travailleurs de notre échantillon : 1,7 emploi occupé avant leur activité actuelle.

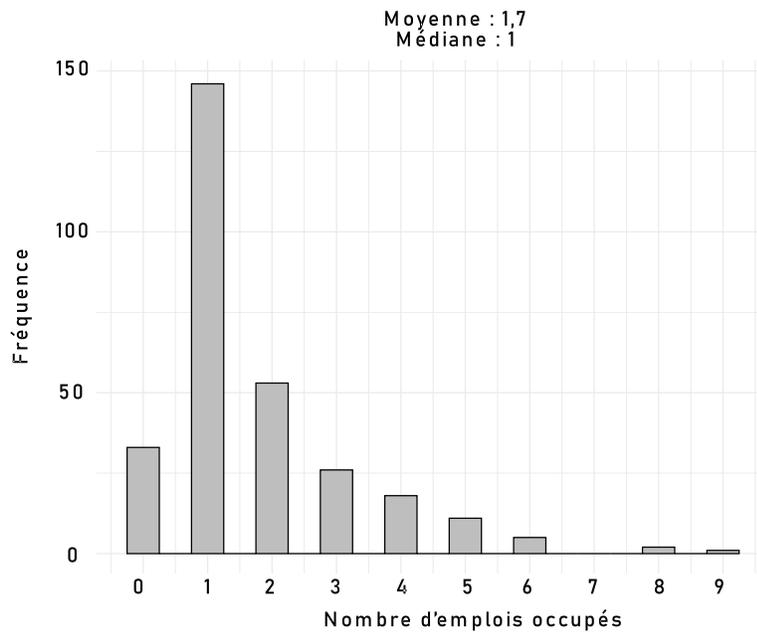


FIGURE 42 – Le nombre d'emplois occupés par les annotateurs de données

Dans le détail, on constate une différence importante entre les entreprises étudiées. Ainsi, en matière d'emplois occupés, la médiane de l'entreprise G est de deux emplois, quand elle n'est que d'un emploi pour les autres. En revanche, l'ancienneté sur le marché de l'emploi est de 4,5 ans pour les travailleurs de l'entreprise G, 4 ans pour l'entreprise A et 6 pour l'entreprise B. Globalement, on constate donc que les travailleurs sont davantage inexpérimentés dans les deux entreprises les plus petites de notre échantillon.

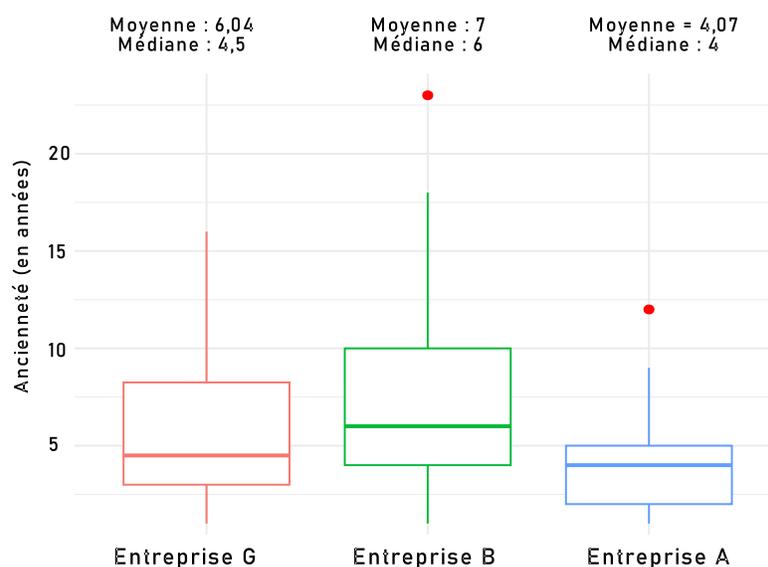


FIGURE 43 – Comparaison de l’ancienneté des annotateurs sur le marché du travail entre les trois entreprises étudiées à travers notre questionnaire

Analysons maintenant les trajectoires professionnelles des annotateurs. La première classe se caractérise par des annotateurs qui commencent leur carrière dans le secteur de l’externalisation directement après leurs études ou le bac, en étant sous contrat ou pour une plus faible proportion sans avoir de contrat (*“entrée directe dans l’externalisation”*). Au fil des années la plupart deviennent annotateurs, certains alternant toujours entre l’annotation et l’externalisation, mais de manière minoritaire. Le deuxième type de trajectoire professionnelle est appelé *“entrée rapide dans l’annotation de données”*, car après un an passé principalement en dehors de l’externalisation comme employé, les travailleurs de cette classe rejoignent le secteur de l’annotation de données. On note toutefois une alternance avec des postes dans le secteur de l’externalisation, ce qui confirme la circulation des travailleurs entre les différentes activités présentes dans les entreprises de sous-traitance de services. La troisième classe comprend des travailleurs qui mettent davantage de temps à intégrer le secteur de l’externalisation. Ils s’illustrent par leur alternance entre des expériences dans l’externalisation et des activités dans d’autres secteurs avant une stabilisation professionnelle dans l’annotation de données. Notre dernière classe se caractérise par des travailleurs qui démarrent sans contrat de travail en dehors du secteur de l’externalisation. La grande majorité des activités à Madagascar se faisant dans le secteur informel, il est courant que les travailleurs ne bénéficient pas d’un contrat de travail. Puis, ces travailleurs rejoignent le secteur de l’annotation et se stabilisent. Pour donner du sens à cette analyse, nous avons utilisé nos entretiens en sélectionnant ceux réalisés avec les travailleurs qui composent chaque classe.

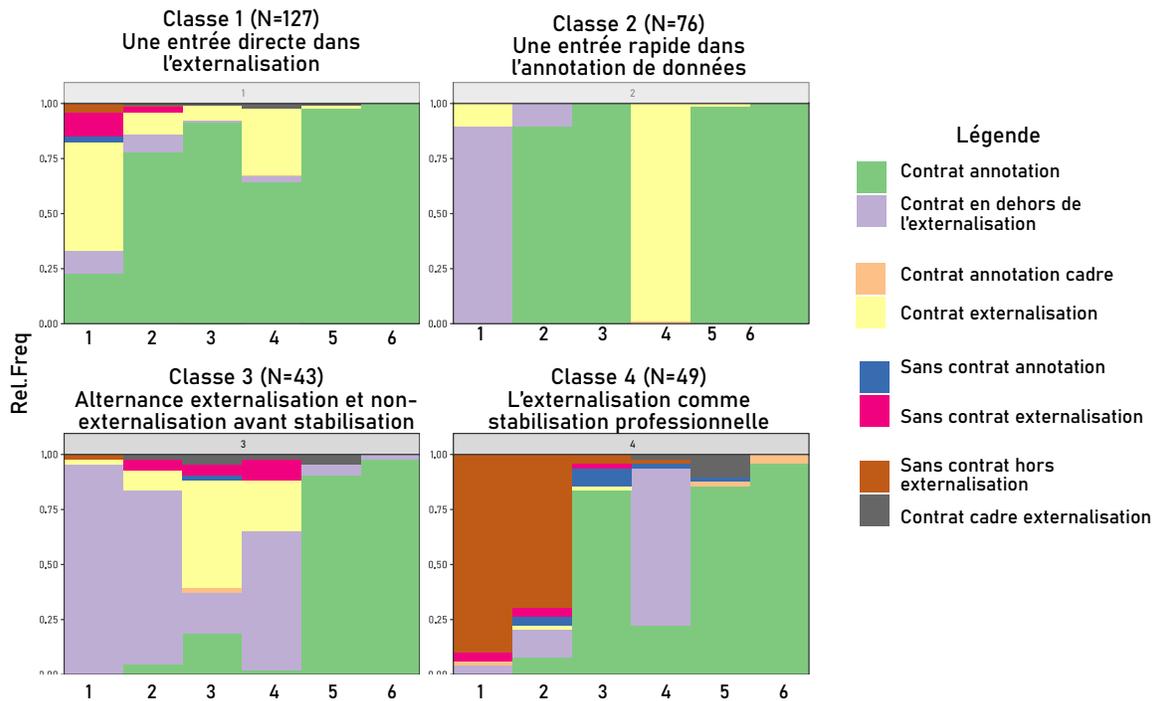


FIGURE 44 – Typologie des trajectoires professionnelles des travailleurs des données. Lecture : chaque classe est constituée d'individus homogènes dans leur trajectoire professionnelle et est déterminée par la fréquence d'individus dans chaque "état"

Pour notre première classe, nous avons choisi Agathe, annotatrice chez le sous-traitant A au moment de l'entretien. Elle obtient son baccalauréat en 2012, puis étudie l'anglais pendant deux ans au sein d'un programme de formation professionnalisant privé. Après cela, elle cherche un travail pendant deux ans sans succès. Puis, elle commence sa carrière professionnelle en intégrant le sous-traitant E au sein de laquelle elle réalise des tâches d'annotations pour Amazon. Elle rejoint ensuite l'entreprise Euphonie où elle doit analyser "les comptes" des nouveaux utilisateurs et "supprimer les comptes fakes" pour un site de rencontre français. Elle décide ensuite de rejoindre son entreprise actuelle, car elle avait "un ami qui travaillait ici avant moi, et il m'a parlé du travail, du salaire et tout, et j'ai postulé" en s'appuyant sur les compétences numériques acquises lors de ses précédents emplois.

En ce qui concerne la deuxième classe, Eja est un annotateur du sous-traitant A. Après le lycée et un diplôme en électronique, il décide d'aller à l'université dans un cursus de "programmation électronique". Malheureusement, à cause d'un "problème de financement" de sa scolarité, il est obligé d'arrêter. Il trouve un premier

emploi de vendeur en CDD qu'il quitte assez rapidement, car : *"le salaire ne [permet pas]de subvenir aux besoins de ma famille"*, il touche alors 150000 ariary et décide ensuite de postuler dans son emploi actuel, par nécessité : *"Vous savez bien que la vie à Madagascar, c'est un peu difficile à vivre"*. Il est alors embauché sur un contrat de longue durée. Il est intéressant d'analyser la manière dont il se projette dans l'avenir. En effet, s'il reconnaît que chez le sous-traitant A, il *"obtient beaucoup de salaire"* pour un travail *"simple"*, mais *"monotone"*, il est conscient de risquer le licenciement : *"y a pas d'avenir de travailler [ici] parce que je vous avais déjà dit qu'il y a des risques très élevés de perdre mon travail l'année prochaine."* Eja exerce en effet des fonctions d'annotateurs sur le projet de vidéo-surveillance évoqué dans nos précédentes parties. Or, ce projet est très délicat, la moindre erreur commise par un annotateur peut être sanctionnée, les employés de la start-up arguant que cela a de fortes conséquences sur leurs clients : au bout de trois erreurs, les annotateurs peuvent ainsi être licenciés. Il considère par ailleurs qu'il sera compliqué pour lui d'atteindre des fonctions managériales au sein de l'entreprise à cause de son manque de diplôme : *"Si je me vois évoluer mais vous voyez bien, y a un problème de financement alors je n'ai pas obtenu beaucoup de diplômes c'est pour ça que je ne peux évoluer [ici] parce que pour être encadrant la limite c'est bac +2 (...) c'est comme responsable de l'activité, il faut au minimum Bac +4"*. Finalement, ils souhaitent *"obtenir plus de diplômes comme [un] master en électronique et informatique et je veux travailler en indépendant"*.

La troisième classe se caractérise par une alternance entre des activités dans le secteur de l'externalisation et d'autres hors de ce secteur. Nous rencontrons par exemple Simon qui travaille comme chef de projet chez le sous-traitant G. Après un baccalauréat obtenu en 2011, il réalise un double BTS en tourisme et en gestion de projet. Il poursuit son cursus de gestion jusqu'à la licence, avant d'arrêter *"à cause de soucis financiers."* A ce jour, il souhaite d'ailleurs toujours obtenir un master. Il est d'abord manutentionnaire de fournitures de bureaux dans une société en CDI, et passe ensuite presque une année en CDD comme magasinier dans une entreprise du secteur textile. Il explique avoir arrêté ce contrat, car : *"c'est assez difficile, bien tout voir, car si on fait une erreur c'est vraiment coûteux. Là c'était aussi difficile de travailler avec d'autres personnes ... Dans le monde des industries textiles, c'est assez difficile ici à Madagascar tant sur des conditions de vie et beaucoup de choses ... Donc j'ai arrêté après 11 mois, justement par rapport à ça."* Il ajoute encore que les conditions de travail difficiles sont très courantes dans ce secteur : *"le souci, c'est surtout par rapport au supérieur hiérarchique qui faisait des abus assez constants, qui directement insultait les gens de par leur race (...)"* Mais ces situations sont acceptées par les travailleurs, car d'après lui, *"les gens n'ont pas le choix"*. Dans son cas, la situation est d'autant plus compliquée qu'il commence à *"entrer dans le monde du mariage, après j'ai eu (...) mon premier enfant, donc là j'étais obligé de prendre ce qu'il y a."* Entre 2018 et 2019, il obtient un poste en CDD pour travailler en tant que superviseur sur le plan de recensement général de la population malgache, mission qui l'amène à souvent partir en déplacement, ce qui lui pose problème vis-à-vis de sa vie de famille. Il décide ensuite de rejoindre le secteur de l'externalisation dans lequel il enchaîne des petites missions, notamment chez le sous-traitant C : *"Je voulais avoir de l'expérience (...), évoluer rapidement, mais*

c'était pas facile (...) dans ce secteur. Il faut vraiment sortir du lot, il faut avoir beaucoup de choses, beaucoup de bagages aussi. J'ai essayé d'autres projets aussi, dans des petites sociétés malgaches qui ont commencé à faire ça aussi un peu partout. Cette expérience se poursuit au sein d'une autre société d'externalisation où il *"commence à s'épanouir dans le travail"* en tant qu'assistant chef de projet. Malheureusement, au bout de neuf mois, *"la société a fait faillite (...) à cause d'un projet ... y a eu des arnaques de services, là le client n'a pas payé, (...) on n'était jamais payés à la fin du mois"*. Il devient ensuite responsable dans son entreprise actuelle.

Notre quatrième classe se caractérise par un début de carrière pendant lequel les travailleurs n'ont pas de contrat. Nous avons choisi Jo pour incarner ce parcours. Après avoir eu son baccalauréat en 2008, il effectue une première année universitaire en droit et réussit même à passer en deuxième année. Il est bloqué lors de cette année, car il a du mal à suivre le rythme de travail à l'université et qu'il a *"vraiment du mal"* d'un point de vue financier, malgré l'aide d'une de ses tantes. Il s'oriente ensuite vers une année d'études de français dont il est finalement recalé. Sur les conseils de ses parents, il choisit alors de poursuivre des études en enseignement à distance tout en travaillant (2011). Il commence par un travail sans contrat de gestionnaire de cybercafé dans lequel il est payé environ 100000 ariary (30 euros). Néanmoins, il n'est pas systématiquement payé : *"Le problème aussi dans le travail... Parfois, quand c'est la fin du mois, (...) ils disaient qu'il fallait attendre un peu parce qu'ils avaient encore l'électricité et le loyer à payer. La recette n'était pas suffisante pour le mois (...). Parfois ils me payaient la moitié, et j'attendais pour recevoir l'autre moitié. Donc ce n'était pas assez"*. Il décide donc d'arrêter ce travail pour chercher autre chose, mais reste sans activité pendant deux ans. Pendant ces deux années, il ne fait *"rien du tout"*, c'est-à-dire qu'il *"faisait le va-et-vient dans le quartier"* pour acheter *"des trucs moins chers"* qu'il revendait. Par ailleurs, comme de nombreux malgaches, il est alors *"un employé de maison"* : *"parce que mes parents comptaient sur moi pour cuisiner et pour aller au marché, parce que je n'avais vraiment rien à faire (...) Mon père disait qu'il préférait me voir rester là à leur rendre service et faire des petites choses. De mon côté, ça m'embêtait, parce que j'étais vraiment comme un bon à rien. J'étais un garçon pas handicapé et sans travail, alors que je n'ai pas arrêté de chercher, mais je n'arrivais pas à trouver"*. Il considère qu'à cette période il manquait d'expérience pour trouver du travail dans les sociétés malgaches, y compris dans des emplois de vendeur en supermarché. Il est ensuite employé comme serveur dans un bar, toujours sans contrat. C'est au cours de cette expérience qu'il opère une nouvelle évolution professionnelle. En effet, son chef d'alors lui propose de lui payer une formation d'éducateur. Il a alors *"la chance de devenir éducateur dans une école préscolaire"* ce qui le conduit à arrêter *"définitivement les études en téléenseignement"*. Il reste quatre ans dans ce poste, avant de subir la crise sanitaire. En effet, pendant le confinement, son école a fermé et il est donc resté sans travail pendant un an et sans salaire pendant neuf mois. Il a donc décidé de rejoindre la capitale, car *"il a entendu dire qu'il y avait du travail à Tana et que c'était mieux payé"*. Après avoir envoyé de nombreux dossiers d'embauches, il finit par être recruté comme annotateur chez le sous-traitant A.

Que nous apprennent ces différents parcours professionnels ? Nos deux premières classes représentent une population de jeunes qui s'insèrent "rapidement" dans le domaine de l'externalisation et de l'annotation. Néanmoins, l'analyse des entretiens montre que cette "rapidité" apparente cache parfois des périodes d'inactivité non renseignées dans nos données quantitatives. Le cas d'Annaelle, qui explique avoir passé deux ans sans activité, illustre les difficultés d'intégration du marché pour les jeunes diplômés. Notre troisième classe représente des travailleurs qui ont expérimenté plusieurs activités avant de trouver un emploi dans le secteur de l'externalisation. Si cette intégration professionnelle se fait dans un cadre plus formel, elle est néanmoins contrastée par les conditions de travail et parfois salariales difficiles des emplois occupés dans les autres secteurs de l'économie malgache. L'exemple de Simon illustre également les difficultés structurelles du secteur de l'externalisation, existantes dans le secteur formel et criantes dans le secteur informel. Lorsqu'un client fait défaut lors d'une de ses expériences sans contrat, il se retrouve sans rien. Dans d'autres cas, certains secteurs, même publics, sont également précaires et instables. C'est ce que nous raconte Holy qui abandonne l'enseignement, car il est "mal payé", avant de rejoindre une grande entreprise malgache de télécommunication et de poursuivre ensuite sa carrière dans l'externalisation. C'est aussi le cas pour Mamy qui explique que le métier de l'enseignement était "dur" et "pas rentable", car cela dépendait des jours travaillés et donc *"des moments ou des jours ou les élèves ne travaillent pas à l'école, on ne peut pas toucher les sous. Si il y a des vacances on reste tout simplement à la maison, et on ne touche pas d'argent."* Pour lui, l'avantage de son travail dans l'externalisation est de *"toujours toucher du salaire, donc c'est très différent."*

Enfin, la dernière classe analysée est composée de travailleurs ayant exercé des activités informelles et donc sans contrat qui trouvent ensuite des débouchés plus "stables" dans le secteur de l'annotation. Jo a ainsi alterné des périodes de travail pendant lesquelles il gagne peu d'argent avec des périodes d'inactivités avant de se stabiliser en tant qu'éducateur, un emploi qu'il perd à cause de la crise sanitaire. C'est alors qu'il décide de rejoindre l'externalisation et la capitale malgache. Son exemple illustre également la mobilité géographique de certains travailleurs qui viennent d'autres régions de Madagascar, le secteur constituant un débouché typique pour des jeunes arrivés récemment dans la capitale. C'est par exemple le cas de Ihary qui arrive à Antananarivo pour faire ses études, car elle souhaite alors devenir *"une femme indépendante"*. Elle nous explique en effet qu'elle vient *"d'une petite ville où il n'y a rien, juste les zones franches et une société française qui fait des vêtements (...). Donc je me suis dit, « pourquoi pas aller à la capitale, parce que là-bas il y a plus de travail qui correspondent à mon niveau »"* Il s'agit donc de mettre ces éléments en perspective avec les difficultés d'entrée sur le marché du travail pour les jeunes diplômés, pour les travailleurs ne venant pas de la capitale ou pour ceux qui n'ont pas d'expériences professionnelles préalables. Lorsque, comme Michael dans l'encadré qui suit, ils cumulent certaines de ces caractéristiques, l'intégration sur le marché du travail est encore plus "laborieuse".

Encadré n°74 : La difficulté de trouver du travail dans la capitale

Q : « Est-ce qu'il y a des choses que tu as apprises dans tes autres expériences professionnelles qui t'ont aidé chez le [sous-traitant A] ? »

R : « Les choses que j'ai apprises, c'est de rester positif (...). Et aussi de trouver une solution pour survivre (...). Parce que quand je suis arrivé ici, je suis resté 3 mois avant de trouver du travail, et il faut rester positif, même si la totalité est négative. Je me suis dit aussi qu'après la pluie, c'est le beau temps. C'est ce que j'ai appris. (...) je suis resté presque 3 mois à Tana sans savoir si j'allais réussir ni trouver du travail. Et mes parents n'arrêtaient pas de m'appeler en me disant de rentrer. En 3 mois, j'ai maigri, je n'avais plus rien à manger, je mangeais n'importe quoi, je vendais tout ce qui était sur moi (...) mes vêtements, mes chaussures, même mon téléphone, pour pouvoir survivre (...) »

Q : « Et pendant ces 3 mois, est-ce que tu as réussi à trouver des activités à faire, par-ci par-là ? »

R : « À Tana, il y a un endroit de bidonvilles. Je passais souvent là-bas et j'achetais des choses que les gens vendent avec le minimum de sous que j'avais, comme des appareils photo qui datent, de vieux jouets des années 80, des choses que je trouvais intéressantes (...). Puis je les revendais pour avoir un petit bénéfice (...). J'ai réussi à tenir pendant 3 mois, avant de trouver du travail. »

Michael - Responsable - Sous-traitant A

Le secteur de l'externalisation et de l'annotation vient remplir un vide structurel fait d'emplois informels, de périodes d'inactivités, de contrats instables hors des entreprises formelles de l'externalisation. Elle permet alors une "stabilisation professionnelle" notamment en matière de revenus. Une situation qu'explique Hejod du sous-traitant A : *"le 2e point positif, c'est que maintenant je peux être dans une autonomie financière plus ou moins constante chaque mois, c'est ça qui est le plus positif pour moi"*. Néanmoins, cette stabilisation est contrastée par les conditions de travail associées aux contrats de travail formels. Ainsi, si nos données ne permettent pas d'analyser formellement le taux de *turn-over* des entreprises d'annotations, on peut tout de même constater que les personnes étudiées sont dans leur emploi actuel depuis relativement peu de temps (19 mois en moyenne) avec 50% des annotateurs qui occupent cet emploi dans l'annotation depuis 12 mois ou moins. Plusieurs hypothèses contradictoires peuvent expliquer ce que l'on peut considérer comme étant une limite de cette stabilisation. Est-ce que les annotateurs restent peu longtemps dans leur emploi pour trouver quelque chose de mieux dans le secteur de l'externalisation ? S'agit-il d'une conséquence du manque de protection des travailleurs qui feraient face à des licenciements économiques ?

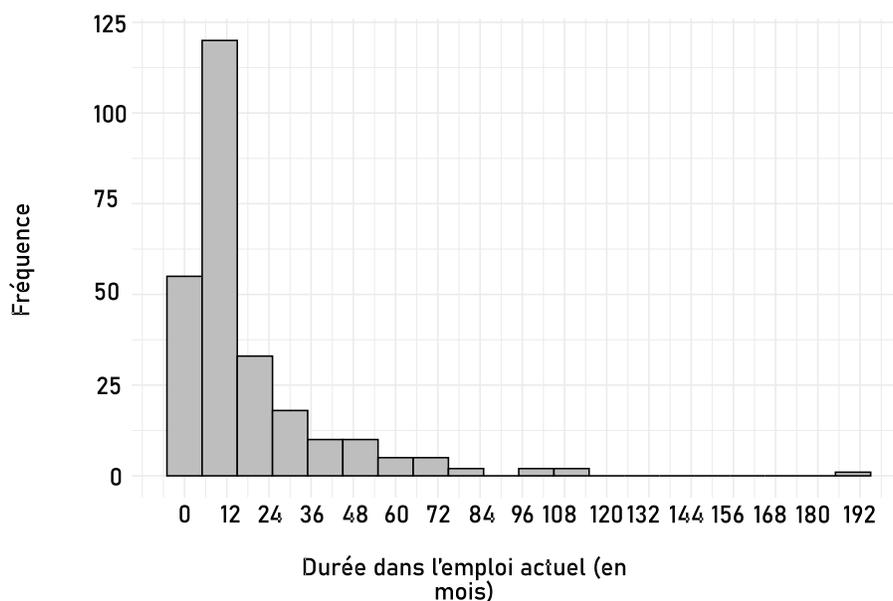


FIGURE 45 – Distribution de la durée dans l’emploi actuel

Deux éléments limitent néanmoins ces résultats. Premièrement, nous avons utilisé une méthode d’imputation pour compléter nos données ce qui limite potentiellement la portée générale de nos résultats. Ensuite, nous avons dû nous focaliser sur un nombre plus restreint de séquences, au risque d’avoir une imputation trop éloignée de la situation réelle de nos annotateurs. En effet, nos données ne proviennent pas d’une analyse longitudinale. Un autre aspect qui limite est cette fois plus cohérent avec la population étudiée : la quasi intégralité des personnes étudiées ont comme dernier état l’annotation de données, car il s’agissait en effet de l’objet initial de du questionnaire d’étudier spécifiquement les annotateurs de données.

Dans cette section, nous avons donc associé une analyse des séquences professionnelles avec l’analyse de profils représentatifs de classes de ces dernières. Deux éléments d’analyse guidaient notre travail : l’externalisation et l’annotation de données comme une forme de stabilisation professionnelle, et la rapidité d’entrée dans ce secteur. Sur ces points, nous constatons que les deux premières classes se caractérisent par une entrée rapide dans le secteur de l’externalisation et par la présence de contrats de travail formels dans l’expérience préalable à leur entrée dans le secteur de l’externalisation. Ces classes de trajectoires professionnelles correspondent aux jeunes travailleurs qui sont encore en formation ou qui viennent d’arrêter leurs études. Que peut-on dire sur cette entrée rapide dans le secteur en termes de bifurcation et en termes de trajectoire personnelle et professionnelle? Nos résultats suggèrent que la trajectoire des travailleurs de ces deux premières classes correspond à une bifurcation personnelle rendue nécessaire par les difficultés

de financement des études. On y retrouve donc énormément de jeunes travailleurs qui ont dû arrêter leurs études soit pour effectuer des formations professionnelles soit pour rejoindre directement le secteur de l'externalisation. Au sein de notre troisième classe, on retrouve des travailleurs qui ont pu intégrer le marché professionnel en ayant accès à des contrats de travail formels (en CDI ou en CDD) mais parfois dans des secteurs dont les conditions de travail sont très difficiles. C'est l'exemple de Simon qui a exercé un travail dans le secteur textile. On retrouve également des travailleurs plus âgés, avec davantage d'expérience professionnelle et qui, pour certains, ont des enfants et donc doivent absolument travailler et avoir une situation professionnelle et économique stable pour élever leurs enfants. Dans ce cas-là, la bifurcation personnelle s'effectue du fait de mauvaises conditions de travail dans d'autres secteurs que l'externalisation ainsi que par la pression de la création de son propre foyer familial. Cette situation les pousse à rejoindre le secteur de l'externalisation, leur permettant d'améliorer leur stabilité financière, en comparaison avec d'autres emplois, du secteur public (par exemple : professeur) ou du secteur privé informel (par exemple : vendeur). Au sein de notre dernière classe, on retrouve des travailleurs qui ont enchaîné des activités professionnelles sans contrat avant de trouver une forme de stabilisation économique par le fait de rejoindre le secteur de l'externalisation. Dans ce cas-là, on retrouve plutôt des profils qui ont eu du mal à intégrer le marché du travail en enchaînant des activités très peu rémunérées dans le secteur informel et avec finalement assez peu de stabilité professionnelle.

* * *

L'analyse développée dans cette sous-section contribue de manière générale à mieux comprendre les effets de l'IA sur le développement et plus précisément à la littérature sur les trajectoires professionnelles des travailleurs des données, et plus généralement des travailleurs précaires dans les pays du Sud global. Nos résultats sur les trajectoires professionnelles des annotateurs de données illustrent l'analyse de [Banerjee et Duflo \(2008\)](#) qui indiquent que rien *"ne semble plus proche de la classe moyenne que le fait d'avoir un emploi stable et bien rémunéré"* et que le *"le sentiment de maîtriser l'avenir (...) que l'on obtient en sachant qu'il y aura un revenu chaque mois - et pas seulement le revenu lui-même [permet] à la classe moyenne de se concentrer sur la construction de sa propre carrière et de celle de ses enfants."* Dans un contexte comme celui de Madagascar, caractérisé par la part importante du travail informel et par les difficultés de l'administration publique à protéger ses citoyennes et ses citoyens et à payer ses fonctionnaires, le secteur privé de l'externalisation constitue une piste pertinente pour des individus recherchant une stabilisation financière. Ce résultat s'articule également avec la littérature sur la constitution d'une classe moyenne dans les pays du continent africain en illustrant le passage rapide entre le secteur de l'agriculture et le secteur des services et la capacité du secteur privé des services à attirer des jeunes diplômés ([Opoku, 2010](#))^{B14}.

314. "les entreprises privées sont devenues plus gratifiantes et plus prestigieuses, et donc plus attrayantes. Alors que les Ghanéens ayant fait des études supérieures recherchaient autrefois des emplois dans le secteur public, beaucoup d'entre eux choisissent aujourd'hui l'entreprise".

Comment analyser nos résultats concernant le rôle éventuel du secteur de l'externalisation dans l'émergence d'une classe moyenne urbaine? Bien sûr, l'émergence de cette classe moyenne, que nous analysons à travers le capital économique, social et culturel, ne doit pas masquer les variations qui existent en sein de notre population. La *"petite précarité urbaine"* (Classe 1) désigne des individus ayant de faibles ressources sociales pour la recherche d'emploi et de faibles moyens économiques, malgré la présence de familles et/ou de couples au sein de cette classe. La *"petite prospérité urbaine en transition"* (Classe 2) est constituée de jeunes encore formation, ce qu'ils envisagent comme une manière d'investir dans leur stabilité financière. Leur "prospérité", contrairement aux petites prospérités agricoles analysées par (Andrianampiarivo, 2018), repose sur la capacité à évoluer dans le secteur de l'externalisation et à valoriser leur "capital humain". Pour ces jeunes malgaches, les principales difficultés à opérer une transition vers une situation plus stable résident dans leur capacité à intégrer un marché du travail très tendu et à être capables de poursuivre leurs études. Enfin, la *"classe moyenne urbaine stabilisée"* est composée de foyers ayant des difficultés économiques du fait de la taille de leur foyer. Néanmoins, ils sont stables économiquement et professionnellement, car ils disposent de davantage de ressources sociales pour trouver du travail.

Nos résultats montrent cependant que le secteur de l'externalisation s'appuie très largement sur l'émergence d'une classe moyenne en profitant de leur formation et des besoins économiques associés à leurs trajectoires personnelles et professionnelles. Le secteur bénéficie également du fonctionnement encore traditionnel de la société malgache : les jeunes habitent avec leurs parents tant qu'ils n'ont pas fondé leur famille. Cette caractéristique permet aux entreprises de sous-traitance de recruter ces jeunes prompts à accepter des salaires moins élevés, n'ayant aucun enfant à leur charge. L'analyse des trajectoires professionnelles suggère finalement que les secteurs de l'externalisation et de l'annotation remplissent un vide structurel lié au marché du travail malgache fait d'activités informelles, de périodes d'inactivité et de contrats instables hors des entreprises formelles de l'externalisation. Ces dernières permettent alors une forme de stabilisation professionnelle et de stabilisation économique qui est bien sûr contrastée par la dimension extractive du secteur.

9.2 Une externalisation protéiforme. Des conditions de travail variables qui dépendent de l'appartenance au secteur formel ou informel ?

Nous avons constaté la variété des trajectoires professionnelles des annotateurs. En plus de jeunes qui se lancent pendant ou à la fin des études, le secteur de l'externalisation accueille des travailleurs de la capitale qui alternent avec d'autres secteurs. On constate une forte circulation entre l'annotation et les autres activités liées à l'externalisation, car ces activités sont bien souvent réalisées dans les mêmes entreprises et nécessitent des compétences similaires. Toutefois, si ces résultats suggèrent que le secteur constitue une des principales voies d'entrée "stables" dans le marché de l'emploi, ils ne permettent pas d'analyser pleinement

le rapport des annotateurs à leur travail. Il convient pour cela de décortiquer les caractéristiques du secteur en matière de conditions de travail.

L'entretien réalisé avec Miky, une ancienne cadre en ressources humaines du sous-traitant C nous semble bien résumer la situation du secteur. Elle nous explique que *"fut un temps ou on percevait les activités d'off-shore [comme] l'équivalent des activités de zone franche."* L'inscription du secteur dans les zones franches est perçue négativement, car cela correspond à *"un travail très dur, parce que c'est des activités de textile (...) C'est se dire que c'est de l'esclavage moderne en fait, entre guillemets, parce que on t'exploite, on t'exploite, mais c'est pas toi qui a les bénéfices. C'était aussi une manière de se dire que y'avait autre chose que ça. Parce que auparavant, quand on sortait des écoles, il y avait que ça sur le marché. C'était une opportunité, mais l'opinion publique ne voyait pas ça comme une bonne chose."* (Miky - RH - Sous-traitant C). Qu'en est-il réellement ? Comment les travailleurs perçoivent-ils leurs conditions de travail ? Est-ce qu'il y a des différences entre les travailleurs ? Et entre les entreprises ?

9.2.1 Analyse objective des conditions de travail dans le secteur de l'externalisation : des variations importantes selon l'appartenance au secteur formel ou informel

Dans cette sous-section, nous revenons tout d'abord sur les conditions "objectives" d'emploi dans les entreprises de sous-traitance étudiées. Nous verrons qu'il existe de fortes différences entre les entreprises selon le caractère formel ou informel de leur activité. Nous détaillons ensuite les différences entre les types de contrats utilisés par les entreprises. Par la suite, nous abordons le ressenti des annotateurs de données à partir de l'analyse des résultats de notre questionnaire et des entretiens. Puis, nous réalisons une typologie des travailleurs en fonction des variables du module "bonheur au travail" et des classes sociales réalisés dans la première sous-section de ce chapitre. L'objectif est d'analyser plus en profondeur le lien entre classe sociale et attitudes vis-à-vis du travail. Nous menons finalement une analyse comparative des conditions de travail dans les trois entreprises étudiées en profondeur afin d'analyser la variation de l'attitude au travail selon le caractère formel ou informel de l'entreprise.

Avant toute chose, il nous faut revenir sur les conditions objectives d'emploi, en analysant les différentes formes de contrats utilisées dans les entreprises de sous-traitance. Nous repartons ici des entreprises que nous avons étudiées plus en profondeur à travers notre questionnaire et nos entretiens (A, B et G) et d'autres entreprises pour lesquelles nous avons obtenu des éléments soit à travers des visites des locaux, soit à travers les entretiens réalisés par le biais de l'échantillon Facebook.

	Taille	Contrat	Projets	Statut	Salaire
A	110	Freelance	IA	Informelle	424K ¹
B	500	Salariés	Variés	Formelle	685K ¹
C	2000	Mixte	Variés	Formelle	250-400K
D	1200	Mixte	Variés	Informelle	300K
E	1000	?	IA	Formelle	300K
F	200	Mixte	IA	Informelle	?
G	70	Salariés	IA	Formelle	445K ¹

¹Pour ces entreprises, il a été possible de calculer la moyenne. Pour les autres, les chiffres se basent sur nos entretiens

TABLE 20 – Conditions d’emploi dans les entreprises de sous-traitance malgaches

Il nous faut ici faire un constat général sur le secteur : il s’appuie globalement sur plusieurs formes de contrats parfois très précaires. Ainsi, même si certaines entreprises ont majoritairement des travailleurs salariés, cela n’empêche pas le recours à des travailleurs ”journaliers” (entreprise B, C, G). Ces derniers sont considérés comme salariés, mais n’ont pas de vacances, car leur salaire est basé sur les jours travaillés. L’entreprise doit quant à elle tenir un registre des travailleurs journaliers. Le contrat se transforme en CDI dès lors que le travailleur a atteint vingt jours de travail consécutif et six mois de présence au même poste. Ce type de contrat est très utilisé par les entreprises pour les projets sans garantie de pérennisation. Ce type de contrat sert aussi à entretenir un vivier de travailleurs qui seront régulièrement appelés, mais de manière discontinue ³¹⁵ ou qui finiront éventuellement par être embauché en CDI ³¹⁶.

D’autres encore ”profitent” de la forte présence du secteur informel pour proposer des contrats de travailleurs indépendants (entreprise A et E). Chez l’entreprise E, le dirigeant français du centre de production nous explique qu’il ”n’a pas encore eu le temps de bien structurer les processus RH” pour justifier le recours à une majorité de travailleurs indépendants Il nous explique également que l’entreprise française qui possède son centre de production travaille également avec d’autres prestataires de services (sous-traitance en cascade) dont les travailleurs ont également ce statut. Au sein de l’entreprise A, il s’avère que les travailleurs

315. ”Chez [sous-traitant B] le contrat de travail ce n’était pas le CDI mais c’est journalier. En tant que journalier je travaille deux ans. Mais quand on travaille [là], on travaille deux mois et puis pas de tâches [il faut attendre de nouvelles tâches].”(Fiary - Sous-traitant B)

316. ”[Q : Depuis 2012 vous avez fait des missions en freelance [journalier] pour cette entreprise ?] A : oui mais c’était pour moins d’un mois à chaque fois, mais ils me recevaient toujours, et quand je pars il n’y a pas de problème (...) c’est l’avantage du contrat freelance ici, il n’y a pas d’entraînement - c’est juste qu’il ne faut pas faire de faute professionnelle, ou des dégâts avec les employés, des petites querelles (...). Quand vous appelez ou que vous venez au bureau (...), ils vous proposent « vous pouvez passer freelance sur ce poste ou celui-là ? », et ils me retenaient toujours. Mais l’année dernière je leur ai dit que j’ai besoin de quelque chose de fixe (...), parce que en fait j’avais déjà 2 enfants l’année dernière, et j’avais besoin d’un emploi stable. Je leur ai demandé s’il y avait un CDI, ils m’ont proposé AlWar (...).”(Aina - Annotateur Sous-traitant B)

se voient proposer de *”faux CDI”*, selon les mots de Jean, un ancien salarié rencontré hors des locaux de l’entreprise à la toute fin de notre enquête chez ce prestataire. Il nous explique alors qu’il a été licencié pour avoir essayé d’informer les salariés sur leur situation contractuelle et que plusieurs d’entre eux ont obtenu une requalification de leur contrat de travail après une saisine de l’inspection du travail. Plus tard, nous avons eu accès à la note d’analyse juridique rédigée par une avocate pour le compte du donneur d’ordres qui précise que *”Bien que le contrat soit intitulé « contrat de prestation de service », le contrat que nous avons en la présente est un contrat de travail dissimulé.”*

Enfin, y compris dans les entreprises qui salarient leurs travailleurs, les possibilités de licenciement sont nombreuses. Que ce soit pour faute lourde (Art 18. Code du travail) ou plus fréquemment pour motif économique, les travailleurs sont finalement assez peu protégés. Dans le second cas, les travailleurs sont fortement dépendants des décisions des donneurs d’ordres comme l’explique Gaëlle : *”[le] projet s’est arrêté et là, ils ont licencié les autres membres de mon équipe, et ensuite ils m’ont appelé pour dire qu’ils ont d’autres choses à me proposer”*. Il arrive en effet que les entreprises changent de prestataires, que certaines entreprises technologiques soient rachetées par un autre groupe et arrêtent leur collaboration avec leur sous-traitant. Dans le cas d’une start-up, une levée de fonds qui se passe mal peut compromettre un contrat de sous-traitance. Les travailleurs se retrouvent alors presque sans rien, puisque le motif de licenciement économique leur permet uniquement de toucher une *”indemnité de licenciement calculée à raison de dix jours de salaire par année complète de service”* (art 28 - Code du travail) et qu’il n’existe pas de droit au chômage. Les travailleurs disposent en revanche d’un certain nombre de droits associés au travail salarié : une durée légale de travail de 42 heures par semaine (Art 75), des congés fixé à 2,5 jours par mois travaillés (Art 86), des indemnités de transport (Art 92), des droit syndicaux (Art 137), le droit d’avoir des délégués du personnel dans les entreprises de plus de de 11 salariés (Art 145).

Par opposition, les contrats de *”freelance”* sont particulièrement problématiques à plusieurs titres. D’abord, les travailleurs ne cotisent pas pour leur retraite et n’ont pas accès à une assurance maladie ni à la médecine du travail. Il s’agit par ailleurs d’un manque à gagner pour l’État malgache qui prélève moins d’impôts sur ces sociétés. Des éléments complémentaires nous permettent d’objectiver la question du contrat de travail. En effet, dans le cadre de notre terrain, nous avons pu suivre une négociation entre une start-up d’IA et son sous-traitant portant sur le passage des annotateurs de prestataires de services à employés en CDI et sur l’augmentation salariale de ces derniers. Dans ce contexte, l’employé de la start-up en charge de la négociation nous a transmis un document permettant de mieux comprendre comment se décompose le salaire des annotateurs. On y constate la différence entre une contractualisation informelle et formelle, que ce soit pour les travailleurs (en termes de droits) et pour la collecte des impôts.

Poste de dépenses	Coûts mode informel	Coûts mode formel
Salaire base	85,3	85,3
Indemnités repas	18,8	18,8
Indemnités transport	12,8	12,8
Primes ¹	19,5	19,5
Ostie salarié ²	0	1,3
Ostie employeur	0	6,83
CNaPS salarié ³	0	1,3
CNaPS employeur	0	17,8
FMFP employeur ⁴	0	1,3
Congés	0	11,4
Assurance maladie privée	19,3	19,3
IRSA ⁵	0	26,8
ISI ⁶	6,8	0
Salaire nette agents	129,7	107,04
Charge salaire employeur	155,8	193,11

¹ La prime n'est pas versé tous les mois à l'ensemble des agents

² L'Organisation Sanitaire Tananarivienne Inter-Entreprise (OSTIE) est un regroupement privé pour la santé au travail

³ La Caisse Nationale de la Prevoyance Sociale est l'organisation de sécurité sociale malgache

⁴ Le Fonds Malgache de Formation Professionnelle est l'organisme public pour la formation professionnelle

⁵ L'Impôt sur les Revenus Salariaux et Assimilés est l'impôt prélevé sur les salaires des travailleurs (à partir de 350000ar)

⁶ L'Impôt Synthétique Intermittent est applicable aux contribuables qui ne payent pas l'IRSA (notamment : travailleurs informels)

TABLE 21 – Structure des coûts de l'entreprise A (en euros) : comparaison mode informel contre mode formel

Si les travailleurs de cette entreprise déclarent travailler en moyenne 37 heures par semaine et bénéficier de 6 semaines de congés, ils sont nombreux à se dire en CDI (89%). Comment expliquer ce résultat alors que les travailleurs d'une des entreprises étudiées sont en fait des prestataires de service? Un premier élément d'explication nous est fourni par Jean, ancien conseiller RH de l'entreprise A, qui nous indique que la propriétaire de l'entreprise trompe ses employés en leur faisant croire qu'ils sont en CDI. La jeunesse des travailleurs de l'annotation et la part importante du travail informel à Madagascar peuvent constituer des facteurs qui expliquent cette situation. Cette dimension se confirme avec l'analyse de l'accès à une assurance employeur par l'entreprise. En effet, on constate qu'au sein de l'entreprise G, de nombreux travailleurs ont répondu exercer en CDI ou en CDD, mais qu'ils n'ont pas accès à une assurance santé, pourtant obligatoire.

Variable	N = 295
Type de contrat	
CDI	235 (89%)
CDD	13 (4,9%)
Travail sans contrat	9 (3,4%)
Autre	5 (2,0%)
N.A	31
Heures travail par semaine ¹	
N.A	37
Nombre semaines de congés ¹	
N.A	36
Assurance	
Oui, payée par mon employeur principal	217 (73,6%)
Oui, autre	4 (1,4%)
Oui, que je paye	15 (5,1%)
Non	51 (17,3%)
N.A	8

¹Moyenne (Médiane)

TABLE 22 – Conditions d'emploi

Cette analyse objective des conditions d'emploi du secteur de l'annotation de données nous offre un paysage contrasté et pointe vers un constat important : même dans les entreprises formelles, les conditions d'emplois restent précaires. Les entreprises, même les mieux établies profitent de la diversité des statuts d'emplois présente à Madagascar pour diminuer les risques liés à la recherche de donneurs d'ordres stables. Ces éléments sont essentiels pour comprendre les conditions de travail des annotateurs. Néanmoins, nos résultats suggèrent que l'application d'un Code du travail dans les entreprises formelles constitue une forme de protection dont ne bénéficient pas les travailleurs informels. En revanche, ces résultats nuancent très largement le constat d'une stabilisation financière par le fait de rejoindre une entreprise formelle. En effet, le Code du travail offre de nombreuses possibilités de licencier les travailleurs (notamment : motif économique). C'est alors la structure du secteur qui renforce la précarité de ces derniers. Il suffit qu'une start-up identifie un problème de qualité dans le travail des données ou tout simplement qu'elle n'ait plus les fonds pour faire appel à des annotateurs pour que les projets s'arrêtent et que les travailleurs soient licenciés. C'est d'ailleurs pour cette raison que certaines entreprises formelles mieux structurées mettent en place des services de *prototypage* afin de tester les projets potentiels et d'identifier leur faisabilité et leur pérennité. Il

s'agit alors analyser deux questions centrales : comment les travailleurs vivent-ils cette situation et qu'est-ce que cela nous indique sur la qualité du travail d'annotateur ?

9.2.2 Des conditions de travail synonymes d'une intégration "laborieuse" sur le marché du travail

Si l'on se focalise sur les aspects qualitatifs de cette activité, les annotateurs sont très peu à considérer que leur activité les expose à des postures pénibles (20,4%) ou à des petits caractères (24,4%)^{B17}. Par ailleurs, ils sont 76,4% à se sentir autonomes et 59,3% à indiquer pouvoir choisir leur rythme de travail (Table 62). Si cela dépend fortement de l'organisation du travail au sein des entreprises, il s'agit également d'une caractéristique des projets d'annotations, qui, en comparaison avec d'autres activités de services comme les relations clients, peuvent être asynchrones ce qui favorise le développement d'une organisation plus flexible. Comme l'indique Fidy lorsqu'il évoque les points positifs de son expérience chez le sous-traitant B : *"Les heures aussi, c'est plus flexible par rapport aux autres sociétés parce que on peut commencer quand on veut"* (Fidy - Annotateur - Sous-traitant B).

Cette flexibilité est beaucoup plus limitée pour des projets d'IA qui nécessitent des activités en temps réel, comme dans le cas du projet AiView dans lequel certains annotateurs travaillent de nuit et le week-end. Sur ces projets, la pression est également très forte, car les erreurs des annotateurs ont un impact direct sur la surveillance des supermarchés français et par ricochet, sur les travailleurs eux-mêmes, puisqu'ils risquent le licenciement en cas d'erreur^{B18} Table 63. C'est également le cas de certaines entreprises, qui organisent le temps de travail en 3/8 afin d'optimiser l'utilisation de leurs locaux et pour s'aligner sur d'autres fuseaux horaires (comme les USA).

Ainsi, si les annotateurs peuvent se sentir autonomes et flexibles, 45% considèrent être soumis à une forte pression liée à la qualité et au volume des données à annoter Table 62. Le premier point fait référence à la temporalité des flux de données envoyés par les entreprises d'IA : Jedy explique par exemple que *"lorsqu'il y a des projets complexes et tellement difficiles, c'est difficile pour eux parce qu'il y a le stress qui monte effectivement si la deadline n'est pas respectée"* (Jedy - Annotatrice - Sous-traitant G)^{B19}. La qualité des données est considérée par les annotateurs comme largement dépendante de leurs performances au travail - 93% des annotateurs considèrent que leur travail a des conséquences sur la qualité - et constitue une source de

317. Nous renvoyons à nos annexes pour la formulation exacte de la question

318. *"il est important d'avoir une bonne précision, car ces vidéos sont traitées en temps réel (...) Si on rate un vol, ça va avoir un impact important sur l'équipe et si l'agent fait une erreur, ça donne une mauvaise qualité à notre équipe, et ça risque de ne pas avoir des primes (...). Chaque agent doit donner le meilleur lorsqu'ils sont sur la plateforme alerte. mais lorsqu'on fait une erreur, ça donne tout de suite du stress, car c'est notre faute si on donne une mauvaise image de notre équipe(...)"* (Maro - Annotateur - Sous-traitant A)

319. Un autre exemple : *"Mais ici, on n'a pas trop de travail, mais on a une pression de traitement, parce que par exemple les deadlines ne sont plus loin, et on fait tous nos efforts, le maximum d'efforts pour atteindre les deadlines. on a beaucoup de travail ici."* (Félix - Annotateur - Sous-traitant G)

stress [Table 62](#). Comme l'explique Hejo, de l'entreprise A cela : *"crée des stress tout le temps. Si on a plusieurs doutes par jour, on est plus stressé stressé, et en fin de semaine, on sera assez crevé"*. Les annotateurs considèrent également que leur travail a des conséquences sur les coûts (66,3%) [Table 62](#). Ce deuxième aspect propre à la qualité des données est lié à la nécessité de prendre en compte les retours des clients en recommençant leur travail, ce qui constitue à nouveau une source de stress : *"Des fois on attend des retours du client (...), est-ce qu'il est satisfait ou est-ce qu'il est pas satisfait ? Dans ces moments-là on est un petit peu stressés. [trop de travail ?] C'était la semaine dernière, il y a eu des changements sur le projet [des amers maritimes]. Il y a eu un malentendu avec le client, et j'ai été obligé de tout refaire."* (Haga - Annotateur - Sous-traitant G).

Au-delà de ces dynamiques générales concernant le travail d'annotateur, nos données nous permettent de constater une différence entre les entreprises. Malgré une présence équilibrée des différentes strates hiérarchiques au sein de chaque échantillon, on constate une variation au niveau du salaire entre le sous-traitant B d'un côté et le sous-traitant G et A de l'autre. Pourtant, à la question de leur ressenti par rapport à leurs salaires, ce sont les annotateurs du sous-traitant A, les moins bien payés qui ont le plus répondu se sentir *"Bien payé"* [Table 23](#). On retrouve quasiment la même hiérarchie entre les entreprises concernant le fait d'être considéré *"à sa juste valeur"* et *"d'être exploité"*. Comment expliquer ces résultats chez le sous-traitant A, qui est pourtant une entreprise informelle au sein de laquelle les travailleurs ne sont pas salariés ?

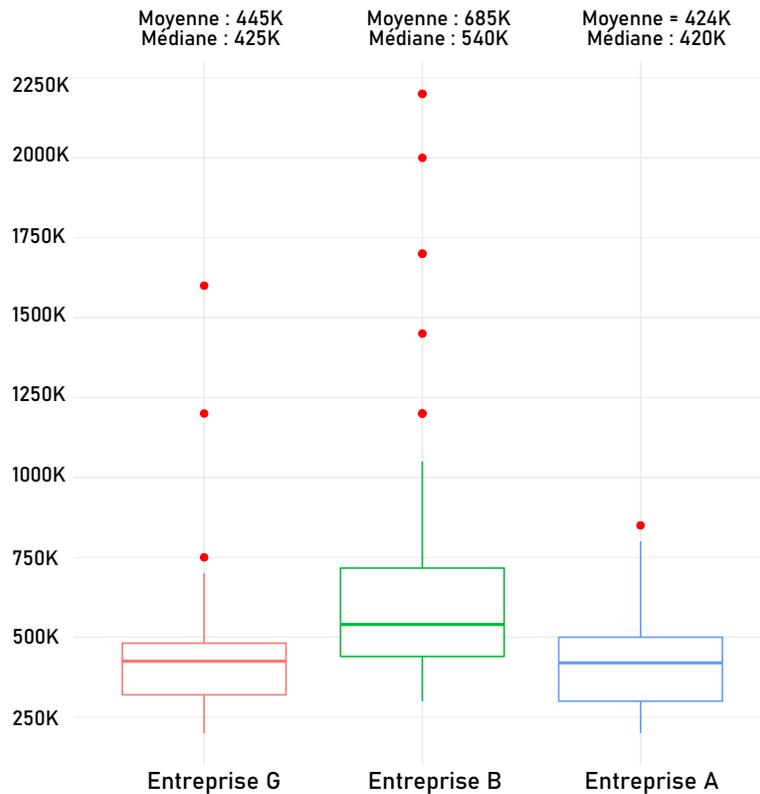


FIGURE 46 – Comparaison des salaires entre les trois entreprises étudiées à travers notre questionnaire. Lecture : 1 euros correspond à 5000 ariary au cours du 15/11/2023

Nos résultats concernant l'âge et l'expérience professionnelle des annotateurs au sein de ces deux entreprises suggèrent deux éléments d'explications. D'abord, comme évoqué précédemment, la jeunesse des annotateurs du sous-traitant A peut expliquer qu'ils acceptent des salaires plus faibles alors que chez les deux autres, les annotateurs plus âgés et ayant plus d'expérience professionnelle considèrent que leur salaire n'est pas suffisant. Toujours au sein du sous-traitant A, il y a en effet bien moins d'annotateurs avec des enfants, ce qui peut justifier cette position vis-à-vis de leur salaire, ceux-ci ayant besoin de moins d'argent pour leur foyer [Table 66](#). Par contraste, cela explique également l'attitude des travailleurs du sous-traitant G qui gagnent un salaire aussi élevé que chez le sous-traitant A, mais sont davantage mariés ou en couple et ont parfois un ou des enfants.

	Pas bien payé	Bien payé	Total
G	68,1	31,9	100
B	55,6	44,4	100
A	35,9	64,1	100
Ensemble	51,2	48,8	100

¹Réponses en pourcentage, Fisher.test = 0,001312

	Pas considéré à sa juste valeur	Considéré à sa juste valeur	Total
G	53,2	46,8	100
B	24,4	75,6	100
A	15,4	84,6	100
Ensemble	27,4	72,6	100

¹Réponses en pourcentage, Fisher.test = 2,062e-12

TABLE 23 – Rapport des travailleurs au salaire

Un autre facteur de différenciation entre les entreprises porte sur les risques de perdre son emploi. Sur ce point, une première variable de notre questionnaire, l'impression d'être "*interchangeable*", illustre cette fois une forte différence entre les deux entreprises formelles de notre échantillon (G et B) et l'entreprise informelle (A). Les travailleurs sont ainsi respectivement 31,9% et 37,8% à se sentir interchangeables contre 62,8%.

	Pas interchangeable	Interchangeable	Total
G	68,1	31,9	100
B	62,2	37,8	100
A	37,2	62,8	100
Ensemble	54,4	45,6	100

¹Réponses en pourcentage, Fisher.test = 0,0005504

TABLE 24 – Interchangeable selon l'entreprise

Nos analyses qualitatives suggèrent que ce résultat peut s'expliquer par l'absence de contrat de travail formel, ainsi que par les risques de licenciement qui pèsent sur les travailleurs du projet AiView. Cet aspect est d'ailleurs confirmé par le nombre important de travailleurs du sous-traitant A à considérer qu'ils ont a minima un "*faible*" risque de perdre leur emploi (62,8% des travailleurs de l'entreprise), par rapport aux deux autres entreprises.

	Inexistante	Faible	Importante	Total
G	61,7	29,8	8,5	100
B	48,4	48,4	3,4	100
A	28,2	62,8	9	100
Ensemble	44	49,5	6,4	100

¹Réponses en pourcentage, Fisher.test = 0,0008977

TABLE 25 – Probabilité de perdre son emploi

Le sous-traitant A est enfin la seule entreprise dirigée par une malgache, et est sans doute impactée négativement par une forme de discrimination envers les entrepreneurs malgaches. C'est ce que peut laisser à penser notre échange avec Jessica, une autre cheffe d'entreprise qui prête ses locaux afin d'interviewer les travailleurs de l'enquête Facebook. Cette dernière nous raconte sa trajectoire professionnelle dans l'industrie de l'externalisation et illustre les difficultés que connaissent les entrepreneurs malgaches. Après plusieurs années chez ComData, un centre d'appels et de saisie de données, elle décide de se lancer en indépendante en cherchant des contrats par le biais de plateformes comme LinkedIn ou Codeur.com. Entre les arnaques au paiement des salaires ou aux frais de connexion à avancer, elle nous raconte ses difficultés à trouver des contrats stables. Elle explique également devoir faire face à l'image négative des travailleurs malgaches auprès de certains intermédiaires. Ces derniers sont en général des chefs d'entreprises tunisiens ou marocains qui récupèrent des contrats auprès de donneurs d'ordres français et les réintermédiat à des entrepreneurs malgaches. Lors de la négociation d'un contrat avec un entrepreneur tunisien, Jessica explique qu'il lui a dit : *"je ne travaille pas avec les Malgaches"*. Ce dernier explique encore qu'une de ses connaissances tunisiennes lui a dit : *«vous n'êtes pas bon pour travailler, vous [ne venez] pas de rendez-vous ... !»*. Jessica comprend alors qu'elle fait face à une discrimination dirigée vers les travailleurs malgaches : *"ils disent, derrière notre dos, que (...) les Malgaches ils ne font pas bien leur travail et cetera, pour que les clients nous acceptent pas (...). Par contre [les clients] ils vont accepter les tunisiens marocains, mais les clients ne savent pas que ce travail, ils vont le sous-traiter avec les prestataires à Madagascar"*.

Par la suite, sa trajectoire professionnelle s'est orientée vers une activité d'achat et de revente de contrats de prestations à des entrepreneurs malgaches, servant alors de garants de confiance entre des entrepreneurs français et malgaches. Cet exemple illustre une discrimination envers les entrepreneurs et les travailleurs malgaches qui se traduit par une moindre capacité à la négociation des prix et des conditions de travail avec les clients. Par ailleurs, sans réseau à l'extérieur du pays, les entrepreneurs malgaches auront du mal à démarcher des clients, ces derniers préférant interagir avec des responsables français. Si l'entrée dans des entreprises franches ne garantit de bonnes conditions de travail, cela se traduit par une formalisation

des contrats, ainsi que par une organisation du travail plus protectrice vis-à-vis des travailleurs. Outre ces entreprises occidentales, le fait de créer une entreprise est rendu difficile par une double inégalité : raciale et financière. Encore, une fois, c'est bien la structure extractive, "raciale" et post-coloniale de ce secteur qui détermine les protections des travailleurs. Ces derniers gagnent en stabilité s'ils peuvent compter sur un entrepreneur français qui, grâce à son personnel, pourra trouver davantage de clients.

	Pas autonome	Autonome	Total
G	19,1	80,9	100
B	17,2	82,8	100
A	39,7	60,3	100
Ensemble	25,7	74,3	100

¹Réponses en pourcentage, Fisher.test = 0,00233

	Ne demande pas de précisions aux clients	Demande précisions aux clients	Total
G	25,5	74,5	100
B	46,2	53,8	100
A	65,4	34,6	100
Ensemble	48,6	51,4	100

¹Réponses en pourcentage, Fisher.test = 5,956e-0,5

TABLE 26 – Autonomie et contacts avec clients

Pour approfondir ces éléments, nous réalisons une régression logistique multinomiale, dans laquelle la variable à expliquer est celle indiquant l'appartenance à telle ou telle entreprise de sous-traitance. Nous reprenons ensuite un certain nombre de variables explicatives liées aux conditions de travail ³²⁰. Les résultats de ce modèle nous permettent de confirmer ce qui différencie les entreprises, mais aussi ce qu'elles partagent ^{Figure 52}. Ainsi l'ennui est ressenti dans toutes les entreprises, du fait de la répétitivité des tâches d'annotation ³²¹. Comme on l'a vu précédemment, le stress constitue également une variable partagée au sein de l'ensemble des entreprises. En revanche, le fait d'être bien payé et pour autant de sentir que son travail est "interchangeable" est davantage relevé par les annotateurs du sous-traitant A, tandis que les annotateurs du sous-traitant G soulignent ne pas être reconnus à leur juste valeur. On constate finalement une différence entre l'entreprise B, la plus grande de notre échantillon et la seule à être en zone franche

320. Sentiment d'être bien payé, sentiment d'être considéré à sa juste valeur, sentiment d'utilité, sentiment d'ennui, sentiment de stress, pénibilité du travail, sentiment d'être interchangeable, probabilité d'être promu, probabilité de perdre son emploi et celle de retrouver un emploi équivalent

321. Manityfy (Sous-traitant B) explique par exemple que "parfois c'est très fatigant, il faut se l'avouer ... moi je fais du contrôle qualité, j'en ai vraiment marre de le faire parfois parce que c'est vraiment répétitif et ça fait mal aux yeux"

avec la G, une petite entreprise privée formelle et la A, une petite entreprise informelle détenue par une dirigeante malgache.

Ces résultats suggèrent que le statut d'entreprise franche peut être associé à de meilleures conditions de travail que le secteur privé formel et que le secteur informel, ce qui confirme les résultats de Razafindrakoto et Roubaud (1997) à propos des zones franches. Néanmoins, nous pensons que cela illustre également les différentes formes d'intégration des travailleurs dans le marché du travail ce qui se traduit par plusieurs types de précarité professionnelle.

Pour analyser cette dimension de l'annotation, nous nous basons sur les recherches de Paugam (2007) sur le développement des formes précaires d'emplois dans lesquelles il présente *“quatre formes d'intégration professionnelle : l'intégration assurée (stabilité de l'emploi et satisfaction au travail), l'intégration incertaine (instabilité de l'emploi et satisfaction au travail), l'intégration laborieuse (stabilité de l'emploi et insatisfaction au travail) et, enfin, l'intégration disqualifiante (instabilité de l'emploi et insatisfaction au travail)”* (Paugam & Vendramin, 2020). Pour ce faire, nous utilisons une méthode similaire à la sous-section 9.1.1, afin de réaliser une typologie des annotateurs selon le rapport à leur travail.

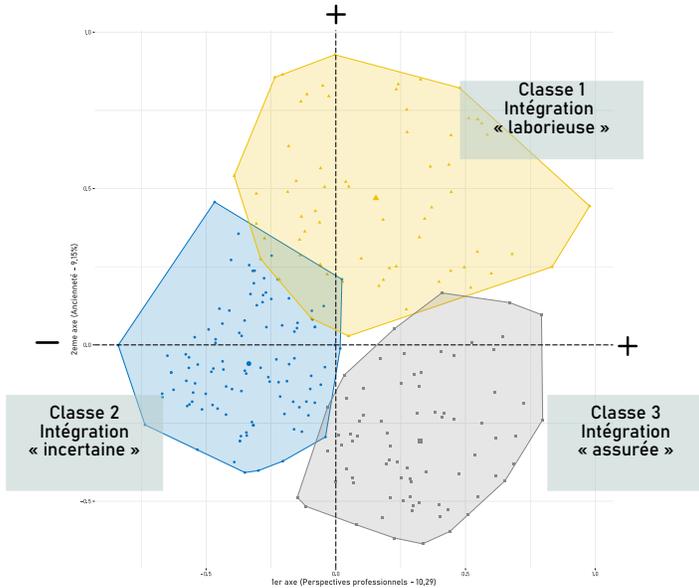
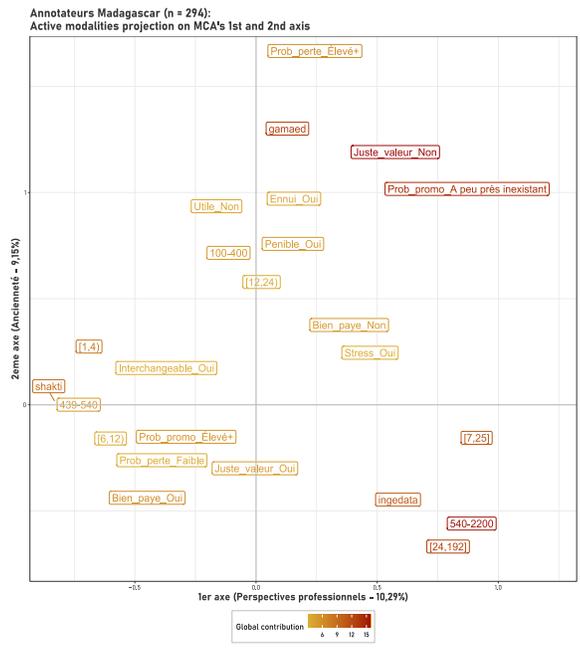


FIGURE 47 – Trois formes d'intégration des annotateurs sur le marché du travail

Le résultat ainsi obtenu confirme les différences entre nos trois entreprises et permet d'identifier trois classes de travailleurs. La première classe est composée de travailleurs de l'entreprise G qui se caractérisent par une ancienneté moyenne sur le marché du travail, davantage d'impression de pénibilité et par le sentiment de ne pas être reconnus à leur juste valeur. Une deuxième classe correspond aux travailleurs de chez le sous-traitant A qui sont plus inexpérimentés et qui comprend le plus d'annotateurs indiquant avoir une probabilité faible ou élevée de perdre leur emploi. Il se considèrent également davantage interchangeables. Enfin, une dernière classe est composée de travailleurs du sous-traitant B qui ont davantage d'ancienneté sur le marché du travail, sont mieux payés, et qui sont davantage satisfaits de leurs conditions de travail.

Entreprise	A peu près inexistant	Probabilité élevé ou faible	Total ¹
G	61.7	38.3	100
B	48.4	51.6	100
A	29.4	70.6	100
Ensemble	44.5	55.5	100

¹Réponses en pourcentage, Fisher.test = 0.000737

TABLE 27 – Probabilité de perdre son emploi

L'analyse des conditions de travail illustre les différences entre les types d'entreprises présentes à Madagascar. Au-delà d'une variation dans les conditions de travail, il s'agit également d'une différence dans les populations impliquées dans chacune de ces entreprises, ce qui correspond également à différents modes d'intégration dans le marché du travail. Dans le cas des travailleurs du sous-traitant A, elle est *incertaine*, car ils sont confrontés à un fort risque de perdre leur emploi à cause de leur contrat de travail informel. Les travailleurs du sous-traitant G se considèrent plus stables, mais sont moins satisfaits de leur travail. Enfin, les travailleurs de l'entreprise B se sentent stables et sont relativement satisfaits de leurs conditions de travail ce qu'on peut expliquer par des salaires plus élevés. Le rapport au salaire est également très différent selon les entreprises. Chez l'entreprise A et B, les annotateurs sont davantage à considérer bien payés que chez l'entreprise G où les annotateurs considèrent qu'ils ne sont pas considérés à leur juste valeur. Comment expliquer ce résultat alors que les salaires des sous-traitants A et B sont pourtant éloignés? D'autres facteurs liés aux conditions de vie des annotateurs constituent des pistes de réflexion centrées cette fois sur la précarité économique des annotateurs.

* * *

Dans ce chapitre, l'étude objective des conditions d'emploi dans le secteur de l'annotation de données

révèle un panorama diversifié et pointe un constat essentiel : même au sein des entreprises formelles, les travailleurs demeurent confrontés à la précarité de l'emploi. Les entreprises de sous-traitance tirent ainsi parti de la variété des statuts de l'emploi à Madagascar pour réduire les risques associés à la recherche de donneurs d'ordres stables. Cette observation est cruciale pour comprendre la réalité des conditions de travail dans l'annotation de données.

Nous avons montré que l'application du Code du travail constitue une forme de protection dans les entreprises formelles, dont ne bénéficient pas les travailleurs informels. Cependant, cette protection sociale est fortement limitée par l'absence de droit au chômage et par les possibilités de licenciement économique. Ces résultats apportent donc une nuance significative au postulat d'une stabilisation financière par l'intégration dans une entreprise formelle. En effet, la simple détection d'un problème de qualité dans le travail des données ou le manque de fonds peut entraîner l'arrêt d'un projet et le licenciement des travailleurs. Pour faire face à cette instabilité, certaines entreprises mieux structurées mettent d'ailleurs en place des services de prototypage des projets qui servent à en évaluer la faisabilité et la pérennité.

Cette dimension inégale du secteur renforce encore davantage le pouvoir des entrepreneurs de la sous-traitance. Nos données mettent en lumière une discrimination envers les entrepreneurs et les travailleurs malgaches qui réduit leur capacité à négocier leurs conditions de travail ou les tarifs éventuels d'une prestation de service comme entrepreneur. Finalement, le travail des données est largement basé sur la capacité d'entrepreneurs blancs et/ou de français originaires de l'élite malgache à utiliser leur réseau personnel extérieur à la Grande Île, pour trouver des clients potentiels. Nous avons pu nous rendre compte de cette dimension des relations avec les travailleurs, car plusieurs d'entre eux nous demandent régulièrement de créer une entreprise pour leur proposer du travail^{B22}. La création d'entreprises locales est alors entravée par une double inégalité, à la fois raciale et financière, soulignant l'impact persistant de la structure raciale et post-coloniale du secteur sur les travailleurs.

Enfin, l'analyse des conditions de travail met en évidence de fortes variations selon la typologie d'entreprise. Les travailleurs vivent des réalités variées selon le degré d'incertitude ou de stabilité d'une entreprise ou d'un projet. En fonction du caractère informel de leurs contrats, de leur satisfaction au travail, des salaires et du sentiment d'être valorisés, ces derniers vivent une variété d'intégration dans le marché du travail. Ces éléments tendent à montrer que l'appartenance à une plus grande entreprise formelle constitue une meilleure protection pour l'avenir. Néanmoins, les facteurs liés aux conditions de vie des annotateurs, en particulier leur précarité économique, représentent des axes de réflexion cruciaux pour la suite de cette recherche, notamment en ce qui concerne leur rapport aux salaires.

322. Un travailleur nous dira même directement qu'il a "besoin d'un blanc" pour monter une entreprise

9.3 Des perspectives personnelles et professionnelles contrariées par la structure extractive du secteur

Le manque de sécurité lié à son travail constitue une des caractéristiques de la précarité (Banki, 2013) et une des aspirations des classes moyennes des pays du Sud global (Banerjee & Duflo, 2008). Au-delà d'une sécurité obtenue par les conditions objectives de travail (protection sociale), un autre aspect concerne les perspectives personnelles et professionnelles de l'emploi d'annotateur. L'objectif de cette dernière sous-section est d'analyser ce manque de sécurité au regard des conditions de vie des travailleurs des données. Ce faisant, nous identifions un certain rapport au salaire qui souligne la non-réciprocité des entreprises françaises en termes de niveau de rémunération. Nous montrons que ce salaire ne constitue pas un salaire d'émancipation économique, puisqu'il ne permet aux travailleurs d'assurer leur avenir personnel et professionnel. Les "emplois sans avenir" ("dead-end jobs") se caractérisent par un manque de perspectives d'avancement professionnel, de mobilité ascendante ou d'enrichissement de compétences pour les travailleurs. En d'autres termes, les personnes qui exercent ces emplois ont souvent l'impression que de stagner, se sentant piégées dans une situation professionnelle sans pouvoir atteindre des postes plus élevés ou améliorer leur statut économique. Les travailleurs occupant ces postes peuvent être coincés dans un cycle où ils acceptent des emplois précaires afin de subvenir à leurs besoins immédiats ce qui les empêche d'avoir le temps de se former ou d'acquérir les compétences nécessaires pour progresser dans leur carrière.

Ces emplois se trouvent généralement dans des secteurs ou des industries caractérisés par une forte précarité, une faible rémunération, des conditions de travail difficiles, et un manque d'opportunités de formation ou de développement de compétences. Dans une recherche basée sur des données allemandes, (Mosthaf, Schnabel, et Stephani, 2011) montrent l'importance des caractéristiques de l'entreprise dans le manque de mobilité professionnelle ascendante. Ainsi, les petites entreprises et celles ayant une proportion élevée de travailleurs à bas salaires semblent souvent être des impasses, tout autant que les entreprises qui nécessitent des postes non qualifiés et qualifiés du secteur des services.

Au-delà des caractéristiques inhérentes au secteur, (Mosthaf et al., 2011) montrent également l'importance des facteurs individuels dans le manque de développement professionnel, en identifiant de fortes différences entre les travailleurs à faibles revenus plus jeunes et mieux qualifiés et les femmes qui rencontrent davantage de difficultés pour obtenir une progression de carrière. Si les stratégies de formation externe à l'entreprise (formation professionnelle, poursuite d'études) peuvent s'avérer payantes (Lynch, 1993), faut-il encore que les travailleurs aient les moyens financiers nécessaires à ces formations, ainsi que suffisamment de temps.

Dans le contexte du secteur des services externalisés des pays du Sud global, des préoccupations majeures ont été soulevées concernant la qualité de l'emploi. Elles mettent en lumière une contradiction apparente où des travailleurs hautement éduqués se retrouvent souvent à exercer des emplois peu qualifiés. À partir

de données empiriques recueillies aux Philippines, Beerepoot et Hendriks (2013) examinent les perspectives professionnelles des travailleurs de services. L'un des principaux arguments avancés dans l'article est la nécessité de déplacer la perspective d'une vision centrée sur l'employeur vers une perspective axée sur les employés. Ce faisant, l'accent est mis sur la compréhension de la manière dont l'emploi dans ce secteur contribue à l'employabilité à long terme et à l'acquisition de compétences génériques pour les travailleurs des pays du Sud global. Cet article met en évidence que ce type de travail peut faire partie d'un plan de carrière en permettant aux travailleurs de renforcer leur employabilité.

En ce qui concerne les perspectives à l'intérieur du secteur des services, Buscatto (2002b) propose constate la "dérégulation de la démographie professionnelle" (Paradeise, 1984). Dans le cas du centre d'appels qu'il étudie, il constate que : "La création de nouveaux postes de chefs de groupe est sévèrement ralentie avec l'arrêt d'expansion" du site qu'il étudie. À partir d'une enquête menée auprès de plus de 2400 centres d'appels dans 16 pays, l'étude de Mañé Vernet et Benner (2009) montre qu'en moyenne, moins de 10 % des agents de centre d'appels sont promus chaque année, dont 5,7 % au sein même du centre d'appels et 4 % dans d'autres entreprises. Celles qui ont des processus de travail plus complexes qui exigent des agents une connaissance spécifique de l'entreprise offrent généralement davantage d'opportunités de promotion. Elles constatent cependant que l'autonomie au travail peut parfois servir de « substitut » aux opportunités d'avancement, et que la syndicalisation est associée à moins d'opportunités d'avancement au sein des centres d'appels, mais à davantage d'opportunités dans d'autres secteurs de l'entreprise. Qu'en est-il au sein des entreprises d'annotation? Nous analysons les perspectives personnelles et professionnelles des annotateurs de données au regard de deux aspects : le rapport des annotateurs à leur salaire ainsi que ce qu'ils envisagent pour leur futur professionnel. Nous verrons qu'il s'agit d'un revenu de subsistance, qui ne leur permet pas nécessairement d'envisager un futur professionnel hors du secteur de l'annotation. Même si leur niveau de qualification les autorise à penser à d'autres activités, qu'ils aimeraient exercer en tant qu'indépendants, le manque de capital, ainsi que la discrimination à laquelle ils font face les limitent dans ces ambitions.

9.3.1 Un revenu de survie en décalage avec la valeur générée par l'industrie de l'IA : la "race" comme variable d'ajustement des marges dégagées par le secteur de l'IA

La littérature a souligné l'importance des activités liées au travail des données dans les pays du Sud global pour la subsistance des foyers (Berg et al., 2018). Dans un autre contexte géographique et au sein d'une industrie structurée différemment des plateformes de micro-travail, il paraît pertinent de se demander ce qu'il en est pour ces travailleurs de Madagascar. Combien perçoivent-ils? Et surtout, que représente ce revenu pour les travailleurs, notamment au regard du contexte de la capitale malgache? Ces questions doivent également être mises en relation avec la production de valeur au sein de la chaîne. Ces questions, classiques dans l'analyse des chaînes de valeur, sont particulièrement difficiles à traiter du fait du manque

de données disponibles sur la structure des coûts des entreprises concernées. Il nous semble pourtant qu'il s'agit d'un facteur important au regard des impacts des chaînes de valeur sur le développement et de la structure du pouvoir au sein du réseau de sous-traitance. Sur ce point, nous faisons appel à des données inédites concernant la structure des coûts de plusieurs entreprises d'annotations de données.

En ce qui concerne le salaire, la plupart des travailleurs gagnent entre 93 euros et 126 euros, mais une partie de ce salaire est dépendante de leurs performances et est versée sous forme de primes (plus ou moins 25% du salaire total). Par ailleurs, ce salaire intègre également des indemnités de transport qui sont obligatoires et correspondent à environ 77000 ariary, sauf si l'employeur met à disposition une navette. Le salaire intègre aussi des indemnités de repas à hauteur de 77000 ariary, sauf en cas de mise à disposition d'un service de cantine ³²³

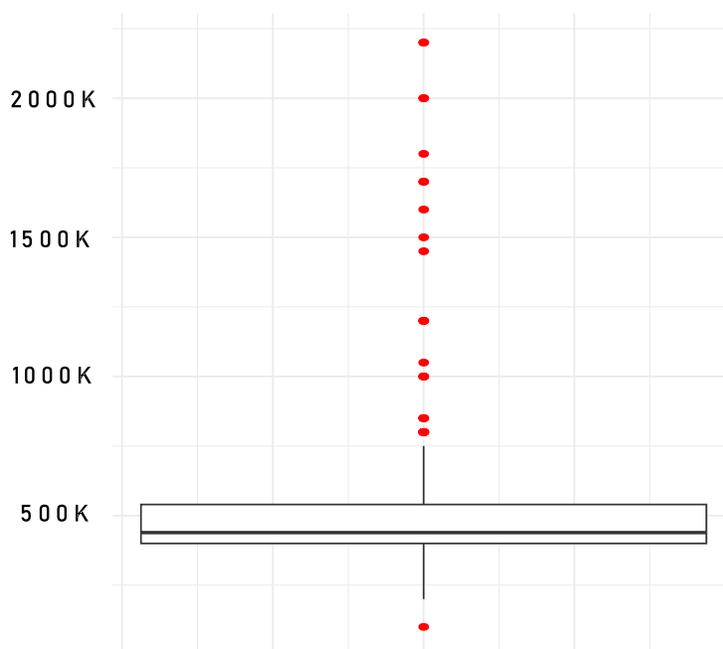


FIGURE 48 – Distribution des salaires des annotateurs de données

Ce niveau de salaire dépend bien évidemment des entreprises, ainsi que de la position occupée dans l'entreprise. Globalement, ces résultats limitent le constat d'entreprises formelles plus protectrices, étant donné que des entreprises formelles comme les sous-traitants C et E, ne sont pas nécessairement les entreprises les plus rémunératrices pour l'annotation de données. Ces entreprises se démarquent d'ailleurs par leur intégration dans des réseaux de production liés à des plateformes numériques de travail comme Mturk

³²³ Oui il y a des indemnités repas et transports qui vont directement avec le salaire. Donc dans la réalité, elle dépense 80K en transport et 80K en repas. (Romane - Annotatrice - Sous-traitant G)

sur lesquels ils proposent leurs services d'annotation de données. Nos résultats suggèrent que ces entreprises façonnent leur service et leurs conditions de travail en fonction des prix pratiqués par les plateformes de travail. Nos estimations de leur rémunération (de 0,41\$ à 1,7\$ par heure) sont nettement inférieures à celles d'Amazon Mechanical Turk, estimées à 3,13\$ par heure en moyenne (Hara et al., 2018). Nos entretiens montrent que les entreprises passant par des plateformes facturent aux alentours de quatre euros de l'heure. Ce tarif est similaire à l'entreprise A, entreprise informelle qui n'a pas recours aux plateformes. Des trois entreprises étudiées systématiquement, l'entreprise B est la seule à ne pas passer par des plateformes numériques de travail. Cela suggère que la différence de revenus entre le travail de plateformes et les revenus des travailleurs malgaches est largement liée à la captation de la valeur par les entreprises de sous-traitance qui gèrent la production, ainsi que par l'intégration ou non dans des réseaux de production plateformes qui ajoutent un intermédiaire supplémentaire qui capte alors une partie de la valeur.

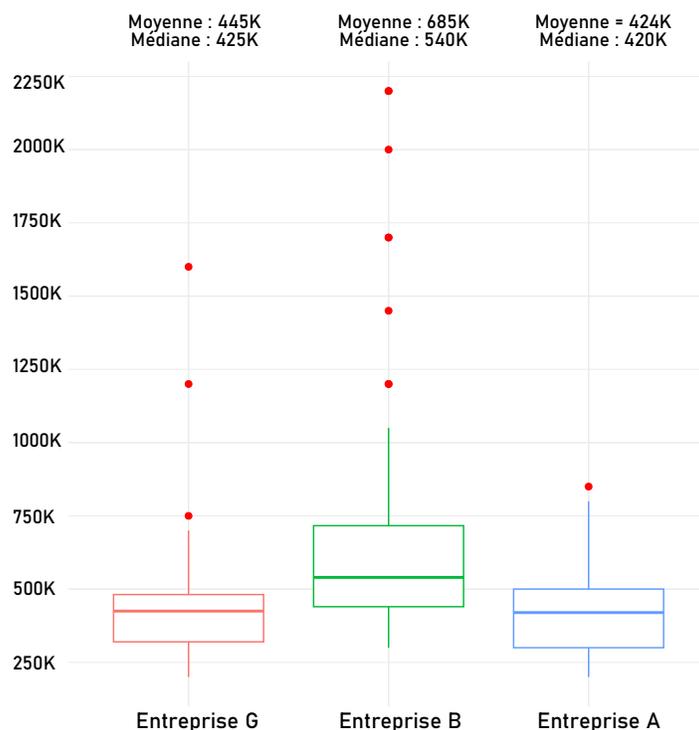


FIGURE 49 – Comparaison des salaires entre les trois entreprises étudiées à travers notre questionnaire

En ce qui concerne la position occupée au sein de l'entreprise, nos résultats suggèrent des différences non négligeables entre les managers et les non managers. Toutefois, la différence entre le salaire moyen et le

salaire médian suggère une variation bien plus importante lorsque les travailleurs atteignent un deuxième niveau hiérarchique (responsable de production) en comparaison avec les autres échelons hiérarchiques (chef d'équipe et annotateurs simples). Il convient de également tenir compte de la mise en place par certaines entreprises de niveaux intermédiaires qui ne sont pas liés à une fonction managériale, comme les agents experts. Dans une moindre mesure, nous avons également pu constater l'existence d'une grille de salaire progressive permettant aux travailleurs de gagner des niveaux de rémunération selon leurs compétences (50000 ariary par niveau). Ces politiques localisées expliquent qu'une partie des annotateurs arrivent à atteindre les niveaux de rémunération de certains managers intermédiaires.

	Manager (N=62)	Non manager (N=196)
Salaire moyen	672909 (139 euros)	482743 (99,96 euros)
Salaire médian	520000 (107,67 euros)	422500 (87,49 euros)
1er quartile	420000 (86,97 euros)	367500 (76,1 euros)
3ème quartile	800000 (165,65 euros)	500000 (103,53 euros)
Manquant	12	7

TABLE 28 – Salaire des annotateurs en fonction de leur position hiérarchique

Essayons à présent d'analyser ces niveaux de salaire en fonction de la situation de Madagascar. Il convient tout d'abord souligner que les derniers chiffres disponibles sont tirés de l'enquête emploi réalisée en 2012 (INSTAT) et qu'ils ne permettent pas de tirer des conclusions pertinentes. Tout juste apprend-on que le secteur formel privé offre des salaires de 200000 ariary en moyenne, soit 41,12 euros au cours actuel. Dans tous les cas, les travailleurs doivent désormais recevoir un minimum de 50 euros par mois, depuis l'augmentation du salaire minimum qui ne s'applique qu'au secteur formel. Il convient toutefois replacer ce niveau de salaire dans le contexte malgache. En effet, un rapport de la banque mondiale souligne que : "Après avoir atteint un plancher de 3,9 pour cent mi-2020, l'inflation a suivi une tendance à la hausse, atteignant (...) 8,8% en moyenne en 2022 contre 5,8% en 2021", surtout, les "dans le secteur alimentaire du fait des conditions du marché domestique".

Pour mieux comprendre le niveau de salaire des annotateurs, il est également possible de s'appuyer sur les données d'autres secteurs économiques. Le Rapport de la Banque mondiale (2020) sur le secteur textile nous semble une bonne base de comparaison³²⁴. Selon le niveau hiérarchique et d'après des chiffres de 2018, les niveaux de salaires se décomposent de la manière suivante : pour les personnels d'encadrement, le salaire est supérieur à 280 euros par mois. Concernant, les techniciens, le salaire se situe entre 141 et 231 euros

324. Le secteur textile est considéré comme équivalent au secteur des services, notamment du fait du fort recours aux zones franches dans les deux secteurs

par mois. Enfin, pour le personnel "simple" qualifié, le salaire se situe entre 74 et 125 euros par mois contre entre 56 et 107 euros par mois pour les salariés non qualifiés.

Nos résultats montrent que les salaires moyens sont relativement similaires à ce qui est observé dans le secteur textile pour les "techniciens" et les travailleurs simples "qualifiés" et "non qualifiés". En revanche, on note une différence importante en ce qui concerne les postes de managers. Par rapport au secteur textile, une piste d'analyse de ce résultat est de considérer que les annotateurs ont moins de perspectives d'évolutions vers des emplois d'encadrement plus élevés, car ces emplois sont davantage occupés par des travailleurs français expatriés, du fait de la nature cognitive du travail et de la nécessité de communiquer avec ces clients français. Cependant, les zones franches textiles qui sont évoquées dans le rapport de la Banque Mondiale se situent à Tulear, Moramanga et Diego Suarez alors que les travailleurs étudiés se situent pour la quasi totalité à Antananarivo, la capitale du pays, où le coût de la vie est plus élevé.

Il convient enfin préciser que les salaires perçus par les annotateurs constituent une partie substantielle des revenus totaux de leur foyer (60% en moyenne). Toutefois, cette moyenne cache d'importantes disparités entre les annotateurs. Ainsi, le poids du salaire représente 100% pour les annotateurs qui habitent seuls, et cette proportion est bien plus restreinte pour les foyers composés d'une famille nucléaire (entre 2 et 5 personnes) avec 65% et l'est encore moins pour les foyers élargis (5 à 29 personnes) pour qui cela représente 50%. La part importante du salaire dans les revenus du foyer suggère qu'il s'agit réellement d'un salaire de subsistance d'autant plus lorsque les annotateurs ont une famille. Comme l'explique Joda, annotateur dans l'entreprise A : *"Le salaire est plutôt bon. On arrive à joindre les deux bouts. Mais faire des économies pour une personne qui est mariée et a des enfants, c'est difficile. On arrive à payer les factures, notre nourriture et la scolarité des enfants, mais c'est très serré comme budget pour une personne mariée. Mais pour une personne [seule], c'est très bien. On arrive à faire beaucoup de marge."* Nos données quantitatives confirment cet extrait d'entretien en montrant que le sentiment d'être bien payé est très largement dépendant de la taille du foyer. Finalement, là où le métier d'annotateur peut permettre à des travailleurs vivant seuls de "joindre les deux bouts", ce n'est pas nécessairement le cas pour les foyers plus nombreux.

Taille foyer	Ne se sent pas bien payé	Se sent bien payé	Total ¹
[1,2)	38,5	61,5	100
[2, 5)	49,3	50,7	100
[5,29]	64,9	35,1	100
Ensemble	54,5	45,5	100

¹Réponses en pourcentage, Fisher.test = 0.03

TABLE 29 – Sentiment d'être bien payé selon la taille du foyer

Au travers de ces premiers éléments d'analyses, on constate qu'à l'exception des annotateurs qui habitent seuls, les travailleurs ne considèrent pas que leurs revenus comme suffisants. Plusieurs aspects liés à leur contexte de vie expliquent ce résultat. Tout d'abord, plusieurs annotateurs soulignent en entretien subir la pression de l'inflation, que ce soit pour le prix de la nourriture ou pour le logement, comme Michael, annotateur qui explique qu'à *"Tana c'est un peu difficile (...). Je dis pas que je suis dans la galère. J'essaye de vivre avec, parce que c'est trop cher ici. La location la moins chère ici, c'est 200 000 ariary. Donc plus le déplacement, la nourriture à la maison, l'éducation des enfants, c'est compliqué."* Cette première dimension du rapport des annotateurs à leur travail s'explique également par des augmentations faibles, voire inexistantes, au fil de leurs carrières. Ainsi, dans l'encadré qui suit, Mevo soulève la problématique de l'augmentation des produits de première nécessité, au regard de l'absence d'augmentation de salaire.

Encadré n°75 : Des produits de première nécessité qui augmentent, des salaires qui stagnent

M : tu es recruté, tu as eu ton bac, tu as quelques compétences en informatique et en bureautique, tu décides de travailler (...) de ne plus étudier, tu décides de stagner [dans ta carrière] tu te poses la question « ça va durer combien de temps ? », ça peut durer jusqu'à 20 ans ! Il y a l'essence, il y a les guerres, la PPN [Produits de première nécessité] (...) Et tu vas réclamer ça aux dirigeants alors que toi tu as accepté au départ, c'est toi qui as accepté [ce salaire] en ne voulant pas évoluer X : oui mais du coup les dirigeants ils pourraient tenir compte de l'inflation par exemple ?

M : oui ils vont tenir compte de l'inflation, mais imagine un peu le truc, tu vas donner 100000a de plus pour 100 personnes sur 12 mois ? imagine un peu tes chiffres et comment tu vas expliquer ça à un fonds d'investissement qui a investi sur une entreprise qui devrait être rentable à 30% minimum (...)

Mevo - Manager - Sous-traitant C

La gestion du budget du foyer est d'autant plus complexe avec la présence d'un enfant. Fidy nous explique qu'elle n'arrive *"pas vraiment"* à vivre dans la capitale malgache. Dans son cas, entre la location de *"son appartement"*, *"les frais de déplacement"* ainsi que *"les rations [alimentaires] mensuelles"*, elle considère que *"c'est un peu serré."*, et ce d'autant plus qu'elle est *"mère célibataire"*, et doit donc s'occuper de son enfant, par exemple *"payer son école"*. Son cas de figure illustre une autre problématique liée aux frais nécessaires pour pouvoir exercer un travail à temps plein : frais de déplacement, frais de nourriture et même frais de garde pour les enfants. Dans le cas de Ranto et de sa femme, parents d'un enfant et tous les deux annotateurs dans l'entreprise B, ces postes de dépenses atteignent jusqu'à 40% de leurs revenus totaux mensuels. Ainsi, sur environ 1 million d'ariary (208 euros) qu'ils gagnent à deux, ils dépensent jusqu'à 60000 ariary de transport pour une personne, 200000 ariary de frais de nourriture et 150000 ariary de frais de garde par mois. Sur ce

dernier point, Ranto précise également que la "nounou" est en fait une tante de sa femme ce qui suggère également l'importance du travail *reproductif* dans le bon fonctionnement du secteur.

Encadré n°76 : La précarité des annotateurs mariés avec enfants

X : avec ça, comment on vit à Tana ? M : c'est difficile. Parce que en ce moment (...), il y a en fait une crise, la montée des PPN ... il y a aussi une annonce qu'ils vont augmenter les transports. Et [moi] en fait, j'ai un enfant à nourrir ... ma femme travaille aussi chez [ce sous-traitant]

X : on l'a eu en entretien ?

M : non. Mais même à nous deux, pour un mois, c'est juste pour notre subsistance, il n'y a pas d'épargne ... on n'a pas d'argent et de quoi économiser. C'est juste un tout petit peu économisé, mais ça part pour les transports, la nourriture, les courses de l'enfant ... C'est assez compliqué en fait. Pour ma femme, 60 000a par mois pour les transports, parce que elle, elle n'arrive pas à marcher à pied (...). X : la nourriture c'est combien le midi ?

M : ... je ne sais pas trop. Je dirais 5000a par jour

Y : j'imagine que vous devez prévoir la nounou pour l'enfant quand vous travaillez ?

M : oui. Le salaire de la nounou (...) ce qu'on s'était fixé c'est à peu près 150k par mois ...

Ranto - Annotateur - Sous-traitant B

Ces éléments portant sur le rapport des annotateurs à leur salaire rejoignent un autre constat de notre enquête sur la part importante des travailleurs qui occupent une double activité. Pour certains, cette deuxième activité est d'ailleurs liée au numérique. Ainsi, une partie de notre questionnaire examine le travail des plateformes à Madagascar. Si le pays est très bien représenté en termes d'exportation de services informatiques, les enquêtes cherchant à localiser les travailleurs de services platformisés n'ont pas mis en évidence une présence importante de travailleurs venant de Madagascar. Pourtant, notre questionnaire montre que 40% des travailleurs interrogés travaillent via des plateformes. Cette question était suivie par des questions portant sur l'usage spécifique de plateformes de micro-travail (comme Appen) ou de *freelancing* (comme Upwork), ces questions n'ont reçu quasiment aucune réponse.

Nos entretiens permettent de mieux expliquer ce résultat. Ainsi, plusieurs annotateurs, comme Navalona et Fenosao (Sous-traitant B) racontent avoir essayé des tâches d'annotations sur des sites comme Remotask, une plateforme américaine spécialisée dans la production de données pour l'IA. Navalona explique avoir testé Remotask en intégrant un groupe de discussion créé par un autre travailleur de son entreprise portant sur l'utilisation du site. Il souligne néanmoins les difficultés à occuper cette deuxième activité, très coûteuse en temps et en argent et finalement peu rémunératrice : *"pour moi le problème [est] matériel, alors je n'ai pas pu faire trop longtemps ce job, mais par contre il m'a bénéficié un peu (...)* Par contre je suis très fatigué,

après ici, c'est là-bas" (Nahary - Annotateur - Sous-traitant G). En effet, en plus de sa journée de travail il doit dédier jusqu'à "six heures" par jour à cette activité qui lui rapporte environ 250000 ariary en un mois et demi, sans compter l'argent qu'il doit reverser au propriétaire du cyber-café qui l'héberge pour cette activité, car il ne possède pas son propre matériel ni de connexion Internet fiable. Cela illustre les freins liés au coût des données Internet dans l'utilisation de certaines plateformes d'annotation qui nécessitent d'afficher des images parfois très lourdes³²⁵. Fidy souligne quant à lui l'investissement nécessaire pour pouvoir travailler sur Remotask : "y a trop de formations et les tâches qu'ils ont attribuées demandent que tu finisses cette formation et des fois ça se bloque sur quelques exemples, ou des ... il y a des exercices avant que tu te mettes à travailler."

Dès lors, comment analyser ce résultat sur l'usage des plateformes? Nous pouvons proposer quelques explications. Premièrement, lors d'un entretien exploratoire, Éric nous a indiqué que le recrutement des travailleurs de l'*outsourcing* se faisait souvent via des groupes d'annonces Facebook, ou via des sites de recrutement similaires à Indeed. Par conséquent, "travailler via des plateformes" pourrait être compris comme le fait d'avoir trouvé un travail ou des tâches courtes (quelques semaines) via Facebook. Par ailleurs, on peut aussi imaginer ce résultat comme la conséquence d'une "continuum de l'informalité" propre à des pays où le travail informel est majoritaire, et où les plateformes numériques de travail commencent à peine à se développer. Dans ce contexte, de nombreuses activités liées aux services numériques sont proposées avec des contrats de *freelancing*, ou même tout simplement sans contrat via des groupes Facebook. D'autres encore, utilisent la plateforme pour développer leur propre entreprise informelle. C'est par exemple le cas pour Nity qui est photographe indépendant en plus de son travail d'annotateur chez le sous-traitant A et qui exerce pour des cérémonies (mariages...). Il "fait sa pub" par le biais de Facebook. C'est également le cas pour Flore, annotatrice dans l'entreprise B et qui nous explique qu'elle vend "des fromages en ligne" sur Facebook. Cette tendance à la création d'activité informelle passant par l'utilisation de groupes Facebook ressemble fortement à ce qui est identifié par Friederici, Wahome, et Graham (2020) à propos du développement de business liés au numérique sur le continent africain.

Plus généralement, nous avons constaté que de nombreux annotateurs occupent deux voire trois activités en même temps. Le cas de Mevo illustre la nécessité pour les annotateurs de diversifier leurs sources de revenus. Il met ainsi en lumière plusieurs aspects importants concernant la question du salaire et de la stabilité financière au sein des entreprises de sous-traitance. Dans l'extrait d'entretien qui suit, il suggère que les salaires chez le sous-traitant C sont insuffisants pour couvrir les besoins de base des employés, ce qui explique la nécessité pour les agents, y compris les cadres d'avoir recours à la multiple activité. Cela souligne finalement les défis financiers auxquels sont confrontés les employés de l'entreprise et met en

325. Un article de Graham (2015) montre d'ailleurs que Madagascar fait partie des rares pays où le coût de la souscription à Internet est supérieur au revenu moyen.

évidence la nécessité pour eux de diversifier leurs sources de revenus pour "survivre".

Encadré n°77 : La double activité généralisée des annotateurs de données

M : oui. Toi si tu vas poser cette question-là à un agent chez Sous-traitant C, «est-ce que ton salaire te suffit ? », ça ne va jamais suffire. [Ici] on est tous en double emploi, même les cadres. Donc tu as un truc que tu fais en parallèle, et qui fait que tu vas aller travailler [ici] (...) à chaque fin de mois ça c'est ton revenu fixe, et quitte à toi de faire progresser tes petits business et cetera [sur Facebook]

X : parce que tout le monde vend des trucs ?

M : tout le monde fait des trucs, c'est ça.

Mevo - Responsable - Sous-traitant C

Dans certains cas cette double activité est complétée par un investissement des annotateurs dans leur formation qui est à la fois financier et temporel. C'est le cas de Bodo, actuellement en master 2 et qui doit réaliser une triple journée. Ainsi, en plus de travailler comme annotateur la nuit dans l'entreprise A, il est étudiant la journée, tout en donnant des cours quelques heures par jour. Investissement dans la formation donc, mais également la nécessité de garantir la survie de son foyer : en tant qu'orphelin, il doit en effet s'occuper de sa soeur. Comme dans le cas des micro-travailleuses françaises que nous avons étudiées dans un article (Tubaro et al., 2022), la précarité des travailleurs de Madagascar les pousse à multiplier les activités rémunérées (leur emploi) et non rémunérées (le travail reproductif).

Pour d'autres annotateurs, cette nécessité d'augmenter leur revenu correspond à un véritable "travail d'à-côté" (Weber, 2009), à l'intersection de l'activité de loisirs qui donne du sens à l'activité principale et du travail marchand pur et simple permettant de gagner plus d'argent. C'est par exemple le cas de Nahary, annotateur chez le sous-traitant B qui, en plus de son temps plein, aide "des amis qui tiennent [un cybercafé]" la semaine et propose des prestations de traiteur le week-end. Il nous indique par ailleurs que "les autres [annotateurs] font de la ferme quelque part, les autres font de la pâtisserie comme ça ... il y a beaucoup, ils font des épiceries ..." Comme pour Nity (sous-traitant A), ou pour Jessica qui nous explique avoir une activité de vente de pâtisserie en plus de son travail d'annotatrice, cette troisième activité est une manière de pratiquer une activité professionnelle par "passion", autant qu'une nécessité induite par la faiblesse du salaire.

Un facteur important tempère néanmoins l'ambiguïté entre le travail et les loisirs et concerne la place de l'activité "d'à-côté" dans les "espoirs professionnels" des travailleurs. Ainsi, autant qu'une manière de diversifier des revenus perçus comme plus stables, mais qui peuvent cependant se tarir soudainement du fait de la structure du secteur, ces activités sont également un moyen pour les travailleurs de s'accomplir professionnellement en vue d'éventuellement exercer cette deuxième activité à temps plein. Jessica nous

confesse d'ailleurs à la fin de son entretien qu'elle souhaiterait exercer son activité de vente de pâtisseries à temps plein suggérant alors ce travail d'à-côté s'inscrit dans une volonté de développement professionnel.

Certes, l'accès à un contrat en CDD ou CDI est synonyme de revenus fixes, mais leur faiblesse les pousse à diversifier leurs revenus. Et c'est sans doute sur ce point que le secteur se révèle particulièrement "extractif" d'autant plus si l'on considère la structure des coûts du secteur et l'accaparement de la valeur produite par les autres acteurs de la chaîne : entreprises de sous-traitance, entreprises d'IA française et entreprises hégémoniques du numérique. Lorsque nous lui demandons pourquoi il a répondu se sentir exploité dans le questionnaire, Nity indique que : *"[sur] le salaire, il y a un intermédiaire, nous ne recevons pas beaucoup parce qu'ils prennent leur part. J'ai l'impression d'être un peu exploitée, parce que je travaille dur et que le salaire n'est pas très bon."* (Nity - Annotateur - Sous-traitant A).

Dans ce contexte de négociations salariales entre une start-up d'IA et son sous-traitant, un employé de la start-up nous a donné accès à la structure des coûts salariaux de l'entreprise de sous-traitance. Ces éléments nous donnent des indications quant à l'appropriation de la valeur produite par les travailleurs. Au moment des entretiens, l'entreprise compte 110 salariés, en contrat de prestataires de services et vend ses services 3,30 euros par heure et par agent, soit environ 530 euros par mois et par agent. Sur cette base, on constate que l'entreprise a un chiffre d'affaires annuel qui se situe entre 650000 euros et 700000 euros, ce qui représente 390000 et 440000 euros de bénéfices nets d'impôts. Les locaux de l'entreprise appartenant à son propriétaire, il faudrait enlever la connexion Internet, qui peut s'élever par exemple à 165 euros par mois pour l'accès à 6 téraoctet de données et un débit de 100Mbits, ainsi que l'entretien des locaux et le matériel informatique. Malgré ces coûts, on constate qu'il s'agit d'une activité qui peut-être très rémunératrice pour les propriétaires d'entreprises de sous-traitance.

Néanmoins, cette situation correspond à une catégorie particulière d'entreprise : informelle et de taille moyenne. Dans le cas d'une autre entreprise, cette fois formelle et de plus grande taille, un annotateur nous a indiqué que la prestation est facturée 75 euros par jour et par agent (soit 1500 euros par mois pour 20 jours ouvrés) et dépend des compétences nécessaires pour le projet d'IA. Même si cette entreprise doit louer des locaux beaucoup plus grands dans la capitale malgache (350 postes à Madagascar), on peut raisonnablement penser que le bénéfice demeure important pour ses propriétaires, ce qui n'échappe d'ailleurs pas à l'un de leurs clients : *"par contre j'en suis convaincu, des boîtes comme ça te proposent un certain nombre de profils, et en fonction des qualifications (...), ils te proposent un budget journalier qui est en général 20 fois supérieur à ce que la personne va réellement toucher, parce que c'est leur business model, et comme tout truc de SaaS ou de vente de service, ils font 80% de profit, donc ce truc-là est complètement disproportionné."* (Cédric - Lead data Team - AIMedic)

Au sujet des autres acteurs de la chaîne, nos éléments sont plus parcellaires, mais permettent tout de même

de mieux comprendre la structure de la chaîne de valeur. Commençons d'abord par les entreprises d'IA en reprenant l'exemple de Automatik. Sur l'année 2021, une employée nous indique qu'ils ont facturé 200000 euros à CanteenCorp pour développer une caisse de cantine automatique. Par ailleurs, le document détaillant la structure des coûts et des bénéfices de l'entreprise sous-traitante indique que Automatik paye environ 100000 euros par an pour ses annotateurs. Les 19 annotateurs affectés à Automatik par le sous-traitant A ne sont cependant pas occupés à temps plein par le projet de cantine. Sur la base de nos entretiens, on peut estimer qu'ils travaillent un quart de leur temps sur ce projet (considéré comme compliqué), pour un coût d'environ 25000 euros pour ce seul projet. Enfin, si l'on considère maintenant les autres acteurs de la chaîne, c'est-à-dire les grandes entreprises du numérique fournissant des services d'hébergement de données et de puissance de calcul (infrastructures), il est déjà plus compliqué de calculer les coûts annuels, car cela dépend de la quantité de données nécessaires et du nombre de paramètres utilisés dans le modèle et donc de la complexité. Une employée d'Automatik nous indique toutefois qu'ils dépensent environ 100000 euros par an pour l'ensemble de leur projet en coûts liés à l'infrastructure.

Cette sous-section nous donne des informations essentielles sur la structure de l'extraction de valeur du secteur de l'IA. Bien qu'étant des éléments prépondérants dans la fabrication des modèles d'IA, les travailleurs malgaches sont très peu valorisés financièrement par les acteurs de la chaîne. Au-delà de leur place dans la structure des coûts, les travailleurs, et notamment ceux qui ont leur propre foyer et des enfants ("les travailleurs expérimentés et bien rémunérés" - Classe 3) vivent le salaire qu'ils tirent de l'activité d'annotation comme un revenu de subsistance qui peine à combler l'ensemble de leurs besoins. Comment interpréter ces résultats au regard de notre analyse de ces travailleurs comme appartenant à une classe moyenne en émergence ?

Pour les travailleurs des données, atteindre la stabilité financière par l'intégration dans le secteur privé de l'externalisation, passe par l'acceptation d'un fort risque de licenciement lié aux possibilités offertes par le droit du travail malgache. Les résultats présentés dans cette sous-section montrent également que cela implique de développer des activités secondaires susceptibles de compléter leur salaire d'annotateurs. Ces activités constituent une stratégie de développement professionnelle autant qu'une manière de gérer la pérennité financière du foyer. Entre un travail "d'à-côté" (Weber, 2009), et le travail entrepreneurial reproductif (Wilson & Yochim, 2017), ces activités constituent un élément essentiel du secteur de l'IA en ce qu'il permet aux travailleurs de faire : gérer la précarité de leur emploi d'annotateur, sans réellement pouvoir subvenir pleinement à leurs besoins ou accumuler de capital. Pendant ce temps-là, les entrepreneurs de la sous-traitance, les entreprises d'IA, et les grandes entreprises du numérique qui vendent l'accès à des infrastructures profitent de ce travail pour accumuler de la plus-value. La conséquence principale de cette chaîne de valeur globale est de limiter les perspectives d'avenir des travailleurs malgaches.

9.3.2 Perspectives personnelles et professionnelles futures des travailleurs. Faire avec des *emplois sans avenir*

Notre précédente sous-section a permis de mieux comprendre les perspectives économiques des travailleurs des données. Nos résultats suggèrent que le travail de l'annotation constitue une voie de subsistance empruntée par des travailleurs qui sont poussés par la nécessité de maintenir leur foyer à flot, notamment dans le cas de ce que nous avons appelé la *"classe moyenne assurée, mais précaire économique"*. Au moment d'aborder les perspectives professionnelles des annotateurs, il nous faut garder ces éléments en tête tant ils sont déterminants dans la mise en place d'une stratégie de carrière.

Une discussion avec l'amie de notre traducteur illustre les difficultés de ces travailleurs. Issue d'études universitaires reconnues et désormais spécialiste des ressources humaines, elle a notamment travaillé pour une entreprise qui propose des services de développement web sous forme de micro-tâches. Son avis sur le secteur est sans appel puisqu'elle considère qu'il s'agit d'une voie sans issue pour les travailleurs. Si elle concède que ce secteur permet à des jeunes de s'intégrer sur le marché du travail, elle nous explique en effet qu'il ne faut pas rester dans cette industrie *"après trente ans"*, au risque de s'y enfermer. Cette discussion illustre alors une question fondamentale des enjeux à long terme en matière de développement et de perspectives professionnelles pour les travailleurs : quelles sont leurs perspectives professionnelles ? Dans quelles mesures s'agit-il d'un secteur sans issue ?

Pour débiter cette analyse, il nous faut revenir sur les pistes d'évolution au sein du secteur de l'annotation et plus généralement de l'externalisation. Les annotateurs sont ainsi plus de la moitié à considérer que leurs probabilités de promotion sont *"à peu près inexistantes"* (15,1%) ou bien *"Faibles"* (38%). Si l'on analyse cette variable au sein des trois entreprises de notre échantillon, on constate que les annotateurs pensent avoir moins de possibilités de promotion au sein de la plus grande entreprise étudiée, le sous-traitant B, ce qui tend à confirmer les constats de [Buscatto \(2002b\)](#) sur le manque de mobilité ascendante à partir d'une certaine taille d'entreprise (autour de 400 salariés). Justin qui travaille chez ce sous-traitant explique d'ailleurs qu'*"avoir une promotion c'est assez rare[ici]"*. Par ailleurs, l'important vivier de travailleurs qualifiés notamment chez les plus jeunes, constitue également un frein pour des travailleurs ayant plus d'ancienneté, mais n'ayant pas le même niveau de diplôme que les plus jeunes travailleurs qualifiés. Ranto d'ailleurs explique que certains agents de ce sous-traitant *"se plaignent qu'ils [le sous-traitant B] ne prennent pas vraiment en compte l'ancienneté des agents [pour les promotions]"*.

A l'inverse, une entreprise comme le sous-traitant G semble avoir une politique de promotion très différente. Le dirigeant exécutif de l'entreprise, un français présent sur place pour gérer la production, nous explique qu'il a mis en place des grilles de progressions salariales basées sur les performances des annotateurs qui permettent à ces derniers de gagner plus au fil du temps. Par ailleurs, comme l'explique Ariel, annotateur :

”c’est la promotion interne qui est priorisée. (...) Parmi nous, pas mal de personnes ont commencé en tant qu’annotateurs et ont progressé au fil des années, pour atteindre le poste de chef de projet”, ce qui pourrait également justifier la part relativement importante d’annotateurs de l’entreprise G qui pensent probable d’obtenir une promotion. Chez le sous-traitant A, même si la promotion interne est valorisée, la politique de l’entreprise consiste à établir des seuils de qualifications pour les postes de management (Bac +2 pour un chef d’équipe et Bac +4 pour un responsable), ce qui complique les possibilités de promotion.

	Faible	Élevé ¹	Total
Sous-traitant G	51.1	48.9	100
Sous-traitant B	66.7	33.3	100
Sous-traitant A	34.6	65.4	100
Ensemble	51.8	48.2	100

¹Réponses en pourcentage, Fisher.test = 0,000149

TABLE 30 – Probabilité de promotion selon l’entreprise

Par ailleurs, les travailleurs sont 67,4% à considérer qu’il serait ”Plutôt difficile” de trouver un emploi similaire, ce qui pourrait suggérer des difficultés à valoriser leur expérience professionnelle dans des activités similaires (annotation) ou connexes (centre d’appels) afin de trouver un emploi ressemblant. Les exemples de parcours professionnels d’annotateurs ayant connu plusieurs entreprises d’annotations nuancent ce résultat. Avec le développement du secteur, des postes s’ouvrent pour des travailleurs ayant déjà une expérience dans ce secteur, mais pas nécessairement à des niveaux plus élevés que leurs précédentes expériences. En revanche, si nous avons pu constater une circulation des travailleurs entre les activités du secteur des services numériques, elle n’est pas forcément accessible à l’ensemble des annotateurs, tant certaines activités demandent des ”qualifications coloniales” plus poussées, notamment pour les activités de rédaction ou de relations clients ³²⁶. Ces activités sont pourtant des facteurs de développement salarial des travailleurs puisqu’elles sont connues comme étant plus rémunératrices que l’annotation, encore associée à la saisie de données dans les entreprises de sous-traitance non spécialisées dans l’annotation.

Plusieurs facteurs ouvrent cependant la porte à des opportunités de mobilité ascendante pour les travailleurs. Tout d’abord, au niveau de l’ensemble de l’externalisation de services informatiques, les chiffres de l’OMC montrent qu’il s’agit d’un secteur qui a doublé entre 2017 et 2021 (70 à 140 millions de dollars d’exportation) ce qui suggère l’existence de possibilités d’évolutions liées à croissance de ce marché. Ensuite, le fait qu’il s’agisse d’un secteur impliquant des tâches variées et des processus complexes pour la

326. ”Mon travail idéal ? Je crois que je suis fait pour un poste de modérateur, ou bien d’opérateur BPO. J’ai jamais pensé à travailler en tant que téléopérateur ou conseiller client. Parce que je dois apprendre beaucoup de choses d’abord (...)” (Jean - Ancien annotateur)

maîtrise de la qualité des modèles de l'IA justifie l'existence de postes équivalents à des ouvriers spécialisés de l'annotation (agents experts). La présence de ces postes intermédiaires, bien que parfois sous-estimés en termes de rémunération, peut servir de substitut aux perspectives de mobilité salariale.

La mobilité ascendante peut néanmoins être entravée par la structure organisationnelle et la répartition des responsabilités au sein des entreprises du secteur. Le fait qu'il n'y ait qu'un manager intermédiaire pour six travailleurs et un responsable de la production pour vingt personnes - seulement 16,6 % des postes sont des postes intermédiaires et 4 % des postes de responsables - peut créer des obstacles à l'avancement pour de nombreux employés. Dans des domaines clés tels que la commercialisation, la recherche et développement, la formation et la gestion de la qualité, certaines grandes entreprises offrent des postes plus élevés, mais ces équipes demeurent restreintes, limitant encore davantage les opportunités de progression pour la majorité des travailleurs.

Le manque de capacité des entreprises sous-traitantes à facturer aux clients des postes de management peut également limiter le développement de postes intermédiaires. Cette limitation découle d'une pression économique inhérente à l'industrie des services, où les coûts doivent souvent être minutieusement gérés pour maintenir la compétitivité. Les entreprises cherchent à maximiser leurs marges bénéficiaires en facturant des services à un tarif compétitif, tout en limitant les coûts de main-d'œuvre supplémentaires. Comme on a pu le voir précédemment, cela dépend également de la capacité à valoriser ces compétences auprès des clients et à outrepasser la discrimination aux compétences que subissent les travailleurs malgaches.

Encadré n°78 : La discrimination envers les travailleurs malgaches

S : [avec ce client] (...) il y a toujours des petits trucs ... j'ai toujours eu des remarques un peu condescendantes (...), quand on a commencé avec ce projet. J'ai juste fait une remarque assez importante, et il m'a remercié, il était étonné de la question aussi parallèlement à cela, il a dit « ah, il y a beaucoup de nuances entre la France et Madagascar aussi, on le voit sur votre visage (...) donc là Mathieu était gêné de la chose, notre gérant était gêné aussi mais il pensait que j'avais pas capté, moi j'étais toujours en sourire ... mais je savais très bien ! il y a toujours des petites choses : « ah, voici Simon, Il est chef de projet sur beaucoup de projets et on le remercie bien » [ton mielleux]... comme ça des petites choses comme ça, on sait qu'il y a un petit peu de ... mais y a pas de souci, c'est le boulot, c'est toujours comme ça

Simon - Chef de projet - Sous-traitant G

En conséquence, l'expansion de postes intermédiaires ou de soutien qui exigent des compétences plus avancées peut être freinée, car ces compétences supplémentaires peuvent nécessiter des salaires plus élevés ou

une augmentation des coûts de formation. Or ces frais supplémentaires sont en contradiction avec la pression à la baisse des tarifs proposés aux entreprises clientes. Les entreprises sous-traitantes doivent donc équilibrer la demande de compétences supérieures avec la rentabilité de leurs opérations. Cette situation peut conduire à une réticence à créer des postes intermédiaires ce qui limite de développement de carrière des travailleurs.

Après avoir parlé des perspectives au sein du secteur de l'annotation et des services informatiques, regardons comment les annotateurs de données perçoivent leur développement professionnel. Pour mieux cerner ce qui favorise ou non leur projection dans le secteur, nous réalisons une régression logistique binomiale, dans laquelle nous utilisons la variable *"Heureux que votre enfant fasse la même activité"* comme *proxy* de l'envie des annotateurs de rester dans le secteur en tant que variable à expliquer. Nous reprenons ensuite un certain nombre de variables explicatives liées aux conditions de travail : sentiment d'être bien payé, sentiment d'être considéré à sa juste valeur, sentiment d'utilité, sentiment d'ennui, sentiment de stress, pénibilité du travail, sentiment d'être interchangeable, probabilité d'être promu, probabilité de perdre son emploi et celle de retrouver un emploi équivalent. Nous utilisons également plusieurs variables de contrôle : l'âge, le statut marital, la présence d'un enfant dans le foyer et le score de position sociale. Le résultat de ce modèle nous permet d'identifier le facteur "salaire" comme étant significatif, en faveur du fait de rester dans l'annotation. En revanche, l'ennui et le stress, que nous avons déjà évoqué dans notre partie sur les conditions de travail constituent des freins aux aspirations professionnels à l'intérieur du secteur de l'annotation.

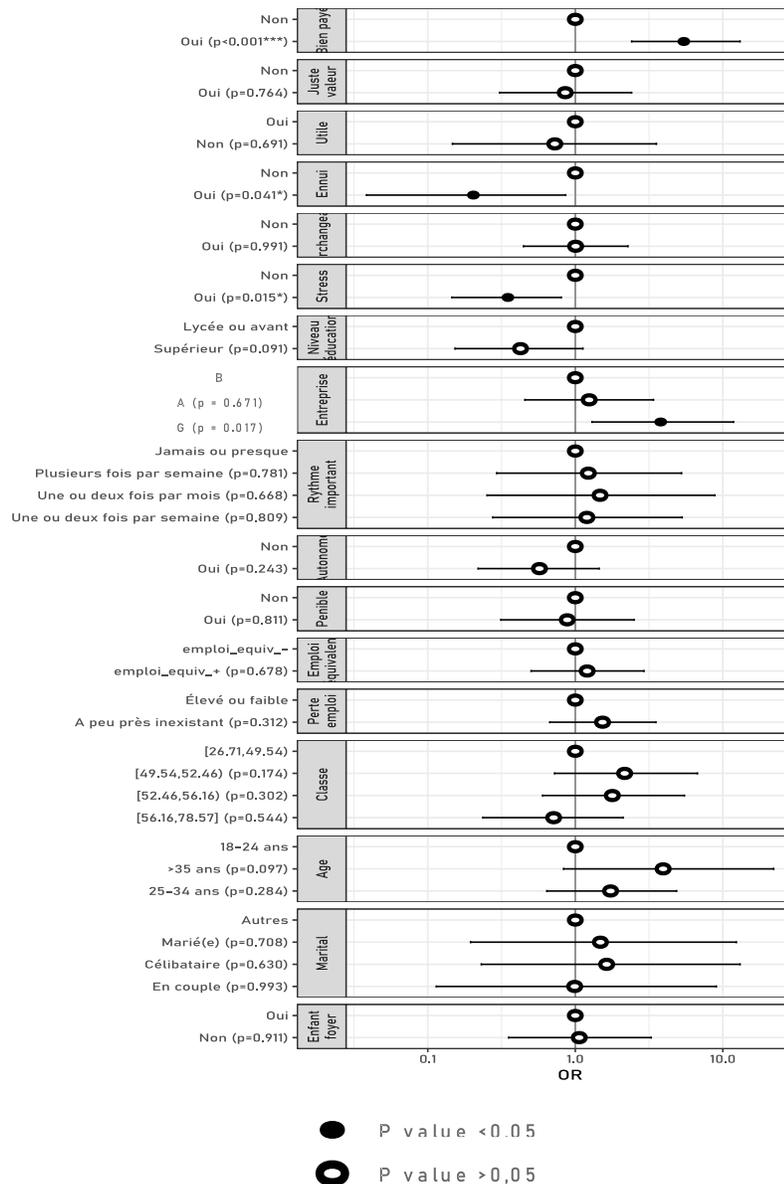


FIGURE 50 – Variable explicative du fait que les annotateurs soient heureux que leurs enfants fassent le même métier qu’eux

Concernant l’ennui, Dafy (sous-traitant G) nous confesse que *”le travail c’est vraiment de la routine, des fois tu commences à oublier toutes les études que tu as apprises, aussi sur la pratique des langues ... ”* (Dafy - Annotateur - Sous-traitant G). Cet extrait d’entretien met en lumière la nature répétitive et routinière de cette activité qui impactent négativement sur la motivation et la valorisation personnelle des travailleurs. Cela

reflète également la déconnexion entre les qualifications potentiellement plus avancées acquises durant les études et les tâches simples et répétitives qu’il accomplit au quotidien.

Cette situation soulève des questions sur l’utilisation des compétences et des qualifications des travailleurs dans le contexte du travail d’annotation de données, ainsi que sur l’éventuelle dévalorisation de ces compétences au sein de ces entreprises de sous-traitance. Si certaines entreprises mettent en place des organisations destinées à enrichir le travail d’annotation par la variété des types d’annotations ou de projets ou bien par la variété des fonctions occupées au sein des projets (vérificateur, annotateur spécialisé), les annotateurs soulignent largement le manque de stimulation intellectuelle de cette activité.

Finalement, l’aspiration des travailleurs à effectuer des tâches plus gratifiantes et stimulantes que l’annotation de données ou les services informatiques met en lumière une question fondamentale : quels sont les rêves et les aspirations des ”petites mains” des algorithmes ? En nous inspirant du travail de [Fournier, Lambert, et Marion-Vernoux \(2020\)](#) sur les aspirations des jeunes travailleurs français, nous avons ainsi classé les extraits d’entretiens en fonction des perspectives affichées par les annotateurs, ce qui nous donne cinq catégories : la poursuite d’études (4/66), le fait de rester dans le secteur de l’externalisation au sens large (8/66), le fait de travailler dans un autre secteur (11/66), le fait de trouver un travail en lien avec les qualifications acquises ou espérées dans le contexte des études (18/66) et le fait de devenir indépendant (18/66).

	Nombre (N=66)
Poursuite études	4
Rester dans l’externalisation	8
Travailler dans un autre secteur	11
Trouver un travail en lien avec les qualifications acquises ou espérées	18
Devenir indépendant	18

TABLE 31 – Les perspectives professionnelles des annotateurs de données

Ces résultats suggèrent que les annotateurs souhaitent majoritairement développer un projet professionnel hors de ce secteur, tout d’abord par la poursuite d’études. Félix (sous-traitant G) nous explique vouloir : *continuer ma formation que j’ai fait en première année à l’université. Mais jusqu’à maintenant je n’ai pas les moyens. Mais je pense que après quelques années ou quelques mois, j’aurai l’occasion de faire ça*. Félix exprime ainsi sa volonté de poursuivre sa formation universitaire, une ambition qui peut être vue comme une réponse aux défis et aux limites de son travail actuel. Néanmoins, son incapacité actuelle à financer cette formation met en lumière les obstacles socio-économiques auxquels les travailleurs malgaches sont confrontés, même

lorsqu'ils ont des aspirations éducatives³²⁷.

Confronté à d'autres difficultés, Eja nous indique vouloir continuer ses études, malgré un *"problème de financement (...) "*. Selon, il n'y a *"pas d'avenir de travailler [chez le sous-traitant A] parce que je vous avais déjà dit qu'il y a des risques très élevés de perdre mon travail l'année prochaine. Si je me vois évoluer mais vous voyez bien, y a un problème de financement (...) "*. Son plan de carrière consiste alors à obtenir un diplôme et à devenir travailleur indépendant. Ce dernier met donc ses préoccupations quant à la stabilité de son emploi actuel. Cette précarité, caractéristique du secteur de l'annotation de données, l'incite à poursuivre ses études, ce qui reflète également une aspiration à l'ascension sociale et à l'indépendance professionnelle, qui peut être considérée comme une réponse aux limites et aux incertitudes de son emploi actuel. Seta manifeste une volonté de travailler comme indépendant, qui est partagée par d'autres travailleurs. Ces derniers considèrent qu'il est possible de gagner plus d'argent de cette manière. Nivo, chef de projet chez le sous-traitant D nous explique : *"en fait je rêve depuis toujours d'être auto-entrepreneur, mais je n'ai pas encore choisi dans quel domaine"*. S'il évoque des pistes dans la boulangerie ou la photographie, il nous raconte qu'il est *"très compliqué"* de créer sa propre entreprise.

Ces différents exemples d'aspirations professionnelles expriment une déception vis-à-vis du travail salarié, notamment en matière de partage de la valeur et des conditions de travail. Ainsi, y compris les salariés qui souhaitent rester dans le domaine de l'externalisation aimeraient y exercer comme indépendant : *"Si je devais partir d'ici ? Comme tout le monde, j'ai l'ambition d'être indépendant. Par exemple, si j'ai des clients en France, c'est le rêve ici. Tout le monde rêve de travailler indépendamment."* (Fiary - Sous-traitant B). Ces perspectives professionnelles s'expliquent en partie par la structure informelle du marché du travail à Madagascar, ainsi que par l'importance de la double activité pour les travailleurs de l'externalisation. Comme on l'a vu dans notre précédente sous-section, ces derniers créent parfois leur propre affaire ou bien exercent des activités liées à l'externalisation en tant qu'indépendants en passant par des groupes ou des plateformes numériques de travail. Ce travail "d'à côté" peut constituer une perspective professionnelle sérieuse, qui se heurte toutefois à deux difficultés : le manque de capital lié à des salaires insuffisants et l'absence d'indemnités de chômage. Par ailleurs, dans le cas d'activités d'externalisation, les travailleurs qui souhaitent développer leur activité se heurtent à des discriminations, ce qui se traduit par des difficultés à trouver des clients ainsi que parfois à être payé.

Encadré n°79 : Une discrimination sur le fait de payer les annotateurs

Z : UpWork vous avez fait aussi ? S : [Oui sur] des projets d'informatique (...) comme les annotations (...) en vidéo plutôt, donc on classifiait les vidéos sur YouTube. Donc les reviews, les ASMR, on classait

327. Le coût des études supérieures constitue le facteur principal expliquant l'arrêt des études

tout ça, mukbang (...). C'était très maigre aussi. Des fois ça explosait, enfin façon de dire ... quand on a été au bout de certains temps, en 2 ou 3 mois, on a commencé à retravailler pour quelqu'un. Parce que là on a discuté, le mieux ici c'est d'être entrepreneur plutôt que de toujours travailler pour quelqu'un, mais ce n'est pas facile

Z : vous avez arrêté au bout de 3 mois pourquoi ?

S : parce que là on était pas payé, au bout d'un certain temps, on a pensé engager quelqu'un mais c'était trop risqué, et puis pour les charges vous voyez, on payait l'électricité, on payait aussi les crédits [Internet]

Z : ça fait beaucoup de frais

S : oui, on a rien, on était comme des cons, on ne savait pas quoi faire, il y avait les charges en même temps, on est marié, on a des enfants ...

Simon - Chef de projet - Sous-traitant G

Les annotateurs qui souhaitent rester dans le secteur de l'externalisation sont pris par la nécessité d'assurer la stabilité économique de leur foyer. Ainsi, Ando (sous-traitant B) explique que *"mon environnement c'est l'informatique, alors tout ce qui est informatique, mais plus rentable. En fait, j'ai 3 enfants en ce moment, et pour moi, subvenir à nos besoins c'est la seule priorité pour moi. Alors si je devais partir de [cette entreprise] je chercherai quelque chose de plus avantageux que ceci, je ne vais pas me rétrograder"*. Cet extrait reflète une préoccupation fondamentale pour la sécurité économique de sa famille. La sortie de son emploi actuel implique de pouvoir améliorer sa rémunération et ses conditions de travail pour subvenir aux besoins de sa famille. Cette attitude met en lumière les aspirations individuelles de travailleurs face à un contexte où la sécurité de l'emploi et la recherche de meilleures opportunités financières sont des préoccupations majeures.

Enfin, pour certains annotateurs, le fait de quitter ce domaine correspond à une volonté de s'aligner avec leurs études et leurs compétences acquises initialement : *"j'aimerais faire toujours le travail qui a un rapport avec mon parcours universitaire [en] agronomie"*. Cette perspective professionnelle correspond alors à la volonté des travailleurs de valoriser leurs qualifications par la création d'entreprises qui correspondent à leurs études. C'est également le cas de Dafy qui a étudié le tourisme et qui souhaite se rapprocher de ce secteur, en créant sa propre entreprise : *"j'ai étudié le tourisme, et (...) ça rapporte beaucoup d'argent et mon projet à long terme c'est de créer un tour opérateur, ou une agence de location de voiture, parce que je trouve que ça me conviendrait bien, d'être le gérant ou bien le propriétaire de ces entreprises (...)"* (Dafy - Annotateur - Sous-traitant G). Il met ainsi en évidence le décalage entre sa formation académique en tourisme et son emploi actuel d'annotateur de données. Il reconnaît ainsi l'écart entre son travail actuel, qui ne correspond pas à son domaine d'études, et ses aspirations à long terme centrées sur son expertise en tourisme.

Dans cette sous-section, nous contribuons à analyser la tension entre la nécessité de créer de l'emploi dans les pays du Sud global et la qualité effective des emplois proposés. Nous mettons en évidence une situation où des individus dotés de solides qualifications et effectuant un travail essentiel se retrouvent à faire face à une minimisation de leurs contributions dans cette industrie. Cette tension se révèle encore davantage lorsque l'on met en balance les rêves d'avenir de ces travailleurs. Un des moments les plus marquants de l'enquête restera d'ailleurs le discours tenu par un jeune annotateur qui nous indique à la fin de l'entretien qu' "ici [à Madagascar] on utilise un jargon, on appelle ce type de travail un *sérum*, parce que c'est [juste] un job qui nous empêche de mourir" (Annotateur - Sous-traitant A).

Nos résultats montrent ainsi que cette activité se caractérise finalement par le manque de perspectives professionnelles dans et hors du secteur. Il s'agit sans doute d'une des caractéristiques les plus problématiques du vécu des travailleurs au sein de ce secteur. Ces derniers peuvent se sentir coincés dans un quotidien où ils doivent accepter ces emplois précaires pour subvenir à leurs besoins immédiats sans avoir le temps de se former ou d'acquérir les compétences nécessaires pour progresser dans leur carrière. Et ce, avec des perspectives limitées d'évolutions professionnelles. Le fort développement du secteur, ainsi que la complexité des processus de production des modèles d'IA laissent toutefois entrevoir des possibilités de structuration de l'industrie et la mise en place de nouveaux postes intermédiaires (Mañé Vernet & Benner, 2009).

Ce constat que l'on retrouve également dans les centres d'appels ou de services des pays occidentaux est d'autant plus difficile à vivre que ces derniers n'ont pas énormément de perspectives d'avenir hors du secteur de l'externalisation. Certes, nos données, principalement focalisées sur les annotateurs actuellement en poste, ne permettent pas de vérifier empiriquement le devenir professionnel de ces travailleurs sur le long terme. Notre matériau au sein duquel on trouve des travailleurs présents parfois depuis plusieurs décennies dans cette industrie et d'autres en attente de trouver un travail, nous donnent néanmoins des indications concernant les perspectives professionnelles de ces derniers hors de l'externalisation. Face à des publications qui mettent en avant le développement d'une employabilité à long terme et l'acquisition de compétences génériques par les travailleurs du secteur, nous avons montré que ce développement professionnel est limité par l'existence d'une identité professionnelle floue (entre valorisation et minimisation des compétences), par la précarité de l'emploi, par le manque de temps et de moyen pour la formation ainsi que par la discrimination subie par les travailleurs de la part des clients.

Il convient par conséquent analyser les perspectives liées au développement de Madagascar au regard des perspectives offertes par ce secteur d'activité. Or, nous constatons qu'il s'agit d'emploi sans avenir, tout juste capable de permettre aux travailleurs d'assurer leur survie. Les évolutions possibles au sein de la même catégorie d'emploi sont très rares et cette activité ne constitue pas non plus un *catalyseur de perspectives professionnelles* dans d'autres secteurs, ou selon d'autres modalités comme l'entrepreneuriat. Pourtant, nos données quantitatives et nos entretiens démontrent très largement les qualifications des travailleurs mal-

gaches. Face à leur précarité, ce désir d'entreprendre illustre à la fois ces compétences et la résilience face à des problèmes structurels qu'ils vivent au quotidien ³²⁸. Face à ces écueils, il existe des pistes de politiques publiques, comme le développement de la formation professionnelle, encouragée par les organisations de développement internationale et par le gouvernement malgache à travers la création en 2017 du Fonds Malgache de Formation Professionnelle. Aucune donnée ne permet pourtant d'affirmer que des "reconversions professionnelles" sont possibles pour les travailleurs d'un pays comme Madagascar, et ce d'autant plus si ces derniers travaillent dans un secteur entièrement vers l'étranger et non vers les besoins locaux. Dès lors, les entreprises du secteur ont un rôle à jouer pour faciliter l'accès aux études et pour faciliter l'accès à des financements pour la création d'entreprise en améliorant substantiellement les conditions salariales.

* * *

Dans notre précédent chapitre, nous avons évoqué les facteurs macro et micro qui contribuent à la formation d'une industrie de la donnée et d'une force de travail qualifiée, capable de produire des données pour l'IA (Pun et al., 2020). Les éléments que nous avons présentés dans ce chapitre contribuent à la littérature sur les effets des chaînes globales de valeurs en montrant les conséquences locales des chaînes de valeur globales. En restituant la part de subjectivité des travailleurs concernant leurs revenus, ainsi que leurs stratégies de *débrouille* face à la "précarité" (Cornet, Cierpe-Torres et Le Ludec, à paraître), nous contribuons à montrer que l'existence des chaînes de valeur globales est consubstantielle à la mise en place de chaînes de précarité locale.

Le développement de cette industrie est facilité par l'offre pléthore de travailleurs (Graham et al., 2017a), ainsi que par la structure encore largement informelle de l'économie malgache qui offre peu de protections aux travailleurs, quelque soit l'activité. Les travailleurs urbains et qualifiés que nous étudions mettent alors en place des stratégies pour travailler dans des industries plus protectrices, comme le secteur de l'externalisation. Néanmoins, cette protection sociale est fortement limitée par l'absence de droit au chômage et par les possibilités de licenciement économique. Ces résultats apportent donc une nuance significative au postulat d'une stabilisation financière par l'intégration dans une entreprise formelle. Une simple détection d'un problème de qualité dans le travail des données ou un manque de fonds pour engager des annotateurs peut entraîner l'arrêt de projets et le licenciement des travailleurs. Pour faire face à cette instabilité, certaines entreprises mieux structurées mettent en place des services de recherche et développement (R&D) pour évaluer la faisabilité et la pérennité des projets.

Cette précarité se traduit alors par des difficultés d'intégration sur le marché du travail dont témoigne la carrière des travailleurs des données. Pour certains, nous parlons d'une *intégration assurée* sur le marché du

328. Reprenant un argument imaginé par Antonio Casilli et Juana Cierpe-Torres (2022), nous parlons de débrouille pour caractériser cette attitude dans un article à paraître (Cornet, Cierpe-Torres et Le Ludec 2024)

travail, car ils disposent d'importantes ressources sociales mobilisables pour trouver un emploi, mais leur revenu est insuffisant en raison de leur contexte familial. Ils ont leur propre foyer, des enfants et subissent donc davantage une précarité économique liée au coût de la vie à Antananarivo (*classe moyenne sous pression*). Pour d'autres, l'intégration sur le marché du travail est *incertaine*. Ils ont un revenu qu'ils considèrent comme suffisant, car ils vivent avec leurs parents. Cette caractéristique leur permet de se davantage projeter dans des études et dans la construction de leur profil professionnel (*classe moyenne en transition incertaine*). Pour d'autres enfin, l'intégration laborieuse car, ils ont un revenu insuffisant pour construire leur vie de couple, ce qui les conduit à vivre mal leur expérience professionnelle actuelle (*petite précarité urbaine*).

Ces résultats permettent alors de réinterpréter nos conclusions sur l'émergence d'une classe moyenne à Madagascar. Si la littérature déjà citée précédemment identifie la volonté de stabilité comme une caractéristique de la classe moyenne dans les pays du Sud global, nous montrons à quel prix cette stabilité financière s'obtient à Madagascar : multiples activités, travail reproductif et activité principale soumise aux aléas du marché (financement des start-ups...) et donc peu sécurisée dans le temps sont le prix d'une stabilité du quotidien. En comparaison avec les "petites prospérités" agricoles analysées par (Andrianampiarivo, 2018), qui détiennent leur moyen de production et qui accumulent du capital, cette classe moyenne urbaine se caractérise par une forte précarité liée à la structure extractive du secteur de l'IA : conditions de travail peu protectrices et incapacité d'accumuler un capital leur permettant de développer leur propre activité.

Ces éléments structurels sont également contrariés par d'autres facteurs. La "démographie professionnelle" du secteur limite les perspectives de mobilité ascendante à l'intérieur du secteur, tandis que la création d'entreprises locales est alors entravée par une double inégalité, à la fois raciale et financière, soulignant l'impact persistant de la structure raciale et post-coloniale du secteur sur les travailleurs. Cette inégalité est également illustrée par la structure du management des entreprises de sous-traitance, avec de nombreux postes de direction réservée aux français des entreprises de sous-traitance ou des start-up d'IA.

Conclusion de la partie 4 - Le travail d'annotateur : limiter la précarité en attendant mieux

Nous avons analysé la chaîne de valeur de l'IA sous l'angle de l'"*extractivisme de données*", en articulant des questionnements centrés sur la qualité des activités professionnelles induites par l'IA et sur les effets de cette industrie sur les travailleurs malgaches. L'adoption d'une perspective *macro* nous a permis d'examiner la structure du secteur à travers les modalités de son émergence. Nous avons remis en question le concept de "*data colonialism*" en tant que conception inadéquate des mécanismes d'extraction de valeur des grandes plateformes numériques, en particulier dans le secteur de l'intelligence artificielle. Puis, suivant l'appel à

un "tournant décolonial du numérique" d'Antonio A. A. Casilli (2017b), nous avons proposé de renverser le concept de colonialisme des données, afin d'analyser le secteur de l'IA à travers sa dimension extractive et de la "colonialité numérique" vécue par les travailleurs malgaches.

En réinscrivant l'industrie de l'IA dans la dynamique post-coloniale à l'oeuvre dans l'externalisation de services liés aux données, nous avons abordé les effets structurels de l'industrie de l'IA, tout en mettant en lumière les stratégies individuelles développées par les travailleurs malgaches pour y répondre. Ces derniers mettent en effet en place des stratagèmes pour assurer leur subsistance en se conformant aux besoins de qualification des entreprises de sous-traitance ou pour construire leur avenir en émigrant dans les pays occidentaux. Les modalités d'apprentissage à Madagascar sont donc façonnées par les besoins des pays du Nord. Cette partie illustre également les perspectives professionnelles contrastées ouvertes aux travailleurs des données. Certes, ils peuvent y obtenir une relative *stabilisation financière*, mais l'inscription du secteur dans une dynamique post-coloniale limite leurs possibilités d'intégration dans le marché du travail, et ce, malgré leurs qualifications et leur appartenance à l'équivalent d'une classe moyenne en émergence. Nous proposons donc à la lumière de ces résultats d'intégrer davantage la notion de précarité à celle de chaîne de valeur trop prompt à souligner les bénéfices en matière de développement, pour évoquer une *chaîne globale de précarité locale*.

La poursuite d'un allongement géographique des chaînes de valeur qui profite des stratégies institutionnelles de développement

Nous parlons d'*extractivisme des données*, car les secteurs du numérique et de l'IA reposent sur des relations de production post-coloniales, en grande partie en raison des disparités de développement entre les pays du Sud global et les pays du nord, ainsi que sur les politiques publiques mises en oeuvre pour attirer des capitaux étrangers. En examinant les dispositifs publics d'attraction des investisseurs, nous avons constaté qu'ils favorisent l'établissement d'investisseurs français et mauriciens au détriment des entrepreneurs locaux, ce qui renforce la dimension extractive du secteur.

Nous avons également défini deux catégories de dirigeants présents dans ce secteur : les "*entrepreneurs de développement*" et les "*entrepreneurs des marges du développement*". Les premiers ont des liens familiaux ou personnels avec Madagascar tandis que les seconds exercent au sein d'entreprises de sous-traitance dont la stratégie repose sur l'exploitation de pays dits *subalternes*. Alors que les premières affichent une *rationalité de développement*, les seconds sont uniquement motivés par des coûts de main-d'oeuvre les plus bas possibles. Les "*entrepreneurs de développement*" combinent une connaissance approfondie de Madagascar avec un réseau extérieur permettant de trouver des donneurs d'ordres. Les "*entrepreneurs des marges du développement*" ont la spécialité d'opérer depuis des zones en périphérie de la mondialisation recherchant le ratio optimal entre la qualification de la force de travail et son coût. Il s'agit à présent de grands groupes qui

possédant des sites de production dans plusieurs pays et qui sont désormais largement financiarisés. Leur perspective de croissance est uniquement fondée sur leur capacité à équilibrer le prix de vente de leur prestation, basée sur les qualifications de leurs travailleurs, et le coût du travail en fonction des caractéristiques des pays où ils installent leur production. Ce modèle d'affaire constitue aujourd'hui un des moteurs d'une mondialisation des services tant ces entreprises se développent : une entreprise de services comme Téléperformance, présente dans de nombreux pays dont Madagascar a par exemple vu son résultat net croître de 71,9% entre 2020 et 2021.

Nous avons également montré que la population de travailleurs des données à Madagascar est bien formée. Malgré cela, les Malgaches sans expérience d'études ou de travail à l'étranger restent en marge de la création d'entreprises de zones franches qui restent le domaine réservé des entrepreneurs français. Cela confirme les travaux précédents sur les zones franches d'exportation, réservées aux personnes ayant des "réseaux à l'extérieur du pays". Nous proposons d'interpréter ces résultats comme révélateur d'une forme de "capitalisme racial" qui repose sur des pays en marge de la mondialisation.

Enfin, nos résultats indiquent que dans le secteur des services, "le monde n'est pas plat". Au contraire, les entrepreneurs de la sous-traitance délocalisent leurs activités dans des pays à la marge pour garantir leurs marges. Au-delà d'un rapport coût/capacité, la localisation géographique de cette chaîne de production s'inscrit tout d'abord dans un processus de réintermédiation. De plus en plus d'entreprises françaises ou américaines délocalisent plus loin géographiquement tout comme d'autres, déjà installées dans des pays comme l'Inde ou le Maroc, qui externalisent à présent à Madagascar. Le développement du numérique à Madagascar ainsi que la remontée d'autres pays du Sud global dans la chaîne de valeur constituent des facteurs de cette localisation géographique. Nous interprétons ces résultats comme l'illustration d'une forme d'intégration qui ne se limite pas aux coûts de main-d'œuvre, mais dépend également des capacités de celle-ci. A la suite de la typologie de Kleibert et Mann (2020), nous pourrions parler d'intégration fonctionnelle pour illustrer ce choix basé sur le rapport coût/capacité. Nous avons introduit le concept d'intégration marginale, considérant que cette typologie ne décrit pas entièrement la rationalité qui guide les entrepreneurs dans leurs choix de localisation. Nous parlons de marginale pour désigner un processus d'intégration influencée par l'état du développement de Madagascar, sa position subsidiaire dans la mondialisation et son intégration des institutions de l'ancien pays colonisateur comme l'Alliance Française, malgré une attention aux compétences des travailleurs. Les travailleurs malgaches, en raison de leur position géographique, reçoivent des salaires inférieurs malgré leurs qualifications, fait révélateur d'un certain "capitalisme racial" ciblant les pays en périphérie de la mondialisation.

L'extractivisme des données par l'orientation des stratégies d'apprentissage à Madagascar

Dans cette partie, nous avons également examiné la tension entre la reconnaissance et la dévalorisation des

compétences des travailleurs. Le rapport qu'entretiennent les travailleurs malgaches avec les pays occidentaux joue un rôle clé dans la mise en place de leurs stratégies de subsistance par l'investissement dans leur formation. Nous avons en effet constaté que les perspectives d'émigration, ainsi que les possibilités offertes par le secteur de la sous-traitance orientent nos enquêtés vers des compétences que l'on pourrait qualifier de *coloniales* (par exemple : linguistiques et culturelles), pour améliorer leurs perspectives de vie.

Ce processus d'*assetisation coloniale* permet aussi de décrire les obstacles auxquels sont confrontés les travailleurs lorsqu'ils envisagent un avenir professionnel dans un domaine, le travail des données, et un pays, Madagascar, offrant peu de perspectives, à moins d'émigrer vers la France et l'Europe. Combinées à la place prépondérante dominante du secteur de l'externalisation dans la capitale malgache, ces éléments restreignent considérablement les perspectives des travailleurs. Le vécu de ces derniers est alors marqué par l'entrelacement entre les vestiges des institutions coloniales comme l'Alliance Française et la présence de nouveaux acteurs français du secteur des services, qui conditionnent les modalités d'acquisition de leurs qualification et leur accès au marché du travail. Nous avons également examiné la tension entre la reconnaissance et la dévalorisation des compétences des travailleurs, en montrant que leurs perspectives restent limitées par le contexte colonial dans lequel ces compétences sont acquises. Les travailleurs malgaches sont contraints de s'adapter à ces dynamiques pour valoriser leurs compétences, tout en étant confrontés à des obstacles persistants : pour émigrer, pour trouver d'autres activités ou bien pour créer leur propre entreprise.

Malgré les discours de certains entrepreneurs de la sous-traitance, il est évident que ces derniers accordent de la valeur aux compétences linguistiques et culturelles héritées de la colonisation détenus par les travailleurs. Cette appréciation est également soulignée par les clients finaux, ce qui contribue à la pérennité du secteur dans le pays. Cette rhétorique de la fiabilité est également une stratégie des entreprises de sous-traitance et d'IA qui essentialisent l'identité coloniale malgache pour se présenter comme concurrentiel sur le marché mondial : ce *rappel de la colonialité* est donc utilisé de manière instrumentale pour les besoins d'entreprises internationales (Padios, 2018). Dans le cas de Madagascar, nous parlons d'une "*rhétorique des qualifications coloniales*" au sens où les entreprises de sous-traitance mettent l'accent sur la complexité et la variété des compétences requises dans ces emplois, tout en essentialisant l'identité des travailleurs malgaches selon leurs compétences coloniales pour les tâches d'annotation, au détriment des autres savoir-faire des travailleurs mobilisés ou non dans l'annotation. Ces compétences sont le plus souvent négligées ou mises en avant par les entrepreneurs de la sous-traitance et de l'IA, uniquement en fonction des besoins spécifiques des clients ou des projets. Nous proposons de considérer cela comme un *paravent des qualifications coloniales*. Face à cet *extractivisme des données* et à cette *colonialité numérique*, les travailleurs malgaches sont contraints de s'adapter pour valoriser leurs compétences, tout en étant confrontés à des obstacles qui limitent les bénéfices qu'ils tirent de la chaîne de valeur. Cette complexité souligne l'importance de prendre en compte les voix marginalisées et les contextes locaux dans l'analyse des enjeux du travail dans l'industrie

de l'externalisation des données.

Une classe moyenne en émergence dont la stabilité financière dépend des possibilités offertes par le secteur de l'externalisation

Le constat de l'émergence éventuelle d'une *classe moyenne urbaine* liée à l'industrie des services illustre les constats de Banerjee et Duflo (2008) sur le lien entre classe moyenne et pérennité financière du foyer. Ce résultat s'inscrit dans la littérature sur la constitution d'une classe moyenne dans les pays du continent africain en illustrant le passage rapide entre le secteur de l'agriculture et le secteur des services et la capacité du secteur privé des services à attirer des jeunes diplômés (Opoku, 2010). Nous montrons que pour des travailleurs malgaches faisant face à nombreuses difficultés structurelles, telles que la difficile intégration dans l'emploi formel, le secteur de l'externalisation de services constitue une piste de stabilisation financière.

Néanmoins, l'émergence de cette *classe moyenne* ne saurait masquer les variations qui existent dans notre population, au sein de laquelle on distingue trois groupes hétérogènes dans leur capacité à faire face aux aléas de la vie.

Nom	Ressources sociales	Ressources économiques	Type foyer	Niveau d'études
Petite précarité urbaine	Faibles	Faibles	Autonome	Moyen
Petite prospérité urbaine en transition	Moyennes	Moyennes	Parental	Important
Classe moyenne urbaine stabilisée	Importantes	Importantes	Autonome avec enfant	Important

TABLE 32 – Trois groupes de travailleurs des données de la classe moyenne urbaine

Au cours de cette partie, nous avons également analysé des séquences professionnelles, afin d'identifier les caractéristiques des trajectoires professionnelles des annotateurs de données. Sur la base de la littérature existante, nous pensions que les difficultés structurelles de Madagascar entraîneraient une entrée rapide dans ce secteur. Nos données empiriques confirment au moins partiellement cette hypothèse. En effet, deux groupes de trajectoires se caractérisent par une entrée rapide dans le secteur de l'externalisation et par la présence de contrats de travail formels dans l'expérience préalable à leur entrée dans le secteur de l'externalisation. Ces classes de trajectoires professionnelles correspondent aux jeunes travailleurs qui sont encore en formation ou qui viennent d'arrêter leurs études. Nos résultats soutiennent donc le fait que la trajectoire de ces travailleurs correspond à une bifurcation personnelle rendue nécessaire par les difficultés de financement des études. Pour pallier cette situation, ces jeunes entreprennent des formations profession-

nelles, soit pour rejoindre directement le secteur de l'externalisation, soit pour tenter d'émigrer en Europe. Les travailleurs de la troisième classe se distinguent par l'accès à des contrats formels, mais parfois dans des secteurs dont les conditions de travail sont très difficiles. On y retrouve également des travailleurs plus âgés qui ont des enfants ce qui nécessite une situation professionnelle et économique stable. Dans ce cas-là, la bifurcation personnelle s'effectue du fait de mauvaises conditions de travail dans d'autres secteurs que l'externalisation ainsi que par une pression liée à l'autonomie de leur foyer. Cette situation les pousse à rejoindre le secteur de l'externalisation, leur permettant d'améliorer leur stabilité financière, en comparaison avec d'autres emplois peu rémunérateurs, du secteur public (par exemple : professeur) ou du secteur privé informel (par exemple : vendeur). Enfin, la trajectoire des travailleurs de la classe quatre marquée par des contrats informels, ce qui explique qu'ils sont encore davantage incités à chercher une forme de stabilisation financière dans le secteur de l'externalisation.

Ces résultats montrent donc que les secteurs de l'externalisation et de l'annotation remplissent un vide structurel lié au marché du travail malgache fait d'activités informelles, de périodes d'inactivité et de contrats instables hors des entreprises formelles de l'externalisation. Ces dernières permettent donc une forme de stabilisation professionnelle et économique, qui est cependant nuancée par la dimension extractive du secteur et par les difficultés structurelles de Madagascar. Finalement, quelle que soit la situation des travailleurs et leurs raisons de s'orienter vers le secteur de l'externalisation, ils y voient surtout le moyen de "survivre" en attendant un "mieux" qu'ils ne sont pas sûrs d'atteindre.

Survivre en attendant mieux : un processus d'assétisation et de développement professionnel limité par la structure extractive de l'industrie

La littérature sur les effets du micro-travail sur le développement a mis en lumière les bénéfices de cette activité en matière d'amélioration des conditions de vie (Berg et al., 2018). Nous avons traité de cette perspective de recherche en analysant les conditions de travail objectives et subjectives avec comme ambition de contribuer à la littérature sur les effets des chaînes de valeur globales en matière de développement.

Le constat fait par certains travailleurs d'une stabilisation financière obtenue dans les entreprises formelles de l'externalisation, ne résiste pas à nos éléments empiriques. L'analyse du Code du travail malgache fait ressortir les nombreuses possibilités de licencier les travailleurs, surtout dans le contexte d'un secteur qui dépend structurellement des remous du marché français. Nous avons également pu constater que certaines entreprises formelles et informelles profitent de l'existence de certains statuts ("*zones grises de l'emploi*") pour rendre les contrats de travail plus flexibles, ce qui renforce encore la précarité des travailleurs. Certaines entreprises formelles mettent d'ailleurs en place des services de prototypage afin de tester les projets de sous-traitance pour évaluer leur faisabilité et leur pérennité. Ce faisant, la chaîne est largement pilotée par les entreprises d'IA, leurs propres donneurs d'ordres, ainsi que par les possibilités de financement de

l'écosystème français de start-up.

Ces éléments expliquent également les différences constatées entre entreprises de sous-traitance. Les entreprises formelles sont plus structurées et surtout très largement détenues par des entrepreneurs français, tandis que les entreprises informelles étudiées appartiennent plutôt à des chefs d'entreprises malgaches. Les risques pour les travailleurs sont paradoxalement bien plus grands au sein de cette dernière catégorie du fait de la nationalité du chef d'entreprise. En effet, seul le réseau extérieur à Madagascar des entrepreneurs de la sous-traitance peut finalement limiter les risques en garantissant un accès à des donneurs d'ordres étrangers. Les différences entre ces entreprises correspondent également différents modes "d'intégration" dans le marché du travail pour les annotateurs malgaches.

Nom	Type entreprise	Stabilité perçue de l'emploi	Satisfaction au travail	Niveau de salaire
Intégration incertaine	Moyenne, informelle	Faible	Moyenne	Faible
Intégration assurée	Grande, formelle	Moyenne	Forte	Fort
Intégration laborieuse	Moyenne, formelle	Forte	Faible	Faible

TABLE 33 – Trois intégrations des travailleurs des données dans le marché du travail malgache

Pour les travailleurs malgaches, atteindre la stabilité financière par l'intégration dans le secteur privé de l'externalisation, passe donc par l'acceptation d'un fort risque de licenciement. Cela justifie également de développer des activités secondaires qui constituent autant une stratégie de développement professionnel qu'une manière de gérer la pérennité financière de leur foyer, notamment pour la *classe moyenne urbaine stabilisée*. Entre un travail "d'à-côté" (Weber, 2009), et le travail entrepreneurial reproductif (Wilson & Yochim, 2017), ces activités constituent donc des éléments essentiels du bout de la chaîne de production l'IA. Que ce soit pour envisager un autre futur professionnel ou bien tout simplement pour gérer la stabilité financière du foyer, ces activités sont autant de moyens de gérer la précarité de leur emploi.

Le développement d'une chaîne locale de précarité : production de valeur mondiale, mais précarité économique et professionnelle locale

Tout au long de cette partie, nous avons vu que le développement de l'industrie de l'IA est facilité par l'offre pléthorique de travailleurs (Graham et al., 2017a), ainsi que par l'informalité du reste de l'économie malgache. En restituant la part de subjectivité des travailleurs concernant leurs revenus, ainsi que leurs stratégies de gestion de la "précarité", nous contribuons à montrer que l'existence des chaînes de valeur globales est consubstantiel à la mise en place de chaînes de précarité locale (Selwyn, 2019).

Tout d’abord, en confrontant nos conclusions sur les conditions de travail au rapport des annotateurs à leur salaire, nous avons précisé nos résultats concernant le vécu des travailleurs. Si certains connaissent une *intégration assurée* sur le marché du travail, car ils disposent de ressources sociales mobilisables pour trouver un emploi, leur revenu est perçu comme insuffisant s’ils ont des enfants : ils ont leur propre foyer avec des enfants. Nous les avons qualifiés de *”classe moyenne sous pression”*, car ils subissent directement le coût de la vie dans la capitale malgache. Pour d’autres, l’intégration sur le marché du travail est *incertaine*, car ils ont un revenu qu’ils considèrent comme suffisant, habitant encore avec leurs parents. Cette caractéristique leur permet de se projeter davantage dans la construction de leur parcours professionnel (*classe moyenne en transition incertaine*), tout en sachant qu’ils ont du mal à financer leurs études et qu’ils risquent de se trouver limités par la structure du marché de l’emploi à Madagascar. Nous avons parlé d’*intégration laborieuse* pour d’autres, car ils considèrent que leur revenu est insuffisant pour construire une vie de couple et éventuellement une vie de famille ce qui les conduit davantage être insatisfaits de leur expérience professionnelle comme annotateurs (*petite précarité urbaine*).

Nom	Type d’intégration	Perception des revenus	Foyer	Satisfaction au travail
Classe moyenne sous pression	Assurée	Mauvaise	Autonome avec enfant	Bonne
Classe moyenne en transition incertaine	Incertaine	Bonne	Dépendance aux parents	Moyenne
Petite précarité urbaine	Laborieuse	Mauvaise	Autonome, sans enfant	Faible

TABLE 34 – Trois intégrations des travailleurs des données dans le marché du travail

Ces résultats permettent ainsi de réinterpréter nos conclusions sur l’émergence d’une classe moyenne urbaine. Si la littérature identifie la volonté de stabilité comme une caractéristique de la classe moyenne dans les pays du Sud global, nous montrons à quel prix cette stabilité financière s’obtient à Madagascar : activités multiples, travail reproductif et activité principale au futur incertaine sont le prix de la stabilité du quotidien. Par rapport aux *”petites prospérités”* agricoles étudiées par (Andrianampiarivo, 2018), qui détiennent leur moyen de production et accumulent du capital, cette classe moyenne urbaine se caractérise par une forte précarité liée à la structure extractive du secteur de l’IA : conditions de travail peu protectrices et incapacité d’accumuler un capital leur permettant de développer leur propre activité.

En reprenant les mots de Ursula Huws (2014) pour qui la précarité se traduit par *”le besoin de subsistance”*, les conditions de travail, ainsi que la *”plus-value expropriée des résultats de ce travail”*, notre travail montre les

différentes facettes de la précarité du secteur de l'IA. En tant qu'acteurs prépondérant de la fabrication des modèles d'IA, les travailleurs malgaches sont cependant très peu valorisés financièrement. Dans le même temps, les entrepreneurs de la sous-traitance, les entreprises d'IA, et les grandes entreprises du numérique qui vendent l'accès à des infrastructures profitent de ce travail pour accumuler de la plus-value. La principale conséquence de cette chaîne de valeur globale est donc de limiter les perspectives d'avenir des travailleurs malgaches, ce qui en limite alors les impacts positifs sur le développement économique et humain du pays.

Enfin, nos données ne permettent pas d'affirmer que cette activité constitue un vecteur de perspectives professionnelles dans d'autres secteurs et/ou dans l'entrepreneuriat. Bien au contraire, la "démographie professionnelle" des entreprises de sous-traitance limite la mobilité ascendante des travailleurs malgaches, les postes les plus valorisés étant majoritairement occupés par des français. La création d'entreprises locales est aussi entravée par une double inégalité subie par les travailleurs qui est à la fois financière et raciale, soulignant à nouveau l'impact persistant de la structure raciale et post-colonial du secteur sur les travailleurs. Pourtant, nos données quantitatives et nos entretiens démontrent très largement les qualifications des travailleurs malgaches. Face à leur précarité, le désir d'entreprendre illustre à la fois ces compétences et la résilience face à des problèmes structurels qu'ils vivent au quotidien.

Conclusion générale

Au moment où nous écrivons ces lignes, l'intelligence artificielle est plus que jamais un objet majeur des débats publics contemporains, quelles que soient ses applications. Suscitant aussi bien les craintes que les espoirs, cette technologie est la cible de vives controverses et par conséquent la source d'intenses discussions quant à sa régulation. Pour déplier cet objet de recherche, notre thèse s'est proposée d'articuler l'analyse ce qu'est et de ce que fait l'IA avec la compréhension de son processus de production.

La vision de l'ensemble de la chaîne de production de l'intelligence artificielle est un résultat en soi de notre travail. Notre matériau empirique permet en effet de réinscrire pleinement l'IA dans l'assemblage des décisions humaines qui guident son développement. Notre méthode d'enquête, basée sur l'analyse d'une pluralité d'acteurs, ainsi que sur la prise en compte de leur *situation sociale et géographique* permet de mieux comprendre ce qui se cache derrière les modèles d'IA déployés dans la société (Christin, 2017; Seaver, 2017). La portée de nos résultats démontre le besoin de structuration d'un discours critique associant la compréhension technique de l'IA avec l'examen des conditions sociales de production de cette technologie.

L'IA comme économie de la qualité. La mise en place de boucles humaines de retro-action pour tenir compte de la volatilité des modèles d'IA

L'ambition de cette thèse était de contribuer à la littérature émergente en sociologie de l'IA en analysant les processus d'annotations dans les entreprises d'IA françaises. Nous avons identifié le caractère *incrémental* du développement de ses modèles et observé les conséquences en termes d'actualisation permanente de la "vérité de terrain" (*ground truth*) (Jaton, 2019) et donc des données. La construction des bases de données repose en effet sur la proximité entre ce que le *modèle fait*, la manière dont il est *problématisé* et la *réalité de ce qui est automatisé*. La forte attention portée à la qualité des données, pour éviter de créer des biais nuisant au bon fonctionnement des modèles, constitue une variable importante de leurs processus de développement. La volatilité des modèles, en fonction de l'évolution des décisions de développement, des remontées des annotateurs sur les données et des évolutions de la vérité terrain, a également posé la question de l'organisation mise en place pour en maintenir la cohérence.

Au cours du travail de catégorisation, un premier enjeu porte sur la nécessité de maîtriser la subjectivité au cours d'un processus de développement devant permettre la traduction du contexte du client ou de l'utilisateur final dans un outil d'automatisation. La chaîne de valeur rassemble donc différents groupes de travailleurs liées par l'objectif de "*maintien de la cohérence*" du résultat : le client final possède la connaissance métier, l'entreprise d'IA celle de la modélisation, et les annotateurs une expertise intermédiaire. Ces derniers développent en effet à la fois une connaissance heuristique des données et une maîtrise des enjeux

de l'annotation. Ce qui se joue dans la chaîne de valeur s'apparente donc à un travail de standardisation de l'IA à travers les échanges sur les données et sur le modèle. Ce processus s'appuie nécessairement sur des coopérations durables au sein d'une communauté de pratiques. Dans ce contexte, nos résultats illustrent le rôle des travailleurs des données comme essentiel à la *stabilisation des modèles d'IA*. Nous montrons que pour maîtriser leurs modèles, les entreprises intègrent le plus possible les travailleurs des données à leurs processus de production à travers différents outils de communication et de contrôle. L'organisation du travail choisie par les entreprises d'IA s'apparente en ce sens à un *double couplage*, d'une part *entre les données et les modèles* et *entre les développeurs et les annotateurs*, d'autre part.

Nos résultats suggèrent également que le marché des données annotées pour l'intelligence artificielle se structure en fonction du prix, mais également en fonction des jugements sur la qualité des données réelle ou perçue. Ce facteur, considéré comme déterminant, dans la forme finale du modèle justifie le recours par les entreprises d'IA au modèle de la sous-traitance directe plutôt qu'à des plateformes numériques de travail. Ainsi, cet agencement du secteur est à la fois poussé par des impératifs de *qualité de la prédiction finale* et par une *éthique de la qualité* lié aux risques de biais.

Une implication directe de ces résultats porte sur la difficulté à dévoiler "*l'opacité organisationnelle*" de la chaîne de valeur de l'IA. Il paraît en effet très compliqué de véritablement auditer l'intégralité de la chaîne de production sans documenter l'ensemble des décisions humaines : comment avoir accès aux modèles et architectures pré-entraînés qui sont développées par les grandes entreprises du numérique et utilisées par les entreprises d'IA françaises ? Comment accéder aux décisions prises lors des phases d'annotations ? Comment documenter ces décisions, qu'elles soient adoptées lors de la création des modèles pré-entraînés ou bien lors de la création des modèles finaux ? Ce faisant, notre travail contribue encore à la littérature en sociologie de l'IA en identifiant plusieurs éléments *d'opacités non-techniques* liés au mode de production de l'IA. En premier lieu, la "*volatilité algorithmique*" ne se matérialise pas uniquement par l'architecture propre au *machine learning*, mais aussi par l'implication de nombreux acteurs humains qui façonnent le modèle final. Cela induit ensuite trois formes d'opacité : la *volatilité des décisions algorithmiques*, la *volatilité des données existantes* et la *volatilité de la vérité terrain*. La conséquence de ce mode de fonctionnement de l'IA est qu'il paraît peu probable de produire des modèles non biaisés tant ils sont le résultat d'une chaîne de production impliquant de nombreuses couches de travail humain (F. A. Schmidt, 2019).

Ce résultat est fondamental au regard des tentatives de régulation de l'intelligence artificielle comme l'*AI Act*. L'article 10 du projet de texte législatif européen prévoit en effet que les systèmes dits à "*haut risques*"³²⁹ répondent à un certain nombre de critères de qualité des données : représentativité, « conception parti-

329. Sont couverts certains produits faisant déjà l'objet de réglementation de l'UE, ainsi que des systèmes ayant « un risque de préjudice pour la santé et la sécurité ou un risque d'incidence négative sur les droits fondamentaux » et des systèmes opérant dans plusieurs secteurs comme l'éducation, le travail... (Art. 6).

nente », « *opérations de traitement pertinents pour la préparation des données, telles que l'annotation* »... Les acteurs devront également mettre en place un « *système de gestion de la qualité* » et détailler le fonctionnement des systèmes d'IA, à travers un mécanisme de publication d'une documentation technique comportant plusieurs points (Article 11.). L'annexe 4 du règlement présente les points devant figurer dans cette documentation. Elle précise que pourra être publié : *"les spécifications de conception du système, à savoir la logique générale du système d'IA et des algorithmes; les principaux choix de conception, y compris le raisonnement et les hypothèses retenues (...) les principaux choix de classification"*. Concernant les données, plusieurs éléments devront également y figurer tels que : *les fiches décrivant les méthodes et techniques d'entraînement et les jeux de données d'entraînement utilisés, y compris des informations sur la provenance de ces jeux de données, leur portée et leurs principales caractéristiques; la manière dont les données ont été obtenues et sélectionnées; les procédures d'étiquetage (par exemple pour l'apprentissage supervisé), les méthodes de nettoyage des données (par exemple la détection des valeurs aberrantes)* ».

Nos résultats permettent d'anticiper la naissance d'une tension entre l'organisation de la chaîne de production et les impératifs de documentation liés aux futures réglementations du secteur. Nos données révèlent en effet les nombreux échanges formels et informels qui ont lieu entre les acteurs pendant le processus de production des modèles d'IA et qui en façonnent le fonctionnement. Comment dès lors appliquer l'*AI Act*? Comment les entreprises pourraient-elles documenter la profondeur des décisions et des échanges informels qui contribuent à l'élaboration des systèmes d'IA? Comment rendre compte des choix de conception effectués dans ces "boîtes noires" organisationnelles? À cet égard, il serait pertinent que les futurs travaux réglementaires prennent en compte les futures perspectives de recherche sur le développement d'une méthodologie d'audit des processus de production des données afin de comprendre ces décisions et à quel niveau de la chaîne elles ont été prises³³⁰. Une telle méthodologie pourrait notamment inclure la documentation des échanges entre concepteurs de systèmes et annotateurs à travers :

- la publication de guides d'annotation *versionnée* pour tenir compte des évolutions du processus d'annotation
- la présentation des choix de classification incluant les décisions concernant les données ambiguës
- un descriptif des modalités d'échanges entre les concepteurs et les annotateurs et par conséquent les modalités d'arbitrage des données ambiguës
- la description de l'organisation et des conditions de travail (par exemple : le temps nécessaire pour réaliser l'annotation)

Par ailleurs, il serait pertinent d'encourager les entreprises à mieux prendre en compte et à rendre transparentes les conditions de travail et les impacts négatifs du développement de l'IA, par exemple, en étendant le modèle de la *"Directive relative au devoir de vigilance des entreprises en matière de développement durable"*.

330. Par exemple, au travers d'actes d'exécutions de l'*AI Act* indiquant la granularité des informations relatives au travail des données censés apparaître sur cette documentation

Il est également possible de les inciter à documenter et à rendre transparentes les activités de leurs sous-traitants en matière de données. À cette fin, sur le modèle du délégué à la protection des données (rôle créé par le règlement général sur la protection des données de l'UE - RGPD), il pourrait également être utile de créer une fonction de gestionnaire des données d'IA (*data stewardship*) qui endosserait ce rôle de documentation des décisions.

Des firmes hégémoniques pour gouverner la chaîne de valeur. Analyser les relations entre les acteurs qui fabriquent l'IA

Cette approche par la chaîne de production nous a également donné l'occasion de questionner les modèles organisationnels autour de l'IA et notamment les plateformes numériques de travail (PNT), très souvent étudiées par la littérature. Plus spécifiquement, nous avons analysé la manière dont les entreprises produisent les données permettant de façonner les modèles d'IA. L'analyse de cette technologie s'articule donc nécessairement avec l'examen des organisations qui la produise.

Il était alors nécessaire de repartir de la position hégémonique de ces *lead firms* pour comprendre l'IA et les modèles organisationnels qui soutiennent son développement. Notre analyse de la position subalterne des entreprises françaises dans la chaîne de valeur tient compte de cette caractéristique du secteur et illustre la division des activités productives entre les acteurs de la chaîne. En tant que firmes focales, l'avantage principale des grands acteurs technologiques est davantage lié à leur capacité à proposer l'ensemble des services nécessaires au développement de l'IA (*intégration verticale*) qu'à un comportement de plateforme capable de connecter plusieurs faces d'un marché. En termes de localisation de la chaîne de valeur, nos résultats permettent également d'identifier d'autres flux géographiques que ceux mis en avant dans la littérature. Ces flux d'échanges intègrent à la fois des entreprises de la Silicon Valley, parfois parmi les plus importantes, des entreprises françaises d'IA (start-up et grand groupes) et des entreprises d'annotations situées à Madagascar.

D'un point de vue empirique, nos résultats paraissent d'autant plus pertinents tant les liens entre les États-Unis et l'Inde ou les Philippines en matière de travail de la donnée ont été examinés par la littérature, par contraste avec les pays du continent africain (Anwar & Graham, 2019). Nous montrons par ailleurs qu'il existe plusieurs autres formes d'organisations du travail que les plateformes numériques de travail. Surtout, nos résultats illustrent la capacité des firmes hégémoniques (*lead firms*) à maîtriser l'ensemble de la chaîne en concentrant des actifs spécifiques. L'infrastructure matérielle permet de proposer l'hébergement de données et de la puissance de calcul. Quant à l'infrastructure logicielle, elle consiste à vendre des modèles pré-entraînés, des logiciels d'annotations et l'accès à des services d'annotations. Ces services sont ensuite vendus aux start-up françaises qui font de l'intégration de modèle en réentraînant ces derniers sur les cas

d'usages de leurs clients (*fine tuning*), en faisant appel à des sous-traitants situés à Madagascar. Dès lors, le déséquilibre des pouvoirs se joue sur la dépendance des entreprises de sous-traitance vis-à-vis des donneurs d'ordres français, et sur la dépendance des entreprises françaises vis-à-vis des grands acteurs américains. *In fine*, de la *French Tech* et de la *Silicon Valley*, les remous se font ressentir avec puissance jusqu'à Antananarivo. Dès lors, comment peut-on parler de souveraineté numérique, alors même que les entreprises françaises se basent d'un côté sur les services des géants américains et de l'autre sur les travailleurs de leur ancien empire colonial ?

La pertinence de la distinction entre l'organisation par la forme hiérarchique et par le marché. Des firmes-plateformes plus proche d'une *hiérarchie électronique étendue* que d'un marché pur

La posture théorique évoquée précédemment à propos du concept de plateforme ne vient pas de nulle part et il nous faut faire un détour par des explications plus personnelles. La question du statut des travailleurs de plateformes a en effet été central dans mon travail de rapporteur au Conseil national du numérique qui donnera lieu à un rapport de politique publique (Régniez et al., 2020). J'ai pu constater le flou conceptuel et juridique entourant cet objet d'étude. Je tire de cette expérience la difficulté d'envisager des alternatives plus protectrices pour ces travailleurs, comme le salariat, tant l'imaginaire de l'indépendance résiste dans la tête de nombreux acteurs. Il en résulte une interrogation : ces entreprises peuvent-elles être réellement considérées comme des plateformes ?

Cela m'a conduit à travailler autour d'un pari théorique développé lors de notre deuxième partie et qui postule de l'inadéquation de l'approche conceptuelle des PNT avec la réalité des organisations platformisées dans le travail des données et plus généralement de l'incapacité du concept de plateformes à saisir la variété des modèles organisationnels présents dans l'économie numérique entre les plateformes comme marché et les plateformes comme entreprise. En effet, si l'on doit revenir aux origines du concept, c'est la capacité des ces entreprises à connecter plusieurs faces du marché à l'aide d'une organisation basée sur la technologie qui constitue la particularité de ces firmes. Que dire alors de l'arrêt de la Cour de justice de l'Union européenne (CJUE) (*Uber v. Elite Taxi* en 2017) qui "qualifie la plateforme Uber comme participant à un service « global » comprenant à la fois une activité d'intermédiation technologique et une activité d'opérateur de service sous-jacent" (Régniez et al., 2020) ? Parler de plateformes ne revient-il donc pas à reprendre le narratif de ces entreprises consistant à se présenter comme des opérateurs d'intermédiation technologique plutôt que comme des entreprises fournissant un service de transport ?

De fait, le concept de plateforme ne raconte que peu de choses de la réalité empirique qui se cache derrière le narratif mis en place par ces firmes, aussi bien au niveau de l'organisation du travail que du vécu des

travailleurs. La participation à ce travail de régulation m'a permis, en plus d'identifier les limites de ce concept, d'entrevoir la nécessité de travailler sur les parcours professionnels des travailleurs des plateformes afin de mieux comprendre comment cette activité s'insérait dans leurs trajectoires et pouvait *in fine* influencer leur manière de percevoir le salariat. Les trajectoires des travailleurs nous donnent donc autant d'indications de ce qu'ils vivent dans des emplois précaires ou platformisés que l'examen de leurs conditions de travail. Lors de la rédaction de ce rapport, les entretiens avec les chauffeurs VTC m'ont incité à davantage considérer cette variable. Je pense notamment à cet échange avec un chauffeur Uber, autrefois salarié comme bagagiste pour l'entreprise aéroport de Paris. S'il exprime sa volonté de redevenir salarié, il considère la flexibilité permise par cette activité comme une amélioration de ses conditions de travail. Lui qui devait jusqu'alors se lever à cinq heures du matin pour se rendre à l'aéroport d'Orly et trier des bagages, commence désormais sa journée à six heures du matin pour aller chercher des clients dans ce même aéroport, mais il a le choix de le faire ou non. Cette expérience m'a laissé une frustration de ne pas avoir pu lier suffisamment les trajectoires professionnelles avec le fait de rejoindre les plateformes. J'ai également pu saisir la variété des plateformes et la nécessité de ne pas analyser d'un seul bloc des modalités d'organisation du travail variées dans des secteurs bien différents.

Ce positionnement a également orienté notre deuxième partie dans laquelle nous avons analysé le processus de production de l'IA au sein des entreprises françaises. Nous avons constaté son caractère *incrémental*, avec des allers-retours réguliers entre les différents acteurs de la chaîne de valeur, justifiant le recours à une force de travail beaucoup plus stable que la main-d'oeuvre platformisée. Nous avons également proposé une analyse des choix d'organisation du travail de la donnée de ces entreprises, au-delà de la dimension *dynamique* de la production des modèles basés sur la *machine learning*.

En repartant des théories de la firme et des chaînes de valeurs, nous avons analysé la manière dont les travailleurs malgaches sont intégrés au processus de production des donneurs d'ordres finaux, à tel point que certains d'entre eux ont tendance à les considérer comme leurs propres employés. Nos résultats démontrent la pertinence du concept *d'intégration* des travailleurs dans des réseaux de sous-traitance, ce que la littérature a analysé comme une caractéristique des firme-réseaux, objets considérés comme hybrides entre la firme et le marché (Baudry & Chassagnon, 2012). Comme nous l'avons vu au cours de cette thèse, la délimitation des frontières des firmes, ainsi que la compréhension de leurs effets sur les travailleurs constituent des questions fondamentales de la littérature et de la régulation lorsque cela touche à la platformisation du travail.

Notre étude montre pourtant que les différentes organisations du travail de la donnée se caractérisent plutôt par le développement de formes de *"hiérarchie électronique étendue"* (Fréry, 1997, Malone, 1987). La numérisation des relations intra-firmes permet d'intégrer des outils supplémentaires de contrôle des travailleurs qui bénéficient à la firme focale. Cette dernière peut alors se comporter comme si les travailleurs

étaient ses propres employés. La mise en place de dispositifs de communication directs entre entreprise d'IA, entreprises de sous-traitance et travailleurs créent les conditions pour que s'établissent des échanges sur la conception du travail des données, mais également sur les concepts mobilisés par les développeurs dans leurs modèles.

Ce résultat prolonge notre argument concernant l'hétérogénéité des plateformes et des modes d'externalisation dans l'IA (Cornet et al., 2022). De notre point de vue, cet intérêt pour ce nouvel "objet", en tant que modèle typique d'organisation masque le fait que les plateformes ne constituent pas un groupe homogène d'organisations, que ce soit en termes de secteur d'activité, de définition du statut juridique et plus généralement de structures organisationnelles. Nous proposons une analyse des relations entre les acteurs de la chaîne qui montre que, malgré des formes d'organisations qui se présentent comme relevant du marché, les relations de pouvoirs entre les acteurs pointent plutôt vers des relations "hiérarchiques", dont le degré diffère selon les modèles d'organisation. Plutôt qu'une distinction entre les organisations, nous montrons qu'il s'agit d'un enchevêtrement, avec une primauté des modèles "hiérarchiques".

Les facteurs évoqués dans le contexte français mettent en lumière des éléments d'explications. Pour tenir compte de la vérité de terrain et créer des modèles applicables dans le monde réel, il est nécessaire de mettre en place des outils de communication et d'avoir accès à une main-d'œuvre stable. En effet, nous constatons également la présence d'entreprises formelles qui sont intégrées à ces plateformes (Appen) et de travailleurs indépendants fortement intégrés dans le processus de production de la plateforme (AIEthic). D'un point de vue purement empirique, cela confirme en apparence les analyses de Tubaro (2021) et A. Casilli et al. (2019a) suggérant l'existence de plateformes intermédiaires entre la hiérarchie et le marché et de mécanismes de coordination du travail partagés entre plusieurs entreprises.

Cependant, cela confirme surtout que ce phénomène de fond ne peut pas uniquement être analysé au prisme de l'émergence d'un "marché mondial du travail" (Anwar et al., 2023). Nous y voyons plutôt une extension numérique de la capacité des entreprises occidentales à standardiser leurs opérations. De ce point de vue, nous pensons qu'il ne faut pas voir des formes de marchés et donc des plateformes partout, mais davantage une variété de formes d'intégration des travailleurs entre la hiérarchie et le marché ce qui souligne la pertinence de cette distinction. L'analyse de nos résultats questionne néanmoins ces formes hybrides, notamment si l'on tient compte de la structure du pouvoir au sein des entités présentées comme relevant du marché. L'étude des relations entre acteurs à l'intérieur des firmes de la chaîne de valeur montre que même dans les cas qui paraissent se rapprocher de la forme du marché, les entreprises intègrent les travailleurs. Ces firmes font coexister trois types de pouvoirs : juridique de par la nature du contrat, *de jure* de par la propriété des actifs de la firme et *de facto* de par la dépendance des acteurs vis-à-vis de ressources critiques, notamment les ressources humaines (Chassagnon, 2019). Nous avons constaté que les plateformes numériques de travail sont soit de simples outils d'annotations soit de réels outils de coordination du travail.

Dans le premier cas, elles délèguent donc le pouvoir juridique et *de jure* aux entreprises de sous-traitance qui partagent le pouvoir *de facto* avec le donneur d'ordre final. Dans le second, elles détiennent le pouvoir juridique et *de facto* et partagent la propriété des actifs avec les travailleurs, ces derniers possédant leur outil de production (un ordinateur) et la plateforme, l'outil d'annotation.

Ces résultats limitent donc la pertinence du concept de *plateforme*, puisque le terme désigne une organisation du travail qui fonctionne en reliant l'offre et la demande de travail, ou du moins en circonscrivent plus précisément la portée. Quelle que soit leur position dans la chaîne de valeur, les plateformes fonctionnent davantage comme des "*catalyseurs de hiérarchie*" contribuant à affirmer le pouvoir qui s'exerce sur les travailleurs de la part des donneurs d'ordres finaux ou bien de la plateforme elle-même. Ces éléments nous conduisent à faire deux propositions théoriques concernant les études sur les plateformes, sur les chaînes de valeur et plus généralement sur l'émergence de formes hybrides de firmes : l'existence de "*réseaux hiérarchiques étendu*" et l'application d'un pouvoir hiérarchique "*de facto*" basé sur les outils de coordination numérique et sur la structure post-coloniale de la chaîne de valeur. Par conséquent, nous considérons la forme purement marchande comme une exception, plutôt que comme une tendance de la plateformes. Évidemment, cela ne veut pas dire que nous n'acceptons pas l'ambiguïté organisationnelle que la création de ces organisations de travail provoque. Néanmoins, il nous semble tout de même nécessaire de reconsidérer l'approche conceptuelle de ces sujets afin de mieux considérer l'hétérogénéité de l'organisation du travail numérique (Cornet et al., 2022), et les modes d'intégration des salariés dans des formes hiérarchiques, qu'elles soient managées par le client final, par la plateforme ou par une société sous-traitante intégrée à la plateforme.

Ces résultats ne sont pas sans conséquences sur la manière d'envisager les entreprises hybrides et plus spécifiquement les plateformes. D'un point de vue purement théorique, nous pensons qu'il est nécessaire de maintenir la distinction entre le marché et la hiérarchie, malgré des analyses qui considèrent certaines entreprises comme hybrides. Le fait même de réutiliser le narratif de ces entreprises, que ce soit pour parler de "*capitalisme de plateforme*" ou pour proposer des régulations spécifiques aux travailleurs de ces entreprises, a des incidences juridiques qui place ces derniers dans une situation intermédiaire les conduisant à vivre le pire des deux mondes entre le salariat et le travail indépendant. On pense notamment au *tiers statut* mis en place pour les travailleurs des plateformes français qui se veut une manière de tenir compte des spécificités de ce mode de distribution du travail, tout en améliorant, à la marge seulement leur protection sociale. Quelle que soit la position de ces *faux intermédiaires* dans la chaîne de valeur, ils disposent d'un pouvoir *de facto* du fait du contrôle d'accès qu'ils exercent, d'un côté sur les clients finaux et de l'autre sur les travailleurs qui passent par l'utilisation d'outils numériques de communication (pouvoir *de jure*) permettant de mettre en place un pouvoir managérial sur les travailleurs du bout de la chaîne (pouvoir *juridique*). Dans le cas de nombreuses plateformes présentes à Madagascar, le cumul de ces trois formes de pouvoir

s'accompagne d'une subordination économique des travailleurs qui facilitent encore davantage l'exercice de ces pouvoirs. En fin de compte, nous considérons qu'il n'existe pas de plateformes, mais seulement des outils qui permettent d'établir une *forme hiérarchique électronique étendue* qui doit donc être traitée comme telle par les régulateurs et la recherche.

Cette caractérisation du modèle plateforme rejoint également ce que nous avons étudié dans le cas des relations de sous-traitance directe et n'est pas sans conséquences juridiques. En effet, si un lien de subordination est démontré, les plateformes agiront dans ce que l'ordre juridique français appelle le *"prêt de main d'oeuvre à titre lucratif"*. Hors des exceptions autorisées, il est strictement interdit et se traduit en deux formes susceptibles de sanctions civiles et pénales : le prêt illicite de main d'oeuvre et le marchandage dont la particularité est le préjudice causé au salarié (notamment par l'éviction des avantages du salariat). Plusieurs indices qui constituent des preuves de ce délit : subordination des travailleurs vis-à-vis du donneur d'ordres, absence de spécialisation de l'entreprise sous-traitante, capacité du donneur d'ordres à réaliser la prestation et précision de la mission. L'intégration des travailleurs à travers les outils numériques et la subordination qui en découle, la conception des donneur d'ordres d'un travail que *"tout le monde peut faire"*, la méconnaissance du savoir-faire des travailleurs malgaches constituent autant d'indices qui pointent vers une mauvaise qualification de ces relations de travail.

La recherche et les régulateurs du secteur pourraient donc emprunter cette voie pour améliorer les conditions des travailleurs. En effet, s'il est possible de vérifier juridiquement qu'il s'agit d'une fausse sous-traitance, il faudrait que les entreprises d'externalisation y remédie par une meilleure protection des travailleurs contre les demandes des donneur d'ordres en matière de gestion des ressources humaines, par la reconnaissance de la spécificité de leur savoir-faire, ainsi que par la mise en valeur de leur rôle essentiel dans la fabrication des modèles d'IA.

Quand le travail des données devient un travail de l'IA. Des travailleurs qualifiés délocalisés pour stabiliser les processus automatisés

Notre perspective de travail centrée sur l'étude de l'industrie de l'IA comme une économie de la qualité, nous a ouvert la voie pour interroger la réalité du travail des données, souvent évoqué pour les conditions de travail, mais peu analysé dans la pratique. Nous avons développé cet aspect dans une troisième partie dans laquelle nous avons pris un autre pari théorique : celui d'analyser le travail des données et des activités numériques comme constitutifs des services numériques "automatisés" proposés par les entreprises occidentales. Face au manque de données empiriques soutenant les publications prospectivistes des économistes sur la fin du travail ou sur sa déqualification, il nous semblait important de réinscrire le travail réalisé depuis Madagascar en tant que témoin et conséquence des projets d'automatisation basés sur les technologies numériques. Les transformations du travail dont nous rendons compte concernent tout autant

les travailleurs occidentaux censés perdre leur emploi, que ceux qui collaborent de manière invisible depuis les pays du Sud global.

Notre travail permet d'observer l'existence d'un "*travail algorithmique supplétif*" à l'échelle de toute une chaîne de valeur entre la France et Madagascar. Qu'il soit central ou périphérique dans le fonctionnement du service algorithmique, il participe activement de ce dernier confirmant l'hypothèse de la continuité d'un travail humain lors des phases d'automatisations algorithmique déjà décrite par [H. R. Ekbia et Nardi \(2017\)](#) et [\(SHESTAKOVSKY, 2015\)](#). Ces chaînes de production se traduisent par une tension entre la qualité et le volume des tâches traitées par les travailleurs. Ce phénomène, déjà discuté dans la littérature sur les centres d'appels a des conséquences qui demeurent controversées sur la qualification des travailleurs : certains considèrent que ces derniers peuvent se réapproprier l'activité en sortant du cadre formulé par les clients finaux, tandis que d'autres enquêtes insistent sur le contrôle exercé sur les travailleurs à travers la mesure de la qualité et de la quantité. Notre travail montre que ces activités *supplétives* nécessitent des compétences plus étendues que ce que les discours des donneurs d'ordres laissent parfois entendre. À Madagascar, ces emplois se traduisent en effet par un *travail relationnel* qui se double d'un *travail culturel*, d'interprétation des tâches à réaliser.

Sur cette base, nous avons analysé les pratiques d'annotations au regard de la qualification de ce travail. Le point de départ de cette piste de recherche provient du constat de l'existence d'un malentendu de la littérature autour du *micro-travail*. Le terme même induit non seulement une faible rémunération et un travail peu qualifié (*micro*). Comme nous avons pu le mentionner, ce terme fait pourtant référence à la micro-finance et à des programmes d'insertion par le travail dans les pays du Sud global. Dès lors, si la rémunération est micro, le travail l'est-il aussi réellement ?

Nous constatons que les services mis en place par les entreprises d'IA étudiées ne répondent pas aux perspectives de généralisation de l'IA, mais plutôt à des cas d'usages très précis qui nécessitent donc des données représentant le mieux possible le monde réel. Les annotateurs effectuent des tâches diverses qui combinent la classification des données (*ce qu'ils voient dans les données*), et la problématisation des modèles (*ce qu'ils voient modifient la forme finale du modèle*) nécessitant un travail d'interprétation. Ce faisant, ce travail n'est pas micro, pas plus qu'il n'est uniquement focalisé sur les données : nous préférons donc parler de *travailleurs de l'IA*. Le développement de l'IA requiert ainsi la mise en place de nombreuses *boucles de rétroaction humaines*. Ainsi, plutôt que d'inventer des systèmes *sans travail*, ce qui est vendu et déployé en tant que solution d'IA n'est pas l'algorithme fini, mais le processus lui-même, y compris l'externalisation du travail humain aux deux extrémités de la chaîne de valeur : vers les consommateurs qui utilisent les systèmes d'IA et vers les travailleurs malgaches.

Un autre aspect de cette question porte sur les processus de professionnalisation des annotateurs. Ici aussi,

ma trajectoire personnelle semble à même d'éclairer l'attention que j'ai voulu porter à la question des qualifications. Issus de familles paysannes et ouvrières, mes parents n'ont pas eu l'opportunité d'obtenir beaucoup de diplômes. Leur trajectoire d'ascension professionnelle et personnelle s'est donc inscrite dans les possibilités de promotion interne offertes au sein de la Société nationale des chemins de fer français (SNCF). C'est donc au travers de la formation professionnelle et de la valorisation de compétences considérées comme subalternes qu'ils ont pu exercer comme cadres durant leur deuxième moitié de carrière. Cet ancrage m'a permis d'intégrer des questionnements liés non pas à l'invisibilité des travailleurs, mais à l'invisibilité de leurs qualifications. Bien connu des sociologues du travail issues d'une perspective féministe (Cockburn & Ploux, 2004; Maruani & Nicole-Drancourt, 1989; A. Soares, 2002), ces mécanismes de minimisation des qualifications s'expliquent par de nombreux facteurs, que nous avons essayé de mettre en lumière dans notre quatrième partie. Néanmoins, ce qui a longtemps été invisible aux yeux des chercheurs l'est beaucoup moins dans l'esprit et les discours des personnes qui emploient ces travailleurs. Ces qualifications sont-elles donc réellement invisibles? Et surtout, au regard de qui le sont-elles?

Pour approfondir cette dimension de notre travail, un détour par le concept de profession et la tension qui existe au regard de ce qu'on peut désigner comme "groupe professionnel" s'impose. En se centrant sur le contenu du travail, sur les activités qui lient des individus entre eux, la sociologie du travail a opéré un renversement théorique à travers le dépassement du concept de profession, jugé trop restrictif, par la notion de "groupes professionnels" qui sont *"des ensembles de travailleurs exerçant une activité ayant le même nom, et par conséquent dotés d'une visibilité sociale, bénéficiant d'une identification et d'une reconnaissance, occupant une place différenciée dans la division sociale du travail et caractérisés par une légitimité symbolique"* (Demazière et Gadéa, 2009, p.20). Un autre enjeu de notre analyse sur les qualifications réside donc dans l'analyse des logiques à l'oeuvre derrière la professionnalisation des annotateurs de données : comment se forment-ils? Comment formalisent-ils des savoirs? Et enfin, comme les entreprises de la chaîne de valeur formalisent les compétences nécessaires pour la réalisation des tâches?

La formation des qualifications dans les entreprises d'annotation de données s'inscrit dans plusieurs logiques de professionnalisation (Fernandes, 2009) qui démontrent l'étendue de ce que les annotateurs savent et de ce qu'ils doivent faire. La *"logique de l'action"* d'abord, car les annotateurs, confrontés à des situations nouvelles, ajustent leurs actions pour réaliser leurs tâches. La *"logique de la réflexion"* s'observe dans la mobilisation de ressources documentaires et de ressources sociales dans l'accomplissement d'une tâche, par exemple, faire des recherches sur Internet ou sur le guide d'annotation. La *"logique de traduction culturelle par rapport à l'action"* renvoie à l'intervention d'un tiers pour l'accompagnement des salariés dans l'accomplissement de leurs missions. De ce point de vue, le caractère socialisé de l'activité d'annotation et la présence de nombreuses couches hiérarchiques dans la chaîne de production constitue une preuve de cette logique de professionnalisation. Enfin, la *"logique de la réflexion sur l'action"* désigne les moments où l'in-

dividu prend de la distance avec son activité pour l'analyser et développer une connaissance qui sera utile par la suite de son activité tandis que la "logique de l'intégration assimilation" caractérise l'utilisation de ressources documentaires pour acquérir des savoirs ou des connaissances. Ces deux dernières dimensions sont beaucoup moins présentes dans l'activité d'annotation et sont davantage réservées aux *team leaders*, aux responsables de production et aux départements qualité et formations qui conçoivent et vérifient les tâches d'annotation pour qu'elles soient réalisables par les simples annotateurs. Par ailleurs, nos résultats montrent que ces logiques de *professionnalisation* relèvent à la fois de l'une intention des entreprises de sous-traitance et d'IA qui passe par la prescription des *bonnes qualifications*, à travers la formation des annotateurs à différents aspects des projets.

Ces éléments ne doivent pas masquer la tension autour de la reconnaissance des qualifications que nous avons abordé dans notre dernière partie. Dire que le travail de l'IA n'est pas un travail déqualifié au regard des pratiques ne doit pas oblitérer la réalité de la position des travailleurs dans la négociation de ces qualifications. S'il semble désormais clair que le corpus de connaissances et de savoirs nécessaires à cette activité témoigne d'un mouvement de *solidification d'un groupe professionnel*, la reconnaissance des qualifications des travailleurs des données demeure limitée par la structure extractive de la chaîne. Le fait de réintroduire cette caractéristique introduit d'ailleurs une variable interprétative importante : le contrôle managérial qui s'exerce sur des travailleurs situés dans les pays du Sud global en constitue le principal facteur limitant.

Des qualifications invisibles ? Le processus de reconnaissance / minimisation comme conséquence de la colonialité de la chaîne de valeur

L'invisibilisation des travailleurs des données a été expliquée par la littérature sur le *micro-travail* comme étant le résultat d'un mode d'organisation du travail platformisé, et la conséquence du jugement négatif qui s'applique aux tâches de collecte de données de la part des entrepreneurs de la tech (Irani, 2013). Dans un article, Star et Strauss (1999), considère que le travail invisible est un sujet profondément politique en se demandant : "qu'est-ce qui "comptera" comme travail productif, travail créatif, travail qui ne peut être externalisé ou remplacé dans la nouvelle société d'aujourd'hui ?" Pour de nombreux acteurs de la chaîne, le travail de l'IA est considéré comme répétitif, routinier et peu important au regard des objectifs finaux de la chaîne de production : automatiser un service. Paradoxalement, ils ont également été les premiers à en reconnaître l'importance par la production de l'IA dans le même mouvement. Nous interprétons cette dissonance comme le résultat d'un paradoxe entre les promesses de l'automatisation et ce que sa réalisation implique. La nécessité de collecter et de transformer des données à de multiples reprises au cours d'un processus de production *incrémental* constitue en ce sens la marque de l'incomplétude des modèles d'IA. Le travail supplétif réalisé à Madagascar constitue la preuve des grains de sables qui persistent dans les rouages des technologies automatiques.

Cependant, le travail de l'IA réalisé à Madagascar n'est pas invisible pour l'ensemble des acteurs de la chaîne. Nos données montrent qu'il est particulièrement visible pour les entrepreneurs de l'IA et les entreprises de sous-traitance. C'est auprès des utilisateurs finaux de ces systèmes qu'il demeure caché. La visibilité peut ainsi servir à contrôler la qualité des données et les travailleurs, mais pas à valoriser le travail des annotateurs.

Malgré les discours tenus par certains entrepreneurs dans le domaine de la sous-traitance, il est manifeste qu'ils attribuent une réelle valeur aux compétences linguistiques et culturelles héritées de l'époque coloniale. Cette appréciation est également mise en avant par les clients. La capacité des travailleurs malgaches à saisir les subtilités de la langue française, et parfois même les aspects culturels français, tout en ayant un niveau de français suffisamment élevé pour discuter d'éventuelles ambiguïtés liées aux tâches d'annotation, revêt un rôle essentiel dans la fabrique française de l'IA.

La notion de *"rhétorique de la fiabilité"* paraît particulièrement pertinente lorsqu'on l'applique à la mise en avant de compétences héritées de l'époque coloniale. Dans le cas des Philippines, (Padios, 2018) cette stratégie est utilisée par les entreprises de sous-traitance qui essaient d'essentialiser l'identité philippine pour obtenir un avantage concurrentiel sur le marché mondial, dans le contexte de la recherche d'investissements nationaux et de la création d'emplois décents. En un sens, cela implique que le rappel de la colonialité est instrumentalisé pour répondre aux besoins des entreprises internationales. Dans le cas de Madagascar, nous identifions une *"rhétorique des qualifications coloniales"* par laquelle les entreprises de sous-traitance mettent en avant la complexité et la diversité des compétences requises dans le domaine de l'annotation, tout en essentialisant l'identité des travailleurs malgaches au regard des compétences coloniales nécessaires pour les tâches d'annotation. Nous suggérons de qualifier cette approche de *"paravent des qualifications coloniales"*. En plus de la fiabilité dans la prestation de services informatiques, qui nécessite en soi des compétences considérables, les travailleurs possèdent d'autres qualifications dans divers domaines. Cependant, ces compétences sont parfois négligées ou mises en avant par les entrepreneurs de la sous-traitance et de l'IA en fonction des besoins spécifiques des clients ou des projets.

En remontant la chaîne de production de l'IA au delà du lien entre le travail humain et l'automatisation, nous avons pu constater l'ensemble des étapes et qualifications requises pour produire un système d'IA. Nous avons déjà évoqué le *travail relationnel, de coordination et de traduction* réalisé par les travailleurs. Il nous semble, après analyse de leurs compétences, que l'invisibilité porte davantage sur leurs qualifications que sur l'importance de leur travail. Précisément, notre analyse de l'*assetisation coloniale* et du *rappel colonial* illustre ce travail de qualification et d'assujettissement du Sud global. Toute une partie de ce travail liées aux données est réalisée dans les *coulisses des coulisses* de l'IA pour reprendre le concept de *"backstage"* que Susan Star emprunte à Goffmann.

Dans les coulisses de la chaîne de valeur, les travailleurs mettent en place des stratégies d'acquisition des compétences nécessaires à leur insertion dans l'industrie de l'externalisation des données. Ils se concentrent notamment sur l'apprentissage du français et de l'informatique, manifestant leur intention délibérée de se positionner sur ce secteur du marché de l'emploi. Par ailleurs, lorsque les jeunes travailleurs envisagent une carrière professionnelle dans un secteur et un pays qu'ils perçoivent comme offrant des perspectives limitées, l'option du départ s'impose souvent comme une alternative. Cependant, cette décision soulève des questions cruciales : quelles destinations choisir et quelles voies emprunter pour réussir cette migration professionnelle ? Une partie de la solution réside dans l'adoption de stratégies d'apprentissage qui mettent l'accent sur l'ouverture vers l'extérieur, l'intégration à l'économie mondiale et l'établissement de liens avec les nations occidentales.

Du côté des entreprises de sous-traitance, ce processus fait l'objet d'une valorisation, car il leur permet de répondre aux besoins de leurs clients. Pour les travailleurs, nous qualifions ce processus d' "*assetisation coloniale* ", car ces qualifications facilitent l'intégration des travailleurs sur le marché du travail, améliorant leur stabilité financière. Cependant, la chaîne de valeur demeure marquée par les stigmates de la colonisation et l'émergence de nouveaux acteurs français dans le secteur des services, qui incarnent une nouvelle forme de colonialité en exerçant une influence déterminante sur le processus d'acquisition de compétences et sur l'accès au marché du travail pour les jeunes travailleurs d'Antananarivo. La valorisation de ces "actifs coloniaux" se trouve donc considérablement entravée par le contexte post-colonial au sein duquel ces compétences sont acquises. Les difficultés d'émigration vers la France et l'Europe, ainsi que le secteur de l'externalisation, qui apparaît comme la seule option pour valoriser ces compétences, limitent fortement les perspectives des travailleurs.

Rendre compte de ces voix marginalisées de par leur position géographique vis-à-vis d'une centralité occidentale a constitué un autre enjeu de cette thèse. C'est sur ce point également que l'on a pu constater l'importance de la localisation de la chaîne de valeur, car elle détermine presque entièrement la structure de la captation de la valeur. C'est finalement parce que ces travailleurs malgaches sont malgaches qu'ils toucheront un salaire moins élevé. Même si ces derniers possèdent des qualifications, ceux qui n'ont pas eu l'opportunité de poursuivre leurs études ou de travailler à l'étranger demeurent exclus de la création d'entreprises dans les zones franches, qui semble être réservée aux entrepreneurs français ou occidentaux. Ces constats suggèrent une forme de capitalisme qui s'appuie sur des distinctions raciales et au détriment des pays situés en périphérie de la mondialisation contemporaine.

L'extractivisme de données : valeur globale, mais précarité locale

Cette thèse a également permis de traiter des effets du secteur de l'IA sur le développement de Madagascar en articulant différents niveaux de la chaîne pour en comprendre les effets structurels, tout en tenant compte

de l'*agentivité* des travailleurs. Cet aspect de notre travail constitue un défi pratique en soi. Comment appréhender le parcours et le vécu de travailleurs socialement et géographiquement très éloigné de notre propre position ? Aux yeux de nombreux travailleurs, nous représentions l'homme blanc venu à Madagascar pour créer une entreprise et susceptible de leur donner du travail. Au cours de cette expérience de terrain, nous avons pu nous rendre compte de ce qu'incarne l'homme blanc sur la Grande Île, qu'il s'agisse de retraités venus profiter des travailleuses du sexe malgaches, de chefs d'entreprises du secteur de la sous-traitance, ou des entrepreneurs du secteur associatif.

Pendant un mois et demi, c'est au travers du regard d'Andry, notre chauffeur, traducteur puis ami que nous avons perçu notre propre position. Que dire aux travailleurs déçus de comprendre que nos recherches ne pourraient pas changer grand chose à leurs conditions de vie et de travail ? Que dire lorsque nous rencontrons les dirigeants d'une association de sauvegarde des baobabs qui nous expliquent que leur objectif est d'en replanter pour pouvoir lancer leur marque de jus à l'international. Que dire lorsque Andry évoque une ex-copine qu'il a croisé avec un blanc à notre hôtel profitant d'un repas offert par ce dernier en échange de faveurs sexuelles ? Notre position d'homme blanc vient certes avec l'autorité de l'ancien colon, mais également avec la méfiance de celui qui profite de tout ce qu'offre la Grande Île.

C'est encore en discutant avec Andry que nous avons pu mieux comprendre le rapport particulier qu'entretiennent les travailleurs malgaches avec les français et qui mêle la résilience vis-à-vis de cette domination et de leurs conditions et la critique du profit à tout prix. Au moment de repartir et après une dernière soirée, Andry nous dira pourtant que ça ne lui était jamais arrivé d'apprécier des "blancs" auparavant. Il fallait pour cela intégrer un aspect de notre positionnement de chercheur qui nous a paru très important : l'acceptation de ce que l'on incarne et le fait de tenir compte de la *colonialité* des travailleurs en essayant d'opérer un travail de traduction fidèle à ce qu'ils vivent au quotidien.

Notre défi était également théorique. Il est difficile de contribuer à la littérature qui porte sur les enjeux de développement des pays du Sud global, tant la notion est sujette à controverse. Il nous a alors fallu déplier toutes les nuances de vécu de nos enquêtés, au travail et dans leur vie quotidienne. L'expérience du terrain réalisé à Madagascar a contribué à choisir comme point de départ la critique du concept de "colonialisme des données" (*data colonialism*). Sur la base d'une théorie de la valeur qui lie les mécanismes d'extraction des travailleurs des données rémunérés et des utilisateurs routiniers qui fournissent gratuitement des contenus aux grandes plateformes du Web, cette théorie tend à mettre sur un même plan d'exploitation des individus dont le vécu est très éloigné. Comment admettre cette conceptualisation après un mois et demi passé à interroger des travailleurs précaires à Madagascar ? Comment concevoir que l'extraction de la valeur issue du travail gratuit des consommateurs constitue une forme d'exploitation coloniale au même titre que ce que vivent nos enquêtés ? Comment articuler cette colonialité pour redéfinir la notion de création de valeur ?

Ces questionnements se sont approfondis à mesure que le terrain progressait et nous ont incité, du moins en partie, à tenter de rendre compte du vécu des travailleurs en tentant, autant que faire se peut, de nous débarrasser de certains concepts ancrés dans les pays du nord. Nous ne sommes pas tous des sujets coloniaux des grandes plateformes, pas plus que nous ne sommes des travailleurs des plateformes. Par nos vécus, nos conditions de travail, nos perspectives d'avenir, nous nous distinguons des travailleurs précaires du numérique dont la valeur réside bien souvent dans la distinction raciale qui s'opère entre nous et eux et qui déterminent nos positions dans la chaîne de valeur. Ce positionnement explique le choix de nous concentrer sur les conditions matérielles d'existence des travailleurs et pas seulement sur les mécanismes d'extraction de valeur. En prenant en compte la subjectivité des travailleurs en ce qui concerne leurs revenus et leurs stratégies de gestion de la précarité, nous apportons une contribution à la littérature en démontrant une fois de plus que l'existence des chaînes de valeur globales est intrinsèquement liée à l'émergence de chaînes de précarité locales (Selwyn, 2019).

Cette dimension du secteur se traduit par plusieurs éléments constitutifs de vécus personnels et professionnels précaires. Nous avons évoqué le *rappel colonial* et la captation par les entreprises de sous-traitance d'un "biocapital" qualifié. S'ajoutent à cela : les difficultés d'intégration dans le marché du travail de la classe moyenne urbaine diplômée, le salaire de subsistance qui peine à combler les besoins des foyers des travailleurs, le manque de protection face aux licenciements économiques et les perspectives limitées de développement professionnel dans le secteur et en dehors.

D'après Ursula Huws (2014), la précarité se manifeste à travers la subsistance, les conditions de travail et l'appropriation de la plus-value issue de ce travail. Notre étude montre bien, qu'en dépit de leur rôle central dans la production des modèles d'IA et de leurs qualifications, les travailleurs malgaches demeurent largement sous-rémunérés et méconnus. Que ce soit en considérant leur position au sein de la structure des coûts, leur propre rémunération ou encore leur vision des revenus qu'ils tirent de l'annotation, ces derniers luttent pour satisfaire leur subsistance. Pour assurer leur survie, ils doivent diversifier leurs sources de revenus en adoptant une approche axée à la fois sur le développement professionnel et la gestion de la stabilité financière de leur foyer. Ces activités supplémentaires jouent un rôle déterminant dans la capacité des entreprises d'IA à trouver des travailleurs acceptant des rémunérations modestes pour l'annotation de données. En réalité, que ce soit pour envisager un avenir professionnel différent ou pour garantir la stabilité financière de leur foyer, la multitude d'activités secondaires constitue une manière de faire face à la précarité professionnelle et économique.

Nous montrons par ailleurs que les difficultés structurelles du pays obligent les travailleurs à chercher une stabilité financière au sein d'un secteur marqué par une relation post-coloniale et orienté vers l'extraction de valeur. N'étant pas en capacité de subvenir pleinement à leurs besoins et d'accumuler du capital par leur activité principale, les travailleurs se voient contraints d'entreprendre des activités complémentaires.

Pendant ce temps, les entrepreneurs de la sous-traitance, les entreprises d'IA et les géants du numérique tirent profit de ces efforts pour accroître leur plus-value. La principale conséquence de cette chaîne de valeur mondiale est aussi de réduire les perspectives d'avenir des travailleurs malgaches, ce qui limite également les retombées positives sur le développement du pays. Plutôt que le micro-travail, la micro-finance ou bien des l'aide au développement, il nous paraît ici judicieux de rappeler qu'un meilleur salaire issu d'un partage plus équitable de la valeur pourrait constituer une évolution positive substantielle pour les travailleurs et pour le développement de Madagascar.

Enfin, nos données ne nous permettent pas de conclure que cette activité constitue une passerelle vers d'autres opportunités professionnelles ou vers l'entrepreneuriat dans d'autres secteurs. Au contraire, la "*démographie professionnelle*" des entreprises de sous-traitance limite la mobilité ascendante des travailleurs malgaches, la majorité des postes les mieux rémunérés étant occupés par des Français. La création d'entreprises locales est entravée par une double inégalité, à la fois raciale et financière, subie par les travailleurs, mettant en évidence l'impact persistant de la structure raciale et post-coloniale du secteur sur ces derniers. Pourtant, nos données quantitatives et nos entretiens attestent largement des compétences des travailleurs malgaches. Face à leur précarité, leur désir d'entreprendre reflète à la fois ces compétences et leur résilience face aux difficultés structurelles qu'ils rencontrent quotidiennement. Néanmoins, ils se sentent pris au piège dans une activité qui, selon les mots d'un travailleur, les "*empêche de mourir*", les condamnant à un emploi sans véritable avenir dans un pays offrant un futur bien trop incertain.

Bibliographie

- Alauzen, M. (2019). *Plis et replis de l'État plateforme. Enquête sur la modernisation des services publics en France* (These de doctorat, Paris Sciences et Lettres (ComUE)). Consulté le 2022-02-11, sur <https://www.theses.fr/2019PSLEM037>
- Andrianampiarivo, T. (2018). Structure de classes et transformations rurales et agricoles : analyse des trajectoires des ménages de la Petite prospérité d'Itasy (Madagascar). *Mondes en développement*, 182(2), 153–174. Consulté le 2023-09-13, sur <https://www.cairn.info/revue-mondes-en-developpement-2018-2-page-153.htm> (Place : Louvain-la-Neuve Publisher : De Boeck Supérieur) doi : 10.3917/med.182.0153
- Aneesh, A. (2006). *Virtual migration : the programming of globalization*. Durham : Duke University Press. (OCLC : ocm61748413)
- Aneesh, A. (2015). *Neutral Accent : How Language, Labor, and Life Become Global*. Duke University Press. Consulté le 2020-03-08, sur <http://read.dukeupress.edu/lookup/doi/10.1215/9780822375715> doi : 10.1215/9780822375715
- Anwar, M. A., & Graham, M. (2019, décembre). *Digital Labour at Economic Margins : African Workers and the Global Information Economy* (SSRN Scholarly Paper N° ID 3499706). Rochester, NY : Social Science Research Network. Consulté le 2020-03-31, sur <https://papers.ssrn.com/abstract=3499706>
- Anwar, M. A., & Graham, M. (2020, avril). Between a rock and a hard place : Freedom, flexibility, precarity and vulnerability in the gig economy in Africa. *Competition & Change*, 102452942091447. Consulté le 2020-07-20, sur <http://journals.sagepub.com/doi/10.1177/1024529420914473> doi : 10.1177/1024529420914473
- Anwar, M. A., Schäfer, S., & Golušin, S. (2023, août). Work futures : globalization, planetary markets, and uneven developments in the gig economy. *Globalizations*, 0(0), 1–19. Consulté le 2023-08-22, sur <https://doi.org/10.1080/14747731.2023.2236876> (Publisher : Routledge _eprint : <https://doi.org/10.1080/14747731.2023.2236876>) doi : 10.1080/14747731.2023.2236876
- Appay, B. (1998, février). Economic Concentration and the Externalization of Labour. *Economic and Industrial Democracy*, 19(1), 161–184. Consulté le 2022-01-24, sur <https://doi.org/10.1177/0143831X98191008> (Publisher : SAGE Publications Ltd) doi : 10.1177/0143831X98191008
- Arshad, N., Salleh, S., Aris, S., Janom, N., & Mastuki, N. (2013). Micro sourcing : The SWOT analysis on the demand, supply and platforms. In *Proc. Sci. Inf. Conf., SAI* (pp. 768–773). Consulté sur <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84892505237&partnerID=40&md5=b16b6af829ba27b432f978df54f8d13b> (Journal Abbreviation : Proc. Sci. Inf. Conf., SAI)
- Attard-Frost, B., & Widder, D. G. (2023, juillet). *The Ethics of AI Value Chains : An Approach for Integrating and Expanding AI Ethics Research, Practice, and Governance*. arXiv. Consulté le 2023-08-22, sur <http://>

- arxiv.org/abs/2307.16787 (arXiv :2307.16787 [cs]) doi : 10.48550/arXiv.2307.16787
- Attewell, P. (1987, août). The Deskilling Controversy. *Work and Occupations*, 14(3), 323–346. Consulté le 2023-08-02, sur <https://doi.org/10.1177/0730888487014003001> (Publisher : SAGE Publications Inc) doi : 10.1177/0730888487014003001
- Aubry, A., Kuehni, M., & Scalabrini, L. (2021, avril). Introduction - Pratiques et politiques de la négociation pour accéder et se maintenir sur un terrain d'enquête. *Cambouis, la revue des sciences sociales aux mains sales*. Consulté le 2022-08-25, sur <https://revue-cambouis.org/index.php/cambouis/article/view/71> doi : 10.52983/crev.vi0.71
- Autor, D., & Salomons, A. (2018, juillet). *Is Automation Labor-Displacing? Productivity Growth, Employment, and the Labor Share* (Rapport technique N° w24871). Cambridge, MA : National Bureau of Economic Research. Consulté le 2021-08-12, sur <http://www.nber.org/papers/w24871.pdf> doi : 10.3386/w24871
- Aytes, A. (2012). Return of the Crowds : Mechanical Turk and Neoliberal States of Exception.. doi : 10.4324/9780203145791-12
- Badouard, R. (2017). *Le désenchantement de l'Internet : désinformation, rumeur et propagande*. Limoges : Fyp éditions.
- Badouard, R. (2021). Modérer la parole sur les réseaux sociaux. Politiques des plateformes et régulation des contenus. *Réseaux*, 225(1), 87–120. Consulté le 2023-01-24, sur <https://www.cairn.info/revue-reseaux-2021-1-page-87.htm> (Place : Paris Publisher : La Découverte) doi : 10.3917/res.225.0087
- Bair, J. (2010, avril). Les cadres d'analyse des chaînes globales. *Revue française de gestion*, n° 201(2), 103–119. Consulté le 2020-04-03, sur <https://www.cairn.info/revue-francaise-de-gestion-2010-2-page-103.htm> (Publisher : Lavoisier)
- Balakrishnan, A. (2011). India's IT Industry : The End of the Beginning. *Social Research*, 78(1), 1–20. Consulté le 2020-04-20, sur <https://www.jstor.org/stable/23347201> (Publisher : The New School)
- Banat-Berger, F. (2010). Les archives et la révolution numérique. *Le Débat*, 158(1), 70–82. Consulté le 2023-01-13, sur <https://www.cairn.info/revue-le-debat-2010-1-page-70.htm> (Place : Paris Publisher : Gallimard) doi : 10.3917/deba.158.0070
- Banerjee, A. V., & Duflo, E. (2008, juin). What Is Middle Class about the Middle Classes around the World? *Journal of Economic Perspectives*, 22(2), 3–28. Consulté le 2023-10-23, sur <https://www.aeaweb.org/articles?id=10.1257/jep.22.2.3> doi : 10.1257/jep.22.2.3
- Banki. (2013, décembre). Precarity of place : a complement to the growing precariat literature. *Global Discourse*, 3(3), 450–463. Consulté le 2023-06-21, sur <https://bristoluniversitypressdigital.com/doi/10.1080/23269995.2014.881139> doi : 10.1080/23269995.2014.881139
- Bastin, G., & Tubaro, P. (2018, septembre). Le moment big data des sciences sociales :. *Revue française de sociologie*, Vol. 59(3), 375–394. Consulté le 2022-10-18, sur <https://www.cairn.info/>

- [revue-francaise-de-sociologie-2018-3-page-375.htm?ref=doi](#) doi : 10.3917/rfs.593.0375
- Batt, R., Holman, D., & Holtgrewe, U. (2009, juillet). The Globalization of Service Work : Comparative Institutional Perspectives on Call Centers : Introduction to a Special Issue of the *Industrial & Labor Relations Review*. *ILR Review*, 62(4), 453–488. Consulté le 2023-01-18, sur <http://journals.sagepub.com/doi/10.1177/001979390906200401> doi : 10.1177/001979390906200401
- Baudelot, C., & Gollac, M. (2002). *Travailler pour être heureux? le bonheur et le travail en France*. Paris : Fayard.
- Baudry, B. (2003). *Économie de la firme*. Paris : La Découverte.
- Baudry, B., & Chassagnon, V. (2012, janvier). Responsabilité sociale inter-firmes, coordination et régulation de la firme-réseau multinationale : une analyse économique. *Revue d'économie industrielle*(137), 43–64. Consulté le 2020-03-25, sur <http://journals.openedition.org/rei/5286> (Number : 137 Publisher : De Boeck Supérieur) doi : 10.4000/rei.5286
- Baya-Laffite, N., Beaudé, B., & Garrigues, J. (2018, novembre). Le deep learning au service de la prédiction de l'orientation sexuelle dans l'espace public. *Rezeaux*, n° 211(5), 137–172. Consulté le 2020-10-23, sur <https://www.cairn.info/revue-rezeaux-2018-5-page-137.htm> (Publisher : La Découverte)
- Beauvisage, T., Beuscart, J.-S., Cardon, V., Mellet, K., & Trespeuch, M. (2013). Notes et avis des consommateurs sur le web. Les marchés à l'épreuve de l'évaluation profane. *Rezeaux*, 177(1), 131–161. Consulté le 2023-01-24, sur <https://www.cairn.info/revue-rezeaux-2013-1-page-131.htm> (Place : Paris Publisher : La Découverte) doi : 10.3917/res.177.0131
- Bechmann, A., & Bowker, G. C. (2019, janvier). Unsupervised by any other name : Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society*, 6(1), 2053951718819569. Consulté le 2020-03-10, sur <https://doi.org/10.1177/2053951718819569> doi : 10.1177/2053951718819569
- Beerepoot, N., & Hendriks, M. (2013, octobre). Employability of offshore service sector workers in the Philippines : opportunities for upward labour mobility or dead-end jobs? *Work, Employment and Society*, 27(5), 823–841. Consulté le 2023-09-15, sur <https://doi.org/10.1177/0950017012469065> (Publisher : SAGE Publications Ltd) doi : 10.1177/0950017012469065
- Beerepoot, N., & Lambregts, B. (2016). Competition and Wage Effects in the Global Online Market for Microwork and Services Outsourcing. In B. Nicholson, R. Babin, & M. C. Lacity (Eds.), *Socially Responsible Outsourcing : Global Sourcing with Social Impact* (pp. 119–137). London : Palgrave Macmillan UK. Consulté le 2020-04-03, sur https://doi.org/10.1007/978-1-137-55729-2_7 doi : 10.1007/978-1-137-55729-2_7
- Benanav, A. (2020). *Automation and the future of work*. London ; New York : Verso. (OCLC : on1147891672)
- Benbouzid, B. (2018, novembre). Quand prédire, c'est gérer. *Rezeaux*, n° 211(5), 221–256. Consulté le 2021-06-14, sur <https://www.cairn.info/revue-rezeaux-2018-5-page-221.htm> (Publisher :

La Découverte)

- Benedetto-Meyer, M., & Boboc, A. (2021). *Sociologie du numérique au travail*. Malakoff : Armand Colin.
- Bensidoun, I., & Ünal, D. (2007, août). Mondialisation des services : de la mesure à l'analyse. *CEPII Research Center, Working Paper*.
- Benzécri, J.-P. (Ed.). (1980). *Pratique de l'analyse des données*. Paris : Dunod.
- Berg, J. (2019, mai). *Protecting Workers in the Digital Age : Technology, Outsourcing and the Growing Precariousness of Work* (SSRN Scholarly Paper N° ID 3413740). Rochester, NY : Social Science Research Network. Consulté le 2022-02-14, sur <https://papers.ssrn.com/abstract=3413740> doi : 10.2139/ssrn.3413740
- Berg, J., Furrer, M., Harmon, E., Rani, U., & Silberman, M. S. (2018). *Digital labour platforms and the future of work : Towards decent work in the online world*. International Labour Office Geneva.
- Beunet, L. (2013). Pratiques de modération sur le site de rencontres Meetic. *Ethnologie française*, 43(3), 495–504. Consulté le 2022-08-25, sur <https://www.cairn.info/revue-ethnologie-francaise-2013-3-page-495.htm> (Place : Paris cedex 14 Publisher : Presses Universitaires de France) doi : 10.3917/ethn.133.0495
- Bizeul, D. (1998). Le récit des conditions d'enquête : exploiter l'information en connaissance de cause. *Revue française de sociologie*, 39(4), 751–787. Consulté le 2022-08-25, sur http://www.persee.fr/doc/rfsoc_0035-2969_1998_num_39_4_4840 (Publisher : Persée - Portail des revues scientifiques en SHS) doi : 10.2307/3323009
- Blanchard, P., & Patou, C. (2003, janvier). Les usages de l'analyse factorielle dans les sciences sociales en France. In (pp. 85–110).
- Boden, M. A. (1987). *Artificial intelligence and natural man* (2. ed., expanded éd.). London : MIT Pr.
- Boelaert, J., & Ollion, . (2018). The Great Regression. Machine Learning, Econometrics, and the Future of Quantitative Social Sciences. *Revue française de sociologie*, 59(3), 475–506. Consulté le 2022-10-18, sur <https://www.cairn.info/revue-francaise-de-sociologie-2018-3-page-475.htm> (Place : Paris Publisher : Presses de Sciences Po) doi : 10.3917/rfs.593.0475
- Bonnefond, C., & Andrianampiarivo, T. (2020). Les "classes moyennes" des pays en développement à l'épreuve de la crise COVID-19 : la Chine et Madagascar mis en perspective.
- Bowker, G. C. (1997). *Social science, technical systems, and cooperative work : beyond the great divide*. Mahwah, N.J : Lawrence Erlbaum Associates.
- Bowker, G. C., & Star, S. L. (1999). *Sorting things out : classification and its consequences*. Cambridge, Mass : MIT Press.
- Bowker, G. C., & Star, S. L. (2000). *Sorting Things Out : Classification and Its Consequences*. MIT Press. (Google-Books-ID : xHIP8WqzizYC)
- Braverman, H. (1998). *Labor and monopoly capital : the degradation of work in the twentieth century* (25th

- anniversary ed éd.). New York : Monthly Review Press.
- Brayne, S., & Christin, A. (2021, août). Technologies of Crime Prediction : The Reception of Algorithms in Policing and Criminal Courts. *Social Problems*, 68(3), 608–624. Consulté le 2022-10-24, sur <https://academic.oup.com/socpro/article/68/3/608/5782114> doi : 10.1093/socpro/spaa004
- Brynjolfsson, E., Rock, D., & Syverson, C. (2017, novembre). *Artificial Intelligence and the Modern Productivity Paradox : A Clash of Expectations and Statistics* (Working Paper N° 24001). National Bureau of Economic Research. Consulté le 2022-01-11, sur <https://www.nber.org/papers/w24001> (Series : Working Paper Series) doi : 10.3386/w24001
- Burrell, J. (2016, juin). How the machine ‘thinks’ : Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 2053951715622512. Consulté le 2021-06-30, sur <https://doi.org/10.1177/2053951715622512> (Publisher : SAGE Publications Ltd) doi : 10.1177/2053951715622512
- Burrell, J., & Fourcade, M. (2021). The Society of Algorithms. *Annual Review of Sociology*, 47(1), 213–237. Consulté le 2021-08-31, sur <https://doi.org/10.1146/annurev-soc-090820-020800> (eprint : <https://doi.org/10.1146/annurev-soc-090820-020800>) doi : 10.1146/annurev-soc-090820-020800
- Buscatto, M. (2002a, janvier). Les centres d’appels, usines modernes ? Les rationalisations paradoxales de la relation téléphonique. *Sociologie du travail*, 44(1), 99–117. Consulté le 2022-03-23, sur <https://journals.openedition.org/sdt/32796?lang=en> (Number : 1 Publisher : Association pour le développement de la sociologie du travail) doi : 10.4000/sdt.32796
- Buscatto, M. (2002b, janvier). Les centres d’appels, usines modernes ? Les rationalisations paradoxales de la relation téléphonique. *Sociologie du Travail*, 44(1), 99–117. Consulté le 2020-11-16, sur <http://www.sciencedirect.com/science/article/pii/S003802960101202X> doi : 10.1016/S0038-0296(01)01202-X
- Béraud, M., Colin, T., & Grasser, B. (2008, juin). La qualification dans les centres d’appels : affaiblissement ou recomposition ? *Travail et Emploi*(114), 45–57. Consulté le 2023-01-18, sur <https://journals.openedition.org/travailemploi/3916#tocto1n3> (Number : 114 Publisher : La documentation française) doi : 10.4000/travailemploi.3916
- Calderón, J. (2006, octobre). Le travail face à la restructuration productive : le cas d’un centre d’appels. *Formation emploi. Revue française de sciences sociales*(96), 11–24. Consulté le 2023-01-18, sur <https://journals.openedition.org/formationemploi/2509#tocto2n1> (Number : 96 Publisher : La documentation française) doi : 10.4000/formationemploi.2509
- CALLON, M. (1986). ÉLÉMENTS POUR UNE SOCIOLOGIE DE LA TRADUCTION : La domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc. *L’Année sociologique* (1940/1948-), 36, 169–208. Consulté le 2021-08-06, sur <https://www.jstor.org/stable/27889913> (Publisher : Presses Universitaires de France)
- Cardon, D. (2013). Dans l’esprit du PageRank. Une enquête sur l’algorithme de Google. *Réseaux*, 177(1), 63–

95. Consulté le 2023-01-26, sur <https://www.cairn.info/revue-reseaux-2013-1-page-63.htm> (Place : Paris Publisher : La Découverte) doi : 10.3917/res.177.0063
- Cardon, D. (2015). *A quoi rêvent les algorithmes : nos vies à l'heure des big data*. Paris : La République des idées : Seuil. (OCLC : ocn925890784)
- Cardon, D., & Casilli, A. A. (2015). *Qu'est-ce que le digital labor?* Bry-sur-Marne : INA. (OCLC : 920030990)
- Cardon, D., Cointet, J.-P., & Mazières, A. (2018, novembre). La revanche des neurones. *Reseaux*, n° 211(5), 173–220. Consulté le 2020-03-10, sur <https://www.cairn.info/revue-reseaux-2018-5-page-173.htm>
- Casilli, A., & Posada, J. (2019). *The Platformization of Labor and Society*. Oxford University Press. Consulté le 2020-03-24, sur <https://halshs.archives-ouvertes.fr/halshs-01895137> (Pages : 293)
- Casilli, A., Tubaro, P., Ludec, C. L., Coville, M., Besenval, M., Mouhtare, T., & Wahal, E. (2019a). Le Micro-Travail en France. Derrière l'automatisation, de nouvelles précarités au travail? , 74.
- Casilli, A., Tubaro, P., Ludec, C. L., Coville, M., Besenval, M., Mouhtare, T., & Wahal, E. (2019b, mai). *Micro-work in France. Behind Automation, New Forms of Precarious Labour?* [Le Micro-Travail en France. Derrière l'automatisation, de nouvelles précarités au travail?] (Rapport technique N° hal-02139528). HAL. Consulté le 2020-03-10, sur <https://ideas.repec.org/p/hal/wpaper/hal-02139528.html>
- Casilli, A. A. (2017a). Digital Labor Studies Go Global : Toward a Digital Decolonial Turn. , 21.
- Casilli, A. A. (2017b, septembre). Global Digital Culture| Digital Labor Studies Go Global : Toward a Digital Decolonial Turn. *International Journal of Communication*, 11(0), 21. Consulté le 2020-04-18, sur <https://ijoc.org/index.php/ijoc/article/view/6349> (Number : 0)
- Casilli, A. A. (2017c, juin). Il n'y a pas d'algorithme. In *L'appétit des géants. Pouvoir des algorithmes, ambitions des plateformes* (pp. pp. 10–19.). Caen. Consulté le 2021-08-18, sur <https://hal.archives-ouvertes.fr/hal-01625657>
- Casilli, A. A. (2019). *En attendant les robots : enquête sur le travail du clic*. Paris XIXe : Éditions du Seuil.
- Chagnon, C. W., Durante, F., Gills, B. K., Hagolani-Albov, S. E., Hokkanen, S., Kangasluoma, S. M. J., ... Vuola, M. P. S. (2022, juin). From extractivism to global extractivism : the evolution of an organizing concept. *The Journal of Peasant Studies*, 49(4), 760–792. Consulté le 2022-11-14, sur <https://doi.org/10.1080/03066150.2022.2069015> (Publisher : Routledge _eprint : <https://doi.org/10.1080/03066150.2022.2069015>) doi : 10.1080/03066150.2022.2069015
- Chassagnon, V. (s. d.). Pouvoir et coopération dans la firme et entre les firmes. , 27.
- Chassagnon, V. (2014). Pouvoir et économie politique. L'apport de François Perroux à l'analyse positive de la firme. *Revue économique*, 65(5), 719–742. Consulté le 2023-02-03, sur <https://www.cairn.info/revue-economique-2014-5-page-719.htm> (Place : Paris Publisher : Presses de Sciences Po) doi : 10.3917/reco.655.0719

- Chassagnon, V. (2019, décembre). De la sous-traitance à la firme-réseau multinationale : une nouvelle division du travail pour de nouvelles frontières économiques. In S. Tisseyre (Ed.), *Sécuriser la sous-traitance : quels nouveaux défis ?* (pp. 119–147). Toulouse : Presses de l'Université Toulouse 1 Capitole. Consulté le 2020-04-06, sur <http://books.openedition.org/putc/7156> (Code : Sécuriser la sous-traitance : quels nouveaux défis? Reporter : Sécuriser la sous-traitance : quels nouveaux défis?)
- Chateauraynaud, F. (2019, mars). Petit traité de contre-intelligence artificielle. Retour sociologique sur des expérimentations numériques. *Zilsel*, N° 5(1), 174–195. Consulté le 2021-07-12, sur <https://www.cairn.info/revue-zilsel-2019-1-page-174.htm> (Bibliographie_available : 0 Cairndomain : www.cairn.info Cite Par_available : 0 Publisher : Éditions du Croquant)
- CHERRY, M. A. (2016). Virtual Work and Invisible Labor. In M. G. Crain, W. R. Poster, & M. A. Cherry (Eds.), *Invisible Labor* (1^{re} éd., pp. 71–86). University of California Press. Consulté le 2020-04-28, sur <http://www.jstor.org/stable/10.1525/j.ctv1xxwt7.8>
- Chmielewski, M., & Kucker, S. C. (2020, mai). An MTurk Crisis? Shifts in Data Quality and the Impact on Study Results. *Social Psychological and Personality Science*, 11(4), 464–473. Consulté le 2023-02-01, sur <https://doi.org/10.1177/1948550619875149> (Publisher : SAGE Publications Inc) doi : 10.1177/1948550619875149
- Christin, A. (2017, décembre). Algorithms in practice : Comparing web journalism and criminal justice. *Big Data & Society*, 4(2), 2053951717718855. Consulté le 2021-05-03, sur <https://doi.org/10.1177/2053951717718855> (Publisher : SAGE Publications Ltd) doi : 10.1177/2053951717718855
- Christin, A. (2020a, octobre). The ethnographer and the algorithm : beyond the black box. *Theory and Society*, 49(5), 897–918. Consulté le 2021-09-01, sur <https://doi.org/10.1007/s11186-020-09411-3> doi : 10.1007/s11186-020-09411-3
- Christin, A. (2020b). *Metrics at work : journalism and the contested meaning of algorithms*. Princeton : Princeton University Press.
- Cihuelo, J. (2010). Le cas d'un modèle hybride de centre d'appels. Entre rationalisation et réappropriation professionnelle. *Réseaux*, 164(6), 161–196. Consulté le 2023-01-18, sur <https://www.cairn.info/revue-reseaux-2010-6-page-161.htm> (Place : Paris Publisher : La Découverte) doi : 10.3917/res.164.0161
- Cockburn, C., & Ploux, M. (2004). Le matériel dans le pouvoir masculin. *Cahiers du Genre*, n° 36(1), 89–120. Consulté le 2020-07-01, sur <https://www.cairn.info/revue-cahiers-du-genre-2004-1-page-89.htm> (Publisher : L'Harmattan)
- Coe, N. (1997, juillet). US Transnationals and the Irish Software Industry : Assessing the Nature, Quality and Stability of a New Wave of Foreign Direct Investment. *European Urban and Regional Studies*, 4(3), 211–230. Consulté le 2020-04-08, sur <https://doi.org/10.1177/096977649700400302> (Publisher : SAGE Publications Ltd) doi : 10.1177/096977649700400302

- Coe, N. M., Dicken, P., & Hess, M. (2008, mai). Global production networks : realizing the potential. *Journal of Economic Geography*, 8(3), 271–295. Consulté le 2020-02-17, sur <https://academic.oup.com/joeg/article/8/3/271/938286> doi : 10.1093/jeg/lbn002
- Coe, N. M., Johns, J., & Ward, K. (2007, novembre). Mapping the Globalization of the Temporary Staffing Industry. *The Professional Geographer*, 59(4), 503–520. Consulté le 2020-02-17, sur <https://www.tandfonline.com/doi/citedby/10.1111/j.1467-9272.2007.00638.x> doi : 10.1111/j.1467-9272.2007.00638.x
- Collins, H. (1998, octobre). Socialness and the Undersocialized Conception of Society. *Science, Technology, & Human Values*, 23(4), 494–516. Consulté le 2022-10-11, sur <https://doi.org/10.1177/016224399802300408> (Publisher : SAGE Publications Inc) doi : 10.1177/016224399802300408
- Collins, H. M. (1992). *Experts artificiels : machines intelligentes et savoir social*. Paris :??d. du Seuil. (OCLC : 802722346)
- Congress of the U. S., D. O. o. T., Washington. (1985). *Automation of America's Offices, 1985-2000* (Rapport technique). DIANE Publishing.
- Cornet, M., Le Ludec, C., Wahal, E., & Joulin, M. (2022, mai). Beyond 'platformisation' : Designing a mixed-methods approach to inspect (digital) working conditions through organisational systems. *Work Organisation, Labour & Globalisation*. Consulté le 2022-09-23, sur <https://www.scienceopen.com/hosted-document?doi=10.13169/workorgalaboglob.16.1.0052> (Publisher : Pluto Journals) doi : 10.13169/workorgalaboglob.16.1.0052
- Corporaal, G. F., & Lehdonvirta, V. (2017). Platform sourcing : How Fortune 500 firms are adopting online freelancing platforms. *Oxford Internet Institute : Oxford*. Retrieved from <https://www.oii.ox.ac.uk/publications/platform-sourcing.pdf>. Accessed on June, 1, 2018.
- Couldry, N., & Mejias, U. A. (2019a). *The costs of connection : how data is colonizing human life and appropriating it for capitalism*. Stanford, California : Stanford University Press.
- Couldry, N., & Mejias, U. A. (2019b, juin). Making data colonialism liveable : how might data's social order be regulated? *Internet Policy Review*, 8(2). Consulté le 2020-11-10, sur <https://policyreview.info/articles/analysis/making-data-colonialism-liveable-how-might-datas-social-order-be-regulated>
- Crawford, K. (2021). *Atlas of Ai : power, politics, and the planetary costs of artificial intelligence*. New Haven : Yale University Press.
- Crawford, K., & Paglen, T. (2019). Excavating AI. -. Consulté le 2021-09-01, sur <https://excavating.ai>
- Darmon, M. (2005). Le psychiatre, la sociologue et la boulangère : analyse d'un refus de terrain. *Genèses*, 58(1), 98–112. Consulté le 2022-08-25, sur <https://www.cairn.info/revue-geneses-2005-1-page-98.htm> (Place : Paris Publisher : Belin) doi : 10.3917/gen.058.0098
- Denis, J. (2018). *Le travail invisible des données : éléments pour une sociologie des infrastructures scripturales*.

Paris : Mines ParisTech.

- Denis, J., & Gardey, D. (2018). *Le travail invisible des données : les enjeux pour une sociologie des infrastructures scripturales*. (ISBN : 9782356715241 OCLC : 1153930265)
- Denton, E., Hanna, A., Amironesei, R., Smart, A., Nicole, H., & Scheuerman, M. K. (2020a). Bringing the people back in : Contesting benchmark machine learning datasets. *arXiv preprint arXiv :2007.07399*.
- Denton, E., Hanna, A., Amironesei, R., Smart, A., Nicole, H., & Scheuerman, M. K. (2020b, juillet). Bringing the People Back In : Contesting Benchmark Machine Learning Datasets. *arXiv :2007.07399 [cs]*. Consulté le 2021-02-11, sur <http://arxiv.org/abs/2007.07399> (arXiv : 2007.07399)
- Desrosières, A. (2010). *La politique des grands nombres*. La Découverte. Consulté le 2020-03-10, sur <https://www.cairn.info/la-politique-des-grands-nombres--9782707165046.htm>
- Desrosières, A. (2013, avril). La mesure du développement : un domaine propice à l'innovation méthodologique. *Revue Tiers Monde*, n°213(1), 23–32. Consulté le 2020-04-08, sur <https://www.cairn.info/revue-tiers-monde-2013-1-page-23.htm> (Publisher : Armand Colin)
- De Terssac, G., Soubie, J.-L., & Neveu, J.-P. (1988). Systèmes experts et transferts d'expertise. *Sociologie du travail*, 30(3), 461–477. Consulté le 2023-04-19, sur https://www.persee.fr/doc/sotra_0038-0296_1988_num_30_3_2418 (Publisher : Persée - Portail des revues scientifiques en SHS) doi : 10.3406/sotra.1988.2418
- Diakopoulos, N. (2015, mai). Algorithmic Accountability. *Digital Journalism*, 3(3), 398–415. Consulté le 2021-09-10, sur <https://doi.org/10.1080/21670811.2014.976411> (Publisher : Routledge _eprint : <https://doi.org/10.1080/21670811.2014.976411>) doi : 10.1080/21670811.2014.976411
- Difallah, D., Filatova, E., & Ipeirotis, P. (2018). Demographics and Dynamics of Mechanical Turk Workers. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining - WSDM '18* (pp. 135–143). Marina Del Rey, CA, USA : ACM Press. Consulté le 2020-07-30, sur <http://dl.acm.org/citation.cfm?doid=3159652.3159661> doi : 10.1145/3159652.3159661
- Dimou, M., & Fernand, P. (2008). Attractivité et stratégies de développement de trois zones textiles de l'océan Indien. *Mondes en développement*, 144(4), 115–128. Consulté le 2022-10-07, sur <https://www.cairn.info/revue-mondes-en-developpement-2008-4-page-115.htm> (Place : Louvain-la-Neuve Publisher : De Boeck Supérieur) doi : 10.3917/med.144.0115
- Do, S., Ollion, E., & Shen, R. (2022, août). *The Augmented Social Scientist. Using Sequential Transfer Learning to Annotate Millions of Texts with Human-Level Accuracy*. SocArXiv. Consulté le 2022-10-18, sur <https://osf.io/preprints/socarxiv/3fkzc/> doi : 10.31235/osf.io/3fkzc
- Du Tertre, C., & Santilli, G. (1992). *Automatisation et travail : utopies, réalités, débats : des années cinquante aux années quatre-vingt-dix* (1re éd éd.). Paris : Presses universitaires de France.
- D'Cruz, P., Mulder, R., Noronha, E., Beerepoot, N., & Magala, S. (2019, janvier). The changing role of the nation-state and regulation : Workplace bullying legislation in The Netherlands. *The Economic*

- and Labour Relations Review*, 30(1), 77–98. Consulté le 2020-04-29, sur <https://doi.org/10.1177/1035304618823959> (Publisher : SAGE Publications Ltd) doi : 10.1177/1035304618823959
- Ekbia, H., & Nardi, B. (2017). *Heteromation, and Other Stories of Computing and Capitalism*. doi : 10.7551/mitpress/10767.001.0001
- Ekbia, H. R., & Nardi, B. A. (2017). *Heteromation, and other stories of computing and capitalism*. Consulté le 2021-05-03, sur <https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1513525> (OCLC : 985461245)
- Eubanks, V. (2017). *Automating inequality : how high-tech tools profile, police, and punish the poor* (First Edition éd.). New York, NY : St. Martin's Press.
- EUBANKS, V. (2018). *Automating Inequality How High-Tech Tools Profile, Police, and Punish the Poor*. Saint Martin's press. Consulté le 2020-02-12, sur <https://us.macmillan.com/automatinginequality/virginiaeubanks/9781250074317>
- Feldstein, S. (2019). *The Global Expansion of AI Surveillance*. Consulté le 2022-02-02, sur <https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847>
- Fernandes, S. (2009, janvier). Wittorski, R., Professionnalisation et développement professionnel. Paris : L'Harmattan, 2007. *Questions Vives. Recherches en éducation*(Vol.5 n°11), 315–316. Consulté le 2021-05-07, sur <http://journals.openedition.org/questionsvives/632> (ISBN : 9782912643353 Number : Vol.5 n°11 Publisher : Université de Provence - Département des Sciences de l'éducation)
- Finez, J. (2014, février). La construction des prix à la SNCF, une socio-histoire de la tarification. *Revue française de sociologie*, Vol. 55(1), 5–39. Consulté le 2021-08-31, sur <https://www.cairn.info/revue-francaise-de-sociologie-2014-1-page-5.htm> (Bibliographie_available : 1 Cairndomain : www.cairn.info Cite Par_available : 1 Publisher : Presses de Sciences Po)
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, ♦., & Srikumar, M. (2020). Principled Artificial Intelligence : Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *SSRN Electronic Journal*. doi : 10.2139/ssrn.3518482
- Flecker, J. (Ed.). (2016). *Space, Place and Global Digital Work*. London : Palgrave Macmillan UK. Consulté le 2020-04-06, sur <http://link.springer.com/10.1057/978-1-137-48087-3> doi : 10.1057/978-1-137-48087-3
- Flick, U. (Ed.). (2014). *The SAGE handbook of qualitative data analysis*. Los Angeles : SAGE. (OCLC : ocn861318873)
- Forde, C., Stuart, M., Joyce, S., Oliver, L., Valizade, D., Alberti, G., ... Carson, C. (2017). *The Social Protection of Workers in the Platform Economy, DIRECTORATE GENERAL FOR INTERNAL POLICIES POLICY DEPARTMENT A : ECONOMIC AND SCIENTIFIC POLICY*.
- Fournier, C., Lambert, M., & Marion-Vernoux, I. (2020, juillet). What do Young Employees Dream of? Quality of Work, Career Aspirations and Desire for Mobility Among the Under 30s. *Economie et Statistique /*

- Economics and Statistics*(514-515-516), 113–131. Consulté le 2022-08-25, sur <https://insee.fr/en/statistiques/4621091?sommaire=4621107> doi : 10.24187/ecostat.2020.514t.2007
- Frey, C. B., & Osborne, M. A. (2017, janvier). The future of employment : How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280. Consulté le 2021-03-16, sur <https://www.sciencedirect.com/science/article/pii/S0040162516302244> doi : 10.1016/j.techfore.2016.08.019
- Friederici, N., Wahome, M., & Graham, M. (2020). *Digital Entrepreneurship in Africa : How a Continent Is Escaping Silicon Valley's Long Shadow*. The MIT Press. Consulté le 2023-09-13, sur <https://direct.mit.edu/books/oa-monograph/4850/Digital-Entrepreneurship-in-AfricaHow-a-Continent> doi : 10.7551/mitpress/12453.001.0001
- Friedmann, G. (1955). Le Travail en miettes. *Esprit (1940-)*(232 (10/11)), 1725–1747. Consulté le 2020-11-16, sur <https://www.jstor.org/stable/24253944> (Publisher : Editions Esprit)
- Galpaya, H., Perampalam, S., & Senanayake, L. (2018). Investigating the Potential for Micro-work and Online-Freelancing in Sri Lanka. In L. Pupillo, E. Noam, & L. Waverman (Eds.), *Digitized Labor : The Impact of the Internet on Employment* (pp. 229–250). Cham : Springer International Publishing. Consulté le 2020-04-06, sur https://doi.org/10.1007/978-3-319-78420-5_14 doi : 10.1007/978-3-319-78420-5_14
- Gardey, D. (2008). *Écrire, calculer, classer*. La Découverte. Consulté le 2020-07-01, sur <https://www-cairn-info-s.proxy.bu.dauphine.fr/ecrire-calculer-classer--9782707153678.htm>
- Gefen, D., & Carmel, E. (2008). Is the World Really Flat? A Look at Offshoring at an Online Programming Marketplace. *MIS Quarterly*, 32(2), 367–384. Consulté le 2020-04-21, sur <https://www.jstor.org/stable/25148844> (Publisher : Management Information Systems Research Center, University of Minnesota) doi : 10.2307/25148844
- Gereffi, G., Humphrey, J., & Sturgeon, T. (2005, février). The governance of global value chains. *Review of International Political Economy*, 12(1), 78–104. Consulté le 2020-04-06, sur <http://www.tandfonline.com/doi/abs/10.1080/09692290500049805> doi : 10.1080/09692290500049805
- Gereffi, G., Korzeniewicz, M., & Political Economy of the World-System Conference. (1994). *Commodity chains and global capitalism*. Consulté le 2020-04-09, sur <http://public.eblib.com/EBLPublic/PublicView.do?ptiID=496758> (OCLC : 895043314)
- Gillespie, T. (2020, juillet). Content moderation, AI, and the question of scale. *Big Data & Society*, 7(2), 2053951720943234. Consulté le 2023-01-28, sur <https://doi.org/10.1177/2053951720943234> (Publisher : SAGE Publications Ltd) doi : 10.1177/2053951720943234
- Gitelman, L. (Ed.). (2013). *"Raw data" is an oxymoron*. Cambridge, Massachusetts; London, England : The MIT Press.
- Goeta, S. (2016). *Instaurer des données, instaurer des publics : une enquête sociologique dans les coulisses de*

- l'open data* (phdthesis, Télécom ParisTech). Consulté le 2022-01-03, sur <https://pastel.archives-ouvertes.fr/tel-01458098>
- Gonzalez, R., Gasco, J., & Llopis, J. (2006, janvier). Information systems offshore outsourcing : A descriptive analysis. *Industrial Management & Data Systems*, 106(9), 1233–1248. Consulté le 2020-02-12, sur <https://doi.org/10.1108/02635570610712555> doi : 10.1108/02635570610712555
- Graaf, P., Kalmijn, M., Kraaykamp, G., & Monden, C. (2010). *The Netherlands Longitudinal Lifecourse Study (NELLS Wave 1)*.
- Graham, M. (2015, avril). Contradictory Connectivity : Spatial Imaginaries and Technomediated Positionalities in Kenya's Outsourcing Sector. *Environment and Planning A : Economy and Space*, 47(4), 867–883. Consulté le 2020-02-27, sur <https://doi.org/10.1068/a140275p> (Publisher : SAGE Publications Ltd) doi : 10.1068/a140275p
- Graham, M., & Anwar, M. A. (2019, avril). The global gig economy : Towards a planetary labour market ? *First Monday*, 24(4). Consulté le 2019-04-06, sur <https://firstmonday.org/ojs/index.php/fm/article/view/9913> doi : 10.5210/fm.v24i4.9913
- Graham, M., Hjorth, I., & Lehdonvirta, V. (2017a, mai). Digital labour and development : impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer : European Review of Labour and Research*, 23(2), 135–162. Consulté le 2021-03-24, sur <https://doi.org/10.1177/1024258916687250> (Publisher : SAGE Publications Ltd) doi : 10.1177/1024258916687250
- Graham, M., Hjorth, I., & Lehdonvirta, V. (2017b, mai). Digital labour and development : impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer : European Review of Labour and Research*, 23(2), 135–162. Consulté le 2019-04-06, sur <https://doi.org/10.1177/1024258916687250> doi : 10.1177/1024258916687250
- Graham, M., Hogan, B., Straumann, R. K., & Medhat, A. (2014, juillet). Uneven Geographies of User-Generated Information : Patterns of Increasing Informational Poverty. *Annals of the Association of American Geographers*, 104(4), 746–764. Consulté le 2020-04-23, sur <https://doi.org/10.1080/00045608.2014.910087> (Publisher : Routledge _eprint : <https://doi.org/10.1080/00045608.2014.910087>) doi : 10.1080/00045608.2014.910087
- Graham, M., Sabbata, S. D., & Zook, M. A. (2015). Towards a study of information geographies : (im)mutable augmentations and a mapping of the geographies of information. *Geo : Geography and Environment*, 2(1), 88–105. Consulté le 2020-10-19, sur <https://rgs-ibg.onlinelibrary.wiley.com/doi/abs/10.1002/geo2.8> (_eprint : <https://rgs-ibg.onlinelibrary.wiley.com/doi/pdf/10.1002/geo2.8>) doi : 10.1002/geo2.8
- Gray, M. L., & Suri, S. (2019). *Ghost work : how to stop Silicon Valley from building a new global underclass*. Boston : Houghton Mifflin Harcourt.
- Grossin, W. (1960). L'automation : à la recherche d'un concept. *Sociologie du travail*, 2(1), 76–81. Consulté

- le 2021-02-19, sur https://www.persee.fr/doc/sotra_0038-0296_1960_num_2_1_1296 (Publisher : Persée - Portail des revues scientifiques en SHS) doi : 10.3406/sotra.1960.1296
- Gupta, N., Martin, D., Hanrahan, B., & O'Neill, J. (2014). Turk-life in India. In *Proc Int ACM SIGGROUP Conf Support Group Work* (pp. 1–11). Association for Computing Machinery. Consulté sur <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84911942914&doi=10.1145/2f2660398.2660403&partnerID=40&md5=9baadcbe231fb5952d6652d34225e55a> (Journal Abbreviation : Proc Int ACM SIGGROUP Conf Support Group Work) doi : 10.1145/2660398.2660403
- Gurstein, P. (2002). *Wired to the world, chained to the home : telework in daily life*. Vancouver, B.C.; London : University of British Columbia Press; Eurospan. (OCLC : 49691561)
- Hagendorff, T. (2020, mars). The Ethics of AI Ethics : An Evaluation of Guidelines. *Minds and Machines*, 30(1), 99–120. Consulté le 2022-08-18, sur <https://doi.org/10.1007/s11023-020-09517-8> doi : 10.1007/s11023-020-09517-8
- Hammett, D., & Sporton, D. (2012). Paying for interviews? Negotiating ethics, power and expectation. *Area*, 44(4), 496–502. Consulté le 2022-08-24, sur <https://www.jstor.org/stable/23358205> (Publisher : Wiley)
- Handley, A. (2015). Varieties of Capitalists? The Middle-Class, Private Sector and Economic Outcomes in Africa. *Journal of International Development*, 27(5), 609–627. Consulté le 2023-10-23, sur <https://onlinelibrary.wiley.com/doi/abs/10.1002/jid.3113> (_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jid.3113>) doi : 10.1002/jid.3113
- Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. (2018). A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In *Conf Hum Fact Comput Syst Proc* (Vol. 2018-April). Association for Computing Machinery. Consulté sur <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85046962507&doi=10.1145/2f3173574.3174023&partnerID=40&md5=52a7b6302710ecb00ea1c7e96b867ba4> (Journal Abbreviation : Conf Hum Fact Comput Syst Proc) doi : 10.1145/3173574.3174023
- Harmon, E., & Silberman, M. (2018). Rating Working Conditions on Digital Labor Platforms. *Computer Supported Cooperative Work : CSCW : An International Journal*, 27(3-6), 1275–1324. Consulté sur <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85047406238&doi=10.1007/2fs10606-018-9313-5&partnerID=40&md5=21d97784abaa32faa0d3f6a3b83991fc> (Publisher : Springer Netherlands) doi : 10.1007/s10606-018-9313-5
- Head, E. (2009, octobre). The ethics and implications of paying participants in qualitative research. *International Journal of Social Research Methodology*, 12(4), 335–344. Consulté le 2022-08-24, sur <https://doi.org/10.1080/13645570802246724> (Publisher : Routledge _eprint : <https://doi.org/10.1080/13645570802246724>) doi : 10.1080/13645570802246724
- Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires : A

- quantitative analysis of the published literature. *American Sociological Review*, 43(4), 447–462. (Place : US Publisher : American Sociological Assn) doi : 10.2307/2094771
- Hechiche Salah, L., Ben Radhia, I., & Ben Ammar-Mamlouk, Z. (2009). Les centres d'appels : « eldorado technologique » ou forme moderne de dégradation des conditions de travail ? *Management & Avenir*, 26(6), 74–94. Consulté le 2023-01-18, sur <https://www.cairn.info/revue-management-et-avenir-2009-6-page-74.htm> (Place : Soliers Publisher : Management Prospective Ed.) doi : 10.3917/mav.026.0074
- Heeks, R. (1996). *India's Software Industry : State Policy, Liberalisation and Industrial Development* (1st éd.). USA : Sage Publications, Inc.
- Henin, C. (2021). *Expliquer et justifier les systèmes de décisions algorithmiques* (phdthesis, Université de Lyon). Consulté le 2022-10-24, sur <https://tel.archives-ouvertes.fr/tel-03551798>
- Henneguelle, A., & Jatteau, A. (2021). *Sociologie de la quantification* (N° 762). Paris : la Découverte.
- Hicks, M. (2017). *Programmed inequality : how Britain discarded women technologists and lost its edge in computing*. Cambridge, MA : MIT Press.
- Hicks, M. (2018). *Programmed inequality : how Britain discarded women technologists and lost its edge in computing*. Cambridge, MA London, UK : MIT Press. (OCLC : on1004074349)
- Hirth, M., Hoßfeld, T., & Tran-Gia, P. (2011). Anatomy of a crowdsourcing platform - Using the example of microworkers.com. In *Proc. - Int. Conf. Innovative Mob. Internet Serv. Ubiquitous Comput., IMIS* (pp. 322–329). Consulté sur <https://www.scopus.com/inward/record.uri?eid=2-s2.0-80052346592&doi=10.1109%2fIMIS.2011.89&partnerID=40&md5=93e504fee5dd87667b4bfba1370674a3> (Journal Abbreviation : Proc. - Int. Conf. Innovative Mob. Internet Serv. Ubiquitous Comput., IMIS) doi : 10.1109/IMIS.2011.89
- Hoffman, S. G. (2017, juillet). Managing Ambiguities at the Edge of Knowledge : Research Strategy and Artificial Intelligence Labs in an Era of Academic Capitalism. *Science, Technology, & Human Values*, 42(4), 703–740. Consulté le 2021-04-23, sur <https://doi.org/10.1177/0162243916687038> (Publisher : SAGE Publications Inc) doi : 10.1177/0162243916687038
- Hopkins, T. K., & Wallerstein, I. (1977). Patterns of development of the modern world-system. *Review (Fernand Braudel Center)*, 111–145. (Publisher : JSTOR)
- Horne, J. (2017). To Spread the French Language Is to Extend the Patrie : The Colonial Mission of the Alliance Française. *French historical studies*, 40(1), 95–127. (Place : Durham Publisher : Duke University Press) doi : 10.1215/00161071-3686068
- Hornuf, L., & Vrankar, D. (2022, août). Hourly Wages in Crowdfunding : A Meta-Analysis. *Business & Information Systems Engineering*. Consulté le 2022-11-10, sur <https://doi.org/10.1007/s12599-022-00769-5> doi : 10.1007/s12599-022-00769-5
- Horton, J., & Chilton, L. (2010, janvier). The Labor Economics of Paid Crowdsourcing. *arXiv :1001.0627 [cs]*.

- Consulté le 2020-04-29, sur <http://arxiv.org/abs/1001.0627> (arXiv : 1001.0627)
- Howson, K., Ferrari, F., Ustek-Spilda, F., Salem, N., Johnston, H., Katta, S., ... Graham, M. (2022). Driving the digital value network : Economic geographies of global platform capitalism. *Global Networks*, 22(4), 631–648. Consulté le 2023-08-14, sur <https://onlinelibrary.wiley.com/doi/abs/10.1111/glob.12358> (_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/glob.12358>) doi : 10.1111/glob.12358
- Howson, K., Johnston, H., Cole, M., Ferrari, F., Ustek-Spilda, F., & Graham, M. (2022, octobre). Unpaid labour and territorial extraction in digital value networks. *Global Networks*, glob.12407. Consulté le 2022-11-10, sur <https://onlinelibrary.wiley.com/doi/10.1111/glob.12407> doi : 10.1111/glob.12407
- Husson, F., Josse, J., & Lê, S. (2008, mars). FactoMineR : An R Package for Multivariate Analysis. *Journal of Statistical Software*, 25. doi : 10.18637/jss.v025.i01
- Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022, janvier). Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences*, 119(1), e2025334119. Consulté le 2022-11-03, sur <https://www.pnas.org/doi/10.1073/pnas.2025334119> (Publisher : Proceedings of the National Academy of Sciences) doi : 10.1073/pnas.2025334119
- Huws, U. (2007). The emergence of EMERGENCE : the challenge of designing research on the new international division of labour. *Work Organisation, Labour & Globalisation*, 1(2), 20–35. Consulté le 2021-03-03, sur <https://www.jstor.org/stable/10.13169/workorglaboglob.1.2.0020> (Publisher : Pluto Journals)
- Huws, U. (2014). *Labor in the global digital economy : the cybertariat comes of age*. New York : Monthly Review Press.
- HUWS, U. (2014). The Underpinnings of Class in the Digital Age. In *Labor in the Global Digital Economy* (pp. 149–182). NYU Press. Consulté le 2020-04-28, sur <http://www.jstor.org/stable/j.ctt1287j8b.10>
- Huws, U., Flecker, J., Bandi, R., Dahlmann, S., Eichmann, H., Hirschfeld, K., ... Staden, P. (2004). *Asian Emergence : The World's Back Office?* Institute for Employment Studies Brighton.
- Hällsten, M., Edling, C., & Rydgren, J. (2015, janvier). The effects of specific occupations in position generator measures of social capital. *Social Networks*, 40, 55–63. Consulté le 2021-09-20, sur <https://www.sciencedirect.com/science/article/pii/S0378873314000355> doi : 10.1016/j.socnet.2014.06.002
- Héder, M. (2020, décembre). A criticism of AI ethics guidelines. *Információs Társadalom*, 20, 57. doi : 10.22503/infvars.XX.2020.4.5
- Idowu, A., & Elbanna, A. (2019a). *Bright ICT and Unbounded Employment : Typology of Crowdworkers and Their Lived and Envisaged Career Trajectory in Nigeria* (Vol. 558; Dwivedi Y., Ayaburi E., Boateng R., & Effah J., Eds.). Springer New York LLC. Consulté

- sur https://www.scopus.com/inward/record.uri?eid=2-s2.0-85068485474&doi=10.1007%2f978-3-030-20671-0_32&partnerID=40&md5=2c0d7e5bc5d4c60edae5eaa3e07efd04 (Journal Abbreviation : IFIP Advances in Information and Communication Technology Pages : 486 Publication Title : IFIP Advances in Information and Communication Technology) doi : 10.1007/978-3-030-20671-0_32
- Idowu, A., & Elbanna, A. (2019b). Bright ICT and Unbounded Employment : Typology of Crowdworkers and Their Lived and Envisaged Career Trajectory in Nigeria. In Y. Dwivedi, E. Ayaburi, R. Boateng, & J. Effah (Eds.), *ICT Unbounded, Social Impact of Bright ICT Adoption* (pp. 470–486). Cham : Springer International Publishing. doi : 10.1007/978-3-030-20671-0_32
- Ipeirotis, P. G. (2010, mars). *Demographics of Mechanical Turk* (SSRN Scholarly Paper N° ID 1585030). Rochester, NY : Social Science Research Network. Consulté le 2020-07-16, sur <https://papers.ssrn.com/abstract=1585030>
- Irani, L. (2013, novembre). The cultural work of microwork. *New Media & Society*, 17(5), 720–739. Consulté le 2020-04-29, sur <https://doi.org/10.1177/1461444813511926> (Publisher : SAGE Publications) doi : 10.1177/1461444813511926
- Irani, L. (2015a). The cultural work of microwork. *New Media and Society*, 17(5), 720–739. Consulté sur <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84930517734&doi=10.1177%2f1461444813511926&partnerID=40&md5=5240329882000964cd3a004f5a3bcfdb> (Publisher : SAGE Publications Ltd) doi : 10.1177/1461444813511926
- Irani, L. (2015b, mai). The cultural work of microwork. *New Media & Society*, 17(5), 720–739. Consulté le 2020-03-11, sur <https://doi.org/10.1177/1461444813511926> (Publisher : SAGE Publications) doi : 10.1177/1461444813511926
- Irani, L., & Silberman, M. (2013). Turkopticon : Interrupting worker invisibility in Amazon Mechanical Turk. In *Conf Hum Fact Comput Syst Proc* (pp. 611–620). Consulté sur <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84877960208&doi=10.1145%2f2470654.2470742&partnerID=40&md5=3a2389c08e4c3308c035bf58bc53b075> (Journal Abbreviation : Conf Hum Fact Comput Syst Proc) doi : 10.1145/2470654.2470742
- Jaeger, C. (2002). L'impossible evaluation du travail des teleopérateurs. Le cas de deux centres d'appel. *Réseaux*, 114(4), 51–90. Consulté le 2022-08-25, sur <https://www.cairn.info/revue-reseaux1-2002-4-page-51.htm> (Place : Cachan Publisher : Lavoisier)
- Jarrett, K. (2022). *Digital labor*. Cambridge, UK : Polity. (OCLC : on1273479511)
- Jaton, F. (2017, décembre). We get the algorithms of our ground truths : Designing referential databases in digital image processing. *Social Studies of Science*, 47(6), 811–840. Consulté le 2021-05-20, sur <https://doi.org/10.1177/0306312717730428> (Publisher : SAGE Publications Ltd) doi : 10.1177/0306312717730428

- Jaton, F. (2019, mars). « Pardonnez cette platitude » : de l'intérêt des ethnographies de laboratoire pour l'étude des processus algorithmiques. *Zilsel*, N° 5(1), 315–339. Consulté le 2020-10-07, sur <https://www.cairn.info/revue-zilsel-2019-1-page-315.html> (Publisher : Éditions du Croquant)
- Jaton, F., & Bowker, G. C. (2021). *The constitution of algorithms ground-truthing, programming, formula-tion*. Consulté le 2021-06-10, sur <https://doi.org/10.7551/mitpress/12517.001.0001> (OCLC : 1247401334)
- Jobin, A., Ienca, M., & Vayena, E. (2019, septembre). Artificial Intelligence : the global landscape of ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. Consulté le 2020-03-25, sur <http://arxiv.org/abs/1906.11668> (arXiv : 1906.11668) doi : 10.1038/s42256-019-0088-2
- Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S. G., ... Shestakofsky, B. (2021, janvier). Toward a Sociology of Artificial Intelligence : A Call for Research on Inequalities and Structural Change. *Socius*, 7, 2378023121999581. Consulté le 2021-05-03, sur <https://doi.org/10.1177/2378023121999581> (Publisher : SAGE Publications) doi : 10.1177/2378023121999581
- Joyce, K. A., Darfler, K., George, D., Ludwig, J., & Unsworth, K. (2018). Engaging STEM ethics education. *Engaging Science, Technology, and Society*, 4, 1–7.
- Kano, L., Tsang, E. W. K., & Yeung, H. W.-c. (2020, juin). Global value chains : A review of the multi-disciplinary literature. *Journal of International Business Studies*, 51(4), 577–622. Consulté le 2020-12-17, sur <https://doi.org/10.1057/s41267-020-00304-2> doi : 10.1057/s41267-020-00304-2
- Karpik, L. (1989). L'économie de la qualité. *Revue française de sociologie*, 30(2), 187–210. Consulté le 2023-11-03, sur <https://www.jstor.org/stable/3321761> (Publisher : [Sciences Po University Press, Association Revue Française de Sociologie]) doi : 10.2307/3321761
- Kitchin, R. (2014). *The data revolution : big data, open data, data infrastructures & their consequences*. Los Angeles, California : SAGE Publications. (OCLC : ocn871211376)
- Kleibert, J. M., & Mann, L. (2020, avril). Capturing Value amidst Constant Global Restructuring? Information-Technology-Enabled Services in India, the Philippines and Kenya. *The European Journal of Development Research*. Consulté le 2020-07-24, sur <https://doi.org/10.1057/s41287-020-00256-1> doi : 10.1057/s41287-020-00256-1
- Kroeker, L., O'Kane, D., & Scharrer, T. (2018). *Middle classes in Africa : changing lives and conceptual challenges*. Cham : Palgrave Macmillan.
- Krzywdzinski, M. (2022, décembre). Toward a Socioeconomic Company-Level Theory of Automation at Work. *Weizenbaum Journal of the Digital Society*, 2(1). Consulté le 2023-01-28, sur https://ojs.weizenbaum-institut.de/index.php/wjds/article/view/2_1_5 (Number : 1) doi : 10.34669/wi.wjds/2.1.5
- Kuek, S. C., Paradi-Guilford, C., Fayomi, T., Imaizumi, S., & Ipeiritis, P. (2015). *The Global Opportunity in Online Outsourcing*. Consulté le 2020-07-16, sur </paper/The-Global-Opportunity-in-Online>

- [-Outsourcing-Kuek-Paradi-Guilford/ee45f094fba7737625f30ea8c787b3b29476cfe1](https://www.semanticscholar.org/Outsourcing-Kuek-Paradi-Guilford/ee45f094fba7737625f30ea8c787b3b29476cfe1) (Library Catalog : www.semanticscholar.org)
- Kuek, S. C., Paradi-Guilford, C., Fayomi, T., Imaizumi, S., Ipeirotis, P., Pina, P., & Singh, M. (2015, juin). *The Global Opportunity in Online Outsourcing* (World Bank Other Operational Studies N° 22284). The World Bank. Consulté le 2020-03-11, sur <https://econpapers.repec.org/paper/wbkwboper/22284.htm>
- Kässi, O., Lehdonvirta, V., & Stephany, F. (2021, avril). How Many Online Workers are there in the World? A Data-Driven Assessment. *arXiv :2103.12648 [econ, q-fin, stat]*. Consulté le 2022-02-14, sur <http://arxiv.org/abs/2103.12648> (arXiv : 2103.12648)
- Lagerie, P. (2019). *Les patrons de la vertu : de la responsabilité sociale des entreprises au devoir de vigilance*.
- Lehdonvirta, V. (2016). Algorithms that Divide and Unite : Delocalisation, Identity and Collective Action in 'Microwork'. In J. Flecker (Ed.), *Space, Place and Global Digital Work* (pp. 53–80). London : Palgrave Macmillan UK. Consulté le 2020-03-11, sur https://doi.org/10.1057/978-1-137-48087-3_4 doi : 10.1057/978-1-137-48087-3_4
- Lehdonvirta, V., Barnard, H., Graham, M., & Hjorth, I. (2014). Online labour markets - levelling the playing field for international service markets? *Working Paper*. Consulté le 2020-04-23, sur <https://ora.ox.ac.uk/objects/uuid:5c2b1fc7-04c0-40d1-89a0-f9a28ffb4891> (Publisher : Oxford Internet Institute)
- Lehdonvirta, V., Kässi, O., Hjorth, I., Barnard, H., & Graham, M. (2019, février). The Global Platform Economy : A New Offshoring Institution Enabling Emerging-Economy Microproviders. *Journal of Management*, 45(2), 567–599. Consulté le 2020-03-30, sur <https://doi.org/10.1177/0149206318786781> (Publisher : SAGE Publications Inc) doi : 10.1177/0149206318786781
- Li, Y., & Verhaeghe, P. (2015, août). The position generator approach to social capital research : measurements and results. *The Handbook of Research Methods and Applications on Social Capital*, 166–186. Consulté le 2022-08-25, sur [https://www.research.manchester.ac.uk/portal/en/publications/the-position-generator-approach-to-social-capital-research-measurements-and-results\(dd22c4d1-4382-445c-810e-b19c011f1c63\)/export.html](https://www.research.manchester.ac.uk/portal/en/publications/the-position-generator-approach-to-social-capital-research-measurements-and-results(dd22c4d1-4382-445c-810e-b19c011f1c63)/export.html) (Publisher : Edward Elgar Publishing Ltd)
- Light, J. S. (1999). When Computers Were Women. *Technology and Culture*, 40(3), 455–483. Consulté le 2020-10-06, sur <https://www.jstor.org/stable/25147356> (Publisher : [The Johns Hopkins University Press, Society for the History of Technology])
- Lin, N., & Erickson, B. (2008). *Social Capital An International Research Program*. Oxford University Press. Consulté le 2021-09-15, sur <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199234387.001.0001/acprof-9780199234387> doi : 10.1093/acprof:oso/9780199234387.001.0001

- Lynch, L. M. (1993, septembre). Entry-level jobs : First rung on the employment ladder or economic dead end? *Journal of Labor Research*, 14(3), 249–263. Consulté le 2023-09-15, sur <https://doi.org/10.1007/BF02685684> doi : 10.1007/BF02685684
- Machlup, F. (1972). *The production and distribution of knowledge in the United States* (1. paperback ed éd.). Princeton, NJ : Princeton Univ. Press. (OCLC : 1996438)
- Mackenzie, A. (2017). *Machine Learners : Archaeology of a Data Practice*. The MIT Press. Consulté le 2022-10-18, sur <https://direct.mit.edu/books/book/3129/machine-learnersarchaeology-of-a-data-practice> doi : 10.7551/mitpress/10302.001.0001
- MacKenzie, D., & Wajcman, J. (1999). *The social shaping of technology* (D. MacKenzie & J. Wajcman, Eds.). Buckingham, UK : Open University Press. Consulté le 2021-07-12, sur <http://mcgraw-hill.co.uk/openup/>
- MacKenzie, D. A. (2021). *Trading at the speed of light : how ultrafast algorithms are transforming financial markets*. Princeton, New Jersey : Princeton University Press.
- Malik, F., Nicholson, B., & Heeks, R. (2017). Understanding the Development Implications of Online Outsourcing. In *ICT4D*. doi : 10.1007/978-3-319-59111-7_35
- Mallard, A. (2014, août). Métamorphoses d'une question scientifique. *Rezeaux*, n° 184-185(2), 33–69. Consulté le 2020-11-16, sur <https://www.cairn.info/revue-rezeaux-2014-2-page-33.htm> (Publisher : La Découverte)
- Margaryan, A. (2019, mai). Workplace learning in crowdwork Comparing microworkers' and online freelancers' practices. *Journal of Workplace Learning*, 31(4), 250–273. (WOS :000475933200001) doi : 10.1108/JWL-10-2018-0126
- Marlow, C. A. (2005). *The structural determinants of media contagion* (Thesis, Massachusetts Institute of Technology). Consulté le 2022-08-25, sur <https://dspace.mit.edu/handle/1721.1/33883> (Accepted : 2007-12-07T19 :22 :42Z ISSN : 6646-4978)
- Martin, D., O'Neill, J., Gupta, N., & Hanrahan, B. V. (2016, février). Turking in a Global Labour Market. *Computer Supported Cooperative Work (CSCW)*, 25(1), 39–77. Consulté le 2020-04-24, sur <http://link.springer.com/10.1007/s10606-015-9241-6> doi : 10.1007/s10606-015-9241-6
- Maruani, M., & Nicole-Drancourt, C. (1989). *Au labour des dames : métiers masculins, emplois féminins*. Paris : Syros/Alternatives.
- Mateescu, A., & Nguyen, A. (2019). Algorithmic management in the workplace. *Data & Society*, 1–15.
- Mañé Vernet, F., & Benner, C. (2009). Dead-End Jobs or Career Opportunities? Advancement opportunities in call centers. *Working Papers*. Consulté le 2023-09-15, sur <https://ideas.repec.org//p/urv/wpaper/2072-42870.html> (Number : 2072/42870 Publisher : Universitat Rovira i Virgili, Department of Economics)
- Melber, H. (2022, août). Africa's Middle Classes. *Africa Spectrum*, 57(2), 204–219. Consulté le 2023-10-23,

- sur <https://doi.org/10.1177/00020397221089352> (Publisher : SAGE Publications Ltd) doi : 10.1177/00020397221089352
- Miceli, M., Posada, J., & Yang, T. (2022). Studying up machine learning data : Why talk about bias when we mean power? *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP), 1–14. (Publisher : ACM New York, NY, USA)
- Miceli, M., Schuessler, M., & Yang, T. (2020a, juillet). Between Subjectivity and Imposition : Power Dynamics in Data Annotation for Computer Vision. *arXiv :2007.14886 [cs]*. Consulté le 2020-10-16, sur <http://arxiv.org/abs/2007.14886> (arXiv : 2007.14886)
- Miceli, M., Schuessler, M., & Yang, T. (2020b, octobre). Between Subjectivity and Imposition : Power Dynamics in Data Annotation for Computer Vision. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2), 115 :1–115 :25. Consulté le 2021-02-26, sur <https://doi.org/10.1145/3415186> doi : 10.1145/3415186
- Miceli, M., Yang, T., Garcia, A. A., Posada, J., Wang, S. M., Pohl, M., & Hanna, A. (2022, août). *Documenting Data Production Processes : A Participatory Approach for Data Work*. arXiv. Consulté le 2023-03-24, sur <http://arxiv.org/abs/2207.04958> (arXiv :2207.04958 [cs])
- Mitter, S., & Pearson, R. (1992). *Global information processing : the emergence of software services and data entry jobs in selected developing countries* (Rapport technique N° 992897713402676). International Labour Organization. Consulté le 2020-03-18, sur <https://ideas.repec.org/p/ilo/ilowps/992897713402676.html> (Publication Title : ILO Working Papers)
- Molina, J. L., García-Macías, A., Lubbers, M. J., & Valenzuela-Garcia, H. (2020, juin). The embeddedness of social capital in personal networks. *Network Science*, 8(2), 189–203. Consulté le 2021-09-15, sur https://www.cambridge.org/core/product/identifler/S2050124219000304/type/journal_article doi : 10.1017/nws.2019.30
- Molina, J. L., Tubaro, P., Casilli, A., & Santos-Ortega, A. (2023, avril). Research Ethics in the Age of Digital Platforms. *Science and Engineering Ethics*, 29(3), 17. Consulté le 2023-08-30, sur <https://doi.org/10.1007/s11948-023-00437-1> doi : 10.1007/s11948-023-00437-1
- Moore, P. V., Upchurch, M., & Whittaker, X. (2018). Humans and machines at work : monitoring, surveillance and automation in contemporary capitalism. In *Humans and Machines at Work* (pp. 1–16). Springer.
- Moreschi, B., Pereira, G., & Cozman, F. G. (2020, avril). The Brazilian Workers in Amazon Mechanical Turk : Dreams and realities of ghost workers. *Revista Contracampo*, 39(1). Consulté le 2020-04-20, sur <https://periodicos.uff.br/contracampo/article/view/38252> (Number : 1) doi : 10.22409/contracampo.v39i1.38252
- Mosthaf, A., Schnabel, C., & Stephani, J. (2011, janvier). Low-wage careers : Are there dead-end firms and dead-end jobs? *Zeitschrift für ArbeitsmarktForschung*, 43(3), 231–249. Consulté le 2023-09-15, sur <https://labourmarketresearch.springeropen.com/articles/10.1007/s12651-010>

-0036-4 (Number : 3 Publisher : SpringerOpen) doi : 10.1007/s12651-010-0036-4

Mtsweni, J., & Burge III, L. (2014). The potential benefits of mobile microwork services in developing nations : Research opportunities and challenges. In *IST-Africa Conf. Exhib., IST-Africa*. IEEE Computer Society. Consulté sur <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84906761855&doi=10.1109%2fISTAFRICA.2014.6880636&partnerID=40&md5=57c05b65d5bb511a422f9a266cf457b8> (Journal Abbreviation : IST-Africa Conf. Exhib., IST-Africa) doi : 10.1109/ISTAFRICA.2014.6880636

Muniesa, F. (2003). *Des marchés comme algorithmes : sociologie de la cotation électronique à la Bourse de Paris* (PhD Thesis). Ecole Nationale Supérieure des Mines de Paris.

Muniesa, F. (2019, mars). Société du comportement, information de la sociologie. *Zilsel*, N° 5(1), 196–207. Consulté le 2021-07-12, sur <https://www.cairn.info/revue-zilsel-2019-1-page-196.htm> (Bibliographie_available : 0 Cairndomain : www.cairn.info Cite Par_available : 0 Publisher : Éditions du Croquant)

Möhlmann, M., & Zalmanson, L. (2017). *Hands on the wheel : Navigating algorithmic management and Uber drivers' autonomy*.

Nguyen, N. (2014). Microworkers crowdsourcing approach, challenges and solutions. In *CrowdMM - Proc. Int. Workshop Crowdsourcing Multimed., Workshop MM* (p. 1). Association for Computing Machinery, Inc. Consulté sur <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84915771596&doi=10.1145%2f2660114.2660128&partnerID=40&md5=1cf009ddd04e47580d397bb9cf0ca4df> (Journal Abbreviation : CrowdMM - Proc. Int. Workshop Crowdsourcing Multimed., Workshop MM) doi : 10.1145/2660114.2660128

Noble, S. U. (2018). *Algorithms of oppression : how search engines reinforce racism*. New York : New York University Press.

OCDE. (2019). *L'intelligence artificielle dans la société*. OECD. Consulté le 2022-02-16, sur https://www.oecd-ilibrary.org/science-and-technology/l-intelligence-artificielle-dans-la-societe_b7f8cd16-fr doi : 10.1787/b7f8cd16-fr

O'Neil, C. (2016). *Weapons of Math Destruction : How Big Data Increases Inequality and Threatens Democracy*. USA : Crown Publishing Group.

Opoku, D. (2010). *The Politics of Government-Business Relations in Ghana, 1982–2008*. (Pages : 264) doi : 10.1057/9780230113107

Organization, I. L. (2021, février). *The role of digital labour platforms in transforming the world of work* (Report). Consulté le 2021-03-24, sur http://www.ilo.org/global/research/global-reports/weso/2021/WCMS_771749/lang--en/index.htm (ISBN : 9789220319444)

Padios, J. M. (2018). *A nation on the line : call centers as postcolonial predicaments in the Philippines*. Durham ; London : Duke University Press.

- Paradi-Guilford, C. (2013, février). *Feasibility study : microwork for the Palestinian territories* (Rapport technique N° ACS3685). The World Bank. Consulté le 2020-03-11, sur <http://documents.worldbank.org/curated/en/529481468140639204/Feasibility-study-microwork-for-the-Palestinian-territories>
- Pasquale, F. (2015). *The black box society : the secret algorithms that control money and information*. Cambridge : Harvard University Press.
- Paugam, S. (2007). *Le salarié de la précarité : les nouvelles formes de l'intégration professionnelle*. Paris : Presses universitaires de France.
- Paugam, S., & Vendramin, P. (2020). 24. Le précaire, une nouvelle classe sociale ? In *50 questions de sociologie* (pp. 243–251). Paris cedex 14 : Presses Universitaires de France. Consulté le 2023-06-21, sur <https://www.cairn.info/50-questions-de-sociologie--9782130820673-p-243.htm> doi : 10.3917/puf.pauga.2020.01.0243
- Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2020, décembre). Data and its (dis)contents : A survey of dataset development and use in machine learning research. *arXiv :2012.05345 [cs]*. Consulté le 2021-02-11, sur <http://arxiv.org/abs/2012.05345> (arXiv : 2012.05345)
- Pesole, A., Fernandez-Macias, E., Brancati, C. U., & Herrera, E. G. (2019, juin). *How to quantify what is not seen? Two proposals for measuring platform work* (Rapport technique N° 2019-01). Joint Research Centre (Seville site). Consulté le 2022-08-10, sur <https://ideas.repec.org/p/ipt/laedte/201901.html> (Publication Title : JRC Working Papers on Labour, Education and Technology)
- Posch, L., Bleier, A., Flöck, F., & Strohmaier, M. (2017, novembre). A Cross-Country Comparison of Crowdworker Motivations. *arXiv :1711.03115 [cs]*. Consulté le 2020-04-29, sur <http://arxiv.org/abs/1711.03115> (arXiv : 1711.03115)
- Posch, L., Bleier, A., Flöck, F., & Strohmaier, M. (2018, décembre). Characterizing the Global Crowd Workforce : A Cross-Country Comparison of Crowdworker Demographics. *arXiv :1812.05948 [cs]*. Consulté le 2020-04-29, sur <http://arxiv.org/abs/1812.05948> (arXiv : 1812.05948)
- Pun, N., Tse, T., Shin, V., & Fan, L. (2020, mars). Conceptualising Socio-Economic Formations of Labour and Workers' Power in Global Production Networks :. *Sociology*. Consulté le 2021-01-22, sur <https://journals.sagepub.com/eprint/R5FH5ITBVWGGQH73FNJF/full> (Publisher : SAGE Publications Sage UK : London, England) doi : 10.1177/0038038520908779
- Raji, I. D., Bender, E. M., Paullada, A., Denton, E., & Hanna, A. (2021, novembre). *AI and the Everything in the Whole Wide World Benchmark*. arXiv. Consulté le 2023-03-10, sur <http://arxiv.org/abs/2111.15366> (arXiv :2111.15366 [cs]) doi : 10.48550/arXiv.2111.15366
- Rani, U., & Furrer, M. (2019). On-Demand Digital Economy : Can Experience Ensure Work and Income Security for Microtask Workers? *Jahrbucher fur Nationalökonomie und Statistik*, 239(3), 565–597. Consulté sur <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85068799524&doi=>

- 10.1515%2fjbnst-2018-0019&partnerID=40&md5=213ebe76ecf26171491b09d389ca2c2e (Publisher : De Gruyter Oldenbourg) doi : 10.1515/jbnst-2018-0019
- Razafindrakoto, M., & Roubaud, F. (1996). L'approche à haute intensité de main d'oeuvre (himo) : une opportunité pour Madagascar : essai de cadrage macro-économique. , 30.
- Razafindrakoto, M., & Roubaud, F. (1997). Les entreprises franches à Madagascar : économie d'enclave ou promesse d'une nouvelle prospérité? : nouvel esclavage ou opportunité pour le développement du pays? , 32.
- Razafindrakoto, M., Roubaud, F., & Wachsberger, J.-M. (2018a, novembre). Chapitre 1. La trajectoire malgache à l'aune des théories du développement. In *L'énigme et le paradoxe : Économie politique de Madagascar* (pp. 14–46). Marseille : IRD Éditions. Consulté le 2023-08-24, sur <http://books.openedition.org/irdeditions/22856> (Code : L'énigme et le paradoxe : Économie politique de Madagascar) doi : 10.4000/books.irdeditions.22856
- Razafindrakoto, M., Roubaud, F., & Wachsberger, J.-M. (2018b, novembre). Chapitre 2. Repères pour une économie politique de la trajectoire malgache. In *L'énigme et le paradoxe : Économie politique de Madagascar* (pp. 47–106). Marseille : IRD Éditions. Consulté le 2023-08-24, sur <http://books.openedition.org/irdeditions/22859> (Code : L'énigme et le paradoxe : Économie politique de Madagascar) doi : 10.4000/books.irdeditions.22859
- ROBERTS, S. T. (2019). *Behind the Screen : Content Moderation in the Shadows of Social Media*. Yale University Press. Consulté le 2020-03-25, sur <https://www.jstor.org/stable/j.ctvhrcz0v>
- Robette, N. (2011). *Explorer et décrire les parcours de vie : les typologies de trajectoires*. CEPED. Consulté le 2023-08-30, sur <https://shs.hal.science/halshs-01016125> (Pages : 86)
- Robette, N. (2021, juillet). Mesurer la dissemblance entre trajectoires. In *L'analyse statistique des trajectoires : Typologies de séquences et autres approches* (pp. 25–42). Paris : Ined Éditions. Consulté le 2023-08-30, sur <http://books.openedition.org/ined/16710> (Code : L'analyse statistique des trajectoires : Typologies de séquences et autres approches) doi : 10.4000/books.ined.16710
- Robette, N., & Thibault, N. (2008). Analyse harmonique qualitative ou méthodes d'appariement optimal? *Population*, Vol. 63(4), 621–646. Consulté le 2021-05-07, sur <https://www-cairn-info-s.proxy.bu.dauphine.fr/revue-population-2008-4-page-621.htm> (Publisher : Institut national d'études démographiques)
- Roomaney, Z., Van Belle, J.-P., & Tsibolane, P. (2018). Addressing the Unemployment Challenge through Mobile Digital Microwork. *CPR SOUTH*.
- Rosenblat, A., & Stark, L. (2016). Algorithmic labor and information asymmetries : A case study of Uber's drivers. *International Journal of Communication*, 10, 27.
- Régniez, P., Hurstel, J., Ludec, C. L., Toko, S., Elidrissi, H., Bailly, M., ... Astolfi, C.-P. (2020). *Travail à l'ère des plateformes. Mise à jour requise*. Conseil national du numérique. (Google-Books-ID : einyDwAAQBAJ)

- S. R. S. Aris, N. Janom, N. H. Arshad, S. S. Salleh, & N. Mastuki. (2013, novembre). A preliminary study for critical success factors of micro sourcing in Malaysia. In *2013 International Conference on Research and Innovation in Information Systems (ICRIIS)* (pp. 22–26). (Journal Abbreviation : 2013 International Conference on Research and Innovation in Information Systems (ICRIIS)) doi : 10.1109/ICRIIS.2013.6716680
- Sambasivan, N., & Veeraraghavan, R. (2022, avril). The Deskilling of Domain Expertise in AI Development. In *CHI Conference on Human Factors in Computing Systems* (pp. 1–14). New Orleans LA USA : ACM. Consulté le 2023-04-05, sur <https://dl.acm.org/doi/10.1145/3491102.3517578> doi : 10.1145/3491102.3517578
- Savage, M., Devine, F., Cunningham, N., Taylor, M., Li, Y., Hjellbrekke, J., ... Miles, A. (2013, avril). A New Model of Social Class? Findings from the BBC's Great British Class Survey Experiment. *Sociology*, 47(2), 219–250. Consulté le 2022-08-25, sur <https://doi.org/10.1177/0038038513481128> (Publisher : SAGE Publications Ltd) doi : 10.1177/0038038513481128
- Saxenian, A. (2002, décembre). Transnational Communities and the Evolution of Global Production Networks: The Cases of Taiwan, China and India. *Industry and Innovation*, 9(3), 183–202. Consulté le 2020-04-27, sur <https://doi.org/10.1080/1366271022000034453> (Publisher : Routledge _eprint : <https://doi.org/10.1080/1366271022000034453>) doi : 10.1080/1366271022000034453
- Saxenian, A., Motoyama, Y., & Quan, X. (2002). *Local and global networks of immigrant professionals in Silicon Valley*. San Francisco, CA : Public Policy Institute of California.
- Schmidt, F. A. (2019). *Crowdproduktion von Trainingsdaten : Zur Rolle von Online-Arbeit beim Trainieren autonomer Fahrzeuge* (Research Report N° 417). Study der Hans-Böckler-Stiftung. Consulté le 2020-03-24, sur <https://www.econstor.eu/handle/10419/201850> (ISBN : 9783865933300)
- Schmidt, G. B., & Jettinghoff, W. M. (2016, juillet). Using Amazon Mechanical Turk and other compensated crowdsourcing sites. *Business Horizons*, 59(4), 391–400. Consulté sur <http://www.sciencedirect.com/science/article/pii/S0007681316000306> doi : 10.1016/j.bushor.2016.02.004
- Scholz, T. (2012). *Digital Labor : The Internet as Playground and Factory*. Routledge.
- Schradie, J. (2019). *The revolution that wasn't : how digital activism favors conservatives*. Cambridge, Massachusetts; London, England : Harvard University Press.
- Schriner, A., & Oerther, D. (2014). No really, (crowd) work is the silver bullet. In Shoag D. & Vidan A. (Eds.), *Procedia Eng.* (Vol. 78, pp. 224–228). Elsevier Ltd. Consulté sur <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84914107377&doi=10.1016/j.proeng.2014.07.060&partnerID=40&md5=c01f967249aafaeb803ba97f931499ad> (Journal Abbreviation : Procedia Eng.) doi : 10.1016/j.proeng.2014.07.060
- Seaver, N. (2017, décembre). Algorithms as culture : Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2), 2053951717738104. Consulté le 2021-09-01, sur <https://doi.org/>

- 10.1177/2053951717738104 (Publisher : SAGE Publications Ltd) doi : 10.1177/2053951717738104
- Selwyn, B. (2015, mars). Commodity chains, creative destruction and global inequality : a class analysis. *Journal of Economic Geography*, 15(2), 253–274. Consulté le 2021-02-05, sur <https://doi.org/10.1093/jeg/lbu014> doi : 10.1093/jeg/lbu014
- Selwyn, B. (2019, février). Poverty chains and global capitalism. *Competition & Change*, 23(1), 71–97. Consulté le 2023-11-03, sur <https://doi.org/10.1177/1024529418809067> (Publisher : SAGE Publications Ltd) doi : 10.1177/1024529418809067
- Sengul-Jones, M. (2017). 'Being a Better #Freelancer' : Gendered and Racialised Aesthetic Labour on Online Freelance Marketplaces. In A. S. Elias, R. Gill, & C. Scharff (Eds.), *Aesthetic Labour* (pp. 215–229). London : Palgrave Macmillan UK. Consulté le 2020-07-20, sur http://link.springer.com/10.1057/978-1-137-47765-1_12 doi : 10.1057/978-1-137-47765-1_12
- SHESTAKOFSKY, B. (2015). MORE MACHINERY, LESS LABOR? *Berkeley Journal of Sociology*, 59, 86–91. Consulté sur <http://www.jstor.org/stable/44713549> (Publisher : Regents of the University of California)
- Shestakofsky, B. (2017, novembre). Working Algorithms : Software Automation and the Future of Work. *Work and Occupations*, 44(4), 376–423. Consulté le 2021-05-03, sur <https://doi.org/10.1177/0730888417726119> (Publisher : SAGE Publications Inc) doi : 10.1177/0730888417726119
- Simonet, M. (2018). *Travail gratuit : la nouvelle exploitation ?* textuel. (Google-Books-ID : uQRuDwAAQBAJ)
- Simut, A. (2017). Contemporary Representations of Artificial Intelligence in Science Fiction Films, Visual Arts, and Literature. A Short Introduction. *Ekphrasis. Images, Cinema, Theory, Media*, 17(1), 5–8. Consulté le 2022-02-10, sur <https://www.ceeol.com/search/article-detail?id=543833> (Publisher : Universitatea Babeş-Bolyai, Facultatea de Teatru si Televiziune)
- Smyrnaiois, N., & Marty, E. (2017). Profession « nettoyeur du net ». De la modération des commentaires sur les sites d'information français. *Réseaux*, 205(5), 57–90. Consulté le 2023-01-24, sur <https://www.cairn.info/revue-reseaux-2017-5-page-57.htm> (Place : Paris Publisher : La Découverte) doi : 10.3917/res.205.0057
- Soares, A. (2002, octobre). Les qualifications invisibles dans le secteur des services : le cas des caissières de supermarchés. *Lien social et Politiques*(40), 105–116. Consulté le 2023-11-04, sur <http://id.erudit.org/iderudit/005239ar> doi : 10.7202/005239ar
- Soares, A. S. (1991). Work Organization in Brazilian Data Processing Centres : Consent and Resistance. *Labour, Capital and Society / Travail, capital et société*, 24(2), 154–183. Consulté le 2022-04-08, sur <https://www.jstor.org/stable/43157931> (Publisher : Labour, Capital & Society)
- Star, S. L. (2010, juillet). Ceci n'est pas un objet-frontière! *Revue d'anthropologie des connaissances*, Vol 4, n° 1(1), 18–35. Consulté le 2020-03-31, sur <https://www.cairn.info/revue-anthropologie-des-connaissances-2010-1-page-18.htm> (Publisher : S.A.C.)

- Star, S. L., & Strauss, A. (1999, mars). Layers of Silence, Arenas of Voice : The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work (CSCW)*, 8(1), 9–30. Consulté le 2023-11-14, sur <https://doi.org/10.1023/A:1008651105359> doi : 10.1023/A:1008651105359
- Stroobants, M., & Singly, F. d. (2016). *Sociologie du travail* (4e éd éd.). Malakoff : Armand Colin.
- Tanguy, L. (2011). 1. La fondation de l'ISST : une conjonction d'actions émanant de milieux universitaires et politiques. In *La sociologie du travail en France* (pp. 29–39). Paris : La Découverte. Consulté le 2022-08-25, sur <https://www.cairn.info/la-sociologie-du-travail-en-france--9782707164308-p-29.htm>
- Tine, B. (2018). *Sociologie des centres d'appels : usages sociaux de l'emploi et défis culturels des délocalisations en Afrique* (Thèse de doctorat non publiée). Connaissances et savoirs, Saint-Denis. (ISBN : 9782753905801 Series : Collection TIC (Technologie de l'information et de la communication) et développement)
- Tine, B. (2021). Ethnographie des centres d'appels au Sénégal : des temporalités en crise ? *Sciences & Actions Sociales*, 14(1), 89–106. Consulté le 2023-01-18, sur <https://www.cairn.info/revue-sciences-et-actions-sociales-2021-1-page-89.htm> (Publisher : Association des Chercheurs des Organismes de la Formation et de l'Intervention Sociales) doi : 10.3917/sas.014.0089
- Tubaro, P. (2021, janvier). Disembedded or Deeply Embedded? A Multi-Level Network Analysis of Online Labour Platforms. *Sociology*, 0038038520986082. Consulté le 2021-04-23, sur <https://doi.org/10.1177/0038038520986082> (Publisher : SAGE Publications Ltd) doi : 10.1177/0038038520986082
- Tubaro, P., & Casilli, A. A. (2019, septembre). Micro-work, artificial intelligence and the automotive industry. *Journal of Industrial and Business Economics*, 46(3), 333–345. Consulté le 2020-03-11, sur <https://doi.org/10.1007/s40812-019-00121-1> doi : 10.1007/s40812-019-00121-1
- Tubaro, P., Casilli, A. A., & Coville, M. (2020a, janvier). The trainer, the verifier, the imitator : Three ways in which human platform workers support artificial intelligence. *Big Data & Society*, 7(1), 205395172091977. Consulté le 2020-07-21, sur <http://journals.sagepub.com/doi/10.1177/2053951720919776> doi : 10.1177/2053951720919776
- Tubaro, P., Casilli, A. A., & Coville, M. (2020b, avril). The trainer, the verifier, the imitator : Three ways in which human platform workers support artificial intelligence :. *Big Data & Society*. Consulté le 2020-07-24, sur <https://journals.sagepub.com/doi/10.1177/2053951720919776> (Publisher : SAGE PublicationsSage UK : London, England) doi : 10.1177/2053951720919776
- Tubaro, P., Coville, M., Le Ludec, C., & Casilli, A. A. (2022). Hidden inequalities : The gendered labour of women on micro-tasking platforms. *Internet Policy Review*, 11(1), 1–26. Consulté le 2022-08-10, sur <https://www.econstor.eu/handle/10419/254273> (Publisher : Berlin : Alexander von Humboldt Institute for Internet and Society) doi : 10.14763/2022.1.1623
- Tubaro, P., Le Ludec, C., & Casilli, A. A. (2020). Counting 'micro-workers' : societal and methodological

- challenges around new forms of labour. *Work Organisation, Labour & Globalisation*, 14(1), 67–82. Consulté le 2020-09-23, sur <https://www.jstor.org/stable/10.13169/workorgalaboglob.14.1.0067> (Publisher : Pluto Journals) doi : 10.13169/workorgalaboglob.14.1.0067
- Uzzi, B. (1997). Social Structure and Competition in Interfirm Networks : The Paradox of Embeddedness. *Administrative Science Quarterly*, 42(1), 35–67. Consulté le 2021-02-08, sur <https://www.jstor.org/stable/2393808> (Publisher : [Sage Publications, Inc., Johnson Graduate School of Management, Cornell University]) doi : 10.2307/2393808
- Van Belle, J.-P., Bonne, K., Cocquyt, I., & Garbutt, M. (2019). Attitudes of Belgian Companies towards Outsourcing Microwork to Africa. In *Proceedings of 4th International Conference on the* (Vol. 12, pp. 382–392).
- van der gaag, M., Appelhof, G., & Webber, M. (2012, janvier). Ambiguities in responses to the Position Generator. *Sociologia e Politiche Sociali*, 15, 113–141. doi : 10.13140/2.1.3009.3444
- Van der Gaag, M., & Snijders, T. (2004). Proposals for the measurement of individual social capital. In *Creation and returns of social capital* (pp. 172–187). Routledge.
- Van Der Gaag, M., & Webber, M. (2008). Measurement of Individual Social Capital. In I. Kawachi, S. Subramanian, & D. Kim (Eds.), *Social Capital and Health* (pp. 29–49). New York, NY : Springer New York. Consulté le 2022-08-25, sur http://link.springer.com/10.1007/978-0-387-71311-3_2 doi : 10.1007/978-0-387-71311-3_2
- Vayre, J.-S. (2021, mars). Pour une critique plurielle de l’intelligence artificielle. *Revue d’anthropologie des connaissances*, 15(1). Consulté le 2021-07-12, sur <https://journals.openedition.org/rac/19266> (Number : 1 Publisher : Société d’Anthropologie des Connaissances)
- Veale, M., & Borgesius, F. Z. (2021, août). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), 97–112. Consulté le 2022-07-12, sur <http://arxiv.org/abs/2107.03721> (arXiv :2107.03721 [cs]) doi : 10.9785/cri-2021-220402
- Velkovska, J., & Zouinar, M. (2020). Les relations aux machines « conversationnelles ». Vivre avec les assistants vocaux à la maison. *Réseaux*, 220-221(2-3), 47–79. Consulté le 2023-08-03, sur <https://www.cairn.info/revue-reseaux-2020-2-3-page-47.htm> (Place : Paris Publisher : La Découverte) doi : 10.3917/res.220.0047
- Verhaeghe, P.-P., Van de Putte, B., & Roose, H. (2013, août). Reliability of Position Generator Measures across Different Occupational Lists A Parallel Test Experiment. *Field Methods*, 25, 238–261. doi : 10.1177/1525822X12453227
- Villani, C. (2018). *Donner un sens à l’intelligence artificielle : pour une stratégie nationale et européenne* (Rapport technique). Paris.
- Wang, Y., & Kosinski, M. (2017, février). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. Consulté le 2021-08-16, sur <https://osf.io/zn79k/>

(Publisher : OSF) doi : 10.17605/OSF.IO/ZN79K

- Webb, M. A., & Tangney, J. P. (2022, novembre). Too Good to Be True : Bots and Bad Data From Mechanical Turk. *Perspectives on Psychological Science*, 17456916221120027. Consulté le 2023-02-01, sur <https://doi.org/10.1177/17456916221120027> (Publisher : SAGE Publications Inc) doi : 10.1177/17456916221120027
- Weber, F. (2009). *Le travail à-côté : une ethnographie des perceptions* (Nouvelle éd. revue et augm éd.). Paris : Éditions de l'École des hautes études en sciences sociales.
- Webster, J. (2016). Microworkers of the Gig Economy. *New Labor Forum*, 25(3), 56–64. Consulté le 2020-04-28, sur <https://www.jstor.org/stable/26420011> (Publisher : Sage Publications, Inc.) doi : 10.2307/26420011
- Williamson, V. (2016). *Can crowdsourcing be ethical?* Consulté le 2022-08-24, sur <https://www.brookings.edu/blog/techtank/2016/02/03/can-crowdsourcing-be-ethical-2/>
- Wilson, J. A., & Yochim, E. C. (2017). *Mothering through Precarity : Women's Work and Digital Media*. USA : Duke University Press.
- Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2018, août). Good Gig, Bad Gig : Autonomy and Algorithmic Control in the Global Gig Economy. *Work, Employment and Society*, 33(1), 56–75. Consulté le 2020-04-29, sur <https://doi.org/10.1177/0950017018785616> (Publisher : SAGE Publications Ltd) doi : 10.1177/0950017018785616
- Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019a, février). Good Gig, Bad Gig : Autonomy and Algorithmic Control in the Global Gig Economy. *Work, Employment and Society*, 33(1), 56–75. Consulté le 2020-03-27, sur <https://doi.org/10.1177/0950017018785616> (Publisher : SAGE Publications Ltd) doi : 10.1177/0950017018785616
- Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019b, février). Networked but Commodified : The (Dis)Embeddedness of Digital Labour in the Gig Economy. *Sociology*, 0038038519828906. Consulté le 2019-04-06, sur <https://doi.org/10.1177/0038038519828906> doi : 10.1177/0038038519828906
- Woolgar, S. (1985, novembre). Why not a Sociology of Machines? The Case of Sociology and Artificial Intelligence. *Sociology*, 19(4), 557–572. Consulté le 2022-10-11, sur <http://journals.sagepub.com/doi/10.1177/0038038585019004005> doi : 10.1177/0038038585019004005
- Yin, M., Gray, M. L., Suri, S., & Vaughan, J. W. (2016). The Communication Network Within the Crowd. In *Proceedings of the 25th International Conference on World Wide Web - WWW '16* (pp. 1293–1303). Montré#233;al, Qu#233;bec, Canada : ACM Press. Consulté le 2020-09-18, sur <http://dl.acm.org/citation.cfm?doid=2872427.2883036> doi : 10.1145/2872427.2883036
- Zarifian, P. (1983). Qualification collective et automatisation : le cas de la sidérurgie. *Formation Emploi*, 1(1), 37–51. Consulté le 2023-08-02, sur https://www.persee.fr/doc/forem_0759-6340_1983_num_1_1_1023 (Publisher : Persée - Portail des revues scientifiques en SHS) doi : 10.3406/forem.1983.1023

Zuboff, S. (1988). *In the age of the smart machine : the future of work and power*. New York : Basic Books.

Annexes

A Analyses qualitatives

A.1 Description des enquêtés

A.1.1 Les employés de start-ups

Nom anonyme ¹	Rôle	Date	Entreprise anonymisées
Carole	Responsable des opérations	07/02/2020	MLDupli
Jeanne	Data analyst	18/06/2020	MLDupli
Claire	Sales representative	22/07/2020	MLArchitecture
Arthur	Chargé de communication	19/01/2021	CityVeillance
Julien	Chief entreprise officer	26/01/2021	Automatik
Patrick	Directeur R&D	28/01/2021	CityVeillance
François	Chief entreprise officer	05/03/2021	Clik
Lucine	Directrice AI	18/03/2021	AIVoice
Eleonor	Business development	02/04/2021	AIVoice
Corentin	Ingénieur	14/04/2021	Automatik
Lin	Responsable datasets	02/05/2021	Automatik
Yoann	Chief entreprise officer	11/05/2021	CallAI
Thibault	Chief entreprise officer	19/05/2021	Papers
Guillaume	Chief entreprise officer	19/05/2021	LegalAI
Simon	Product manager	01/07/2021	CCenter
Cédric	Chief Data Officer	25/08/2021	AI Medic
Pauline	Responsable RSE	02/09/2021	Automatik
Bastien	Product manager	03/09/2021	Automatik
Théophile	Chief entreprise officer	06/09/2021	Automatik
Mathilde	Responsable datasets	11/09/2021	BankAI
Clément	Manager annotation	03/11/2021	Clik
Nicolas	Chief Technical Officer	07/11/2021	AI Medic
Arnaud	Ingénieur logiciel	03/11/2021	GitAI
Audrey	Solution Architect	14/12/2021	Automatik
Antoine	Chief entreprise officer	16/11/2021	ModeAI
Jean	Chief entreprise officer	16/11/2021	ModeAI

TABLE 35 – Description des personnes enquêtées au sein des entreprises d'IA

Type	Date	Entreprise anonymisées
Webinar	21/01/2020	AIVoice
Observation en présentiel	09/03/2021	Automatik

TABLE 36 – Observations

A.1.2 Les employés d’entreprises d’externalisation

Nom anonyme	Rôle	Date	Entreprise anonymisées
Dama	Chief entreprise officer	02/01/2021	VirtualAssistant
Loik	Customer Satisfaction officer	26/01/2021	E
Elisoa	Chief entreprise officer	01/02/2021	MicroWorkTana
Dinasoa	Chief entreprise officer	16/02/2021	E
Victor	Chief operations officer	18/03/2021	B
Patrick	Chief entreprise officer	02/04/2021	E
Pierre	Chief entreprise officer	21/05/2021	G
Frédéric	Chief entreprise officer	10/09/2021	C
Eva	Chief entreprise officer	22/04/2022	A
Tsyki	RH	22/04/2022	C
Jean	Vice Président	17/05/2022	C
Quentin	Responsable annotation	19/05/2022	D
Alexandre	Manager	19/05/2022	D
Léo	Responsable annotation	01/08/2022	C
Axelle	Responsable annotation	01/08/2022	C

TABLE 37 – Description des personnes enquêtées au sein des entreprises d’annotation de données

A.1.3 Les entreprises d’externalisation étudiées et les modalités de contrat

A.2 Codage des entretiens

Niveau 1	Niveau 2	Définition
Travail de l’IA	Numérisation	Le travail du clic comme contribution à la numérisation des services

	Problématisation	Le travail du clic comme processus de "problématisation de l'IA"
	Classification	Le travail du clic comme processus de classification de données
	Projets	Description des projets d'IA
<hr/>		
Gestion travail des données	Outils	L'annotation repose sur des outils de production
	Processus	L'organisation de l'annotation
	Qualité	Les processus de contrôle qualité
	Discussion	Les discussions des annotations
	Documents	L'annotation repose sur des documents qui permettent d'inscrire les discussions
	Formation	La formation des annotateurs
<hr/>		
Conditions de travail	Par entreprise d'externalisation	Les conditions de travail par entreprise
	Salaire	Le rapport au salaire
	Travail plateformes	Les conditions spécifiques aux plateformes
	Précarité	Les difficultés des annotateurs par rapport à la structure de l'industrie
	Difficultés annotation	Les difficultés des annotateurs par rapport au travail au quotidien
	Identité professionnelle	La manière dont les annotateurs se définissent
	Carrière	La carrière et le devenir professionnel des annotateurs

TABLE 38 – Codage pour les travailleurs des données

Question	Codage	Définition
Raisons externalisation		

Gestion travail des données	Arbitrage solutions d'externalisation	Expliquer le choix entre plusieurs pays ou type de services
	Normes	Les normes juridiques qui s'appliquent au secteur
	Attraction	La stratégie d'attraction des IDE de Madagascar
	Expertise	Ces entreprises bénéficient de compétences sur place qu'ils vont ensuite former à l'annotation
	Post-colonial	Le lien avec la France
	Télécom	Le développement de l'Internet puis de la fibre à Madagascar
	Outils	L'annotation repose sur des outils de production
	Processus	L'organisation de l'annotation
	Qualité	Les processus de contrôle qualité
	Discussion	Les discussions des annotations
Documents	L'annotation repose sur des documents qui permettent d'inscrire les discussions	
Formation	La formation des annotateurs	
Coût	La structure des coûts de l'annotation	

TABLE 39 – Codage pour les dirigeants des entreprises d'annotation et les salariés d'entreprises d'IA

Question	Codage	Définition
Processus de production	Accès données	Les moyens d'accéder à des données à annoter
	Adaptation contexte clients	Produire des données annotées spécifiques aux clients
	Remontée annotateurs modèles	Modification des modèles selon les retours des travailleurs de Madagascar

Structure chaîne de valeur	Documents	Les entreprises doivent définir et discuter des standards d'annotation
	Outils	Les entreprises doivent définir et discuter des outils d'annotation et de discussions
	Client	Pour qui travaillent les entreprises d'IA
	Computation	Elles s'appuient sur des modèles déjà entraînés et proposés par les GAFAM
	Importance travail données	Elles soulignent l'importance du travail d'annotation

TABLE 40 – Codage pour les salariés des entreprises d'IA

B Passation du questionnaire

B.1 Exemple de négociation d'entrée de terrain avec des entreprises d'annotations

Bonjours, bon je prends le temps de vous écrire quant meme j'aime bien les furteurs ingénieurs

Cher étudiant (à priori sans but détourné)

On ne se comprend pas bien ce n'est pas l'étude que nous facturons mais le temps pris par l'expert pour répondre à vos questions et encore on fait un effort on a pas dit oui

Je pensais que vous étiez financé des deux côtés (d'ailleurs pas qui ? votre école ou tartempion le concurrent) : votre temps et le temps qu'un expert va prendre pour répondre (non rémunéré du côté expert si je comprend bien la démarche).

il y a toujours un risque de fuite , le risque zéro ça n'existe pas sauf dans les fables pour enfants

Ce que vous nous demandez la c'est de vous renseignez en clair gratuitement sur du savoir faire technologique, le fonctionnement interne d'une boite comme la notre , une expérience sur le marché et ses process et ses acteurs pour une thèse qui peut parfaitement être récupérée par des concurrents potentiels ou réels friand d'infos gratuites, cela peut paraître paranoïaque mais vous connaissez le dicton "only the paranoid survive" tout cela pour ? Nos beaux yeux ? je vais vous dire même à 15000 EUR c'est un deal médiocre je dirais même à 100 000 car franchement donneriez vous un savoir faire technique ou une vision claire du marché à un concurrent pour 15 000 ? 100 000 ? A combien valoriser vous la vision sur un marché ? Les gens en général pensent que l'info est gratuite, l'info de qualité est ce qui coûte LE PLUS CHÈRE et une information d'un expert d'un secteur est hors de prix; comme vous êtes jeunes , voyez cet extrait de film c'est une boutade, étant de bonne humeur aujourd'hui.

<https://www.youtube.com/watch?v=n5iwjFSpUw8>

the most valuable commodity I know is ... information

Votre demande d'étude arrive en plus de cela 3 mois après que deux cabinets internationaux nous ait approchés pour "bizarement " (je dirais que je crois au hasard dans 1 % des cas un peu pres) exactement la même chose avec leurs gros sabots, bon eux nous proposait une "publication" dans un journal de BPO afin de nous faire un peu moussé (on n'aime pas trop la bière) , nous avons appris que c'était un concurrent (enfin j'ai poussé pour le savoir car cela m'intriguait) à travers plusieurs sociétés qui avaient en réalité commandé l'étude en question à plusieurs cabinets : nous avons bien évidemment refusé catégoriquement. Vous comprendrez donc notre grande méfiance voir notre agacement sur ce genre de démarche.

Si c'est juste un questionnaire, faut voir déjà le questionnaire et les questions mais bon à priori ... je vous souhaite bonne chance avec d'autres sociétés qui peut être seront plus entrain à vous répondre (généralement ceux qui ne savent rien à rien) conclusion : plus on sait , plus on se tait

Si par contre vous arrivez a faire l'étude et bien on serait preneur contre une petite rémunération :) il est toujours bien de connaitre la vision de la concurrence

Excellente journée

FIGURE 51 – Mail de réponse à une prise de contact pour diffuser le questionnaire

B.2 Brief envoyé aux annotateurs en vue de la réponse à l'enquête

”Nous sommes deux chercheurs français travaillant dans des institutions de recherche publiques. Notre but est de faire progresser la recherche scientifique en Sociologie, en collectant des données sur vous et sur votre travail d’annotateur (via le questionnaire et l’entretien). L’idée est de vous donner la parole, car c’est vous qui connaissez le mieux le sujet.

(Izahay dia mpikaroka frantsay anankiroa miasa amina sampam-pikarohana ho an’ny daholobe. Ny tanjonay dia ny hampivoatra ny fikarohana siantifika ara-Sosiolojia, amin’ny fanangonana ny mombamomba anao sy ny asanao amin’ny maha “annotateur” anao (amin’ny alalan’ily fanadihadiana sy ny tafatafa). Ny hevitra dia ny hanome ny fitenenana anao satria ianao no mahafantatra kokoa ny lohahevitra.)

Pourquoi cette enquête ? (Ny anton’ ity fanadihadiana ity ?)

Car il n’y a pas beaucoup d’enquêtes sur l’annotation de données et plus généralement sur travail du numérique à Madagascar. En fait, alors que vous participez à l’économie mondiale et au développement d’une technologie innovante, ce que vous faites et qui vous êtes est assez peu connu, même des acteurs du numérique.

(Satria tsy misy fanadihadiana betsaka mahakasika ny fikirakiràna ny angona sy ny asa momba ny nomerika amin’ny ankapobeny eto Madagasikara. Raha ny tena marina, mandritry ny fandraisanao anjara amin’ny toe-karen’izao tontolo izao sy ny fampandrosoana teknolojia vaovao, ny zavatra ataonao sy ny hoe iza ianao dia tsy fantatra firy, na dia ny mpisehatra eo amin’ny nomerika aza.)

Qu’est-ce qu’une enquête ? Quel est votre rôle ? (Inona no atao hoe fanadihadiana ? Inona no anjara asanao ?)

Le questionnaire (Ny fanadihadiana) Il nous permet d’avoir une idée générale de qui vous êtes, de votre parcours, et de votre travail au quotidien. Le but est de répondre aux questions, le plus fidèlement possible, même quand vous n’êtes pas sûr de maîtriser tous les aspects de la question. C’est votre point de vue personnel qui nous intéresse.

(Ity dia ahafahanay manana hevitra amin’ny ankapobeny momba anao, ny lalana nodiavinao, sy ny asanao amin’ny andavan’andro. Ny tanjona dia ny hamaliana ny fanontaniana, amim-pahatokiana araka izay azo atao, na dia tsy azonao antoka azan y fahafehezanao ny lafiny rehetra amin’ily

fanontaniana. Ny fomba fijerinao no tena mahaliana anay.)

Ce questionnaire dure environ 30 minutes, vous pourrez y répondre en suivant le lien suivant :
(Ity fanadihadiana ity dia maharitra eo amin'ny 30 minitra eo, afaka mamaly izany ianao manaraka ity rohy ity :) [Lien du questionnaire]

Secret des données (Tsiambaratelon'ny angon-drakitra)

Votre identité sera gardée secrète et seuls les membres de l'équipe de recherche auront accès aux données. A aucun moment ce que vous nous direz en entretien ou via le questionnaire ne sera communiqué tel quel à qui que ce soit. Toutes les données seront anonymisées, et seul des rapports synthétiques, décrivant des tendances générales, et où vos noms n'apparaissent pas, seront publiés. Les données seront stockées sur des serveurs sécurisées. Votre participation ne comporte donc aucun risque.

(Ny mikasika anao dia ho voatazona ho tsiambaratelo ary ny hany mahita izany dia ny mpikambana ato amin'ny ekipan'ny sampam-pikarohana ireny ihany. Tsy misy na fotoana iray aza ka hampitana tahak'izay voalazanao tamin'ny fanadihadiana sy ny tafatafa, izay nambaranao na amin'iza na amin'iza. Ny angon-drakitra rehetra dia tsy ampahafantarina anarana ary ny tati-tratra voafintina, mamaritra ny fironana ankapobeny, sy tsy misy anaranao mipoitra no avoaka. Ny momba anao dia angonina amina « serveur » voaro. Ny fandraisanao anjara izany dia azo antoka tanteraka ary tsy misy atahorana.)”

B.3 Annonce de recrutement publiée sur Facebook

”Pour le compte d'un organisme public de recherche se trouvant à Paris, nous cherchons des personnes travaillant ou ayant travaillé dans le milieu du BPO (annotation de données, modération de contenus, saisie de données) afin de remplir un questionnaire et éventuellement participer à un entretien pour en savoir plus sur le métier et ce qui vous a poussé à rejoindre ce secteur d'activité. A noter que nous ciblons plus précisément les annotateurs ou anciens annotateurs mais les autres métiers se trouvant dans le domaine seront triés selon leur CV.

Nous vous invitons à nous envoyer vos CV au préalable afin que nous puissions planifier les entretiens.”

B.4 Le générateur de position avec les scores associés

Professions	ISCO88_3d	Score
High		
Médecins	222	78,57
Instituteur	233	63,79
Chef d'entreprise	121	66,87
Professeur à l'université	2310	73
Digital		
Développeur logiciel web	213	75,13
Data analyst, data scientist	213	75,13
Consultant informatique	213	75,13
Middle		
Manager dans un hôtel ou restaurant	1225	58,52
Manager dans l'agriculture	1311	41,55
Manager dans la pêche	1311	41,55
Manager dans l'industrie (par exemple : mine)	1222	60,13
Manager dans les services ¹	122	62,86
Commerçants	131	57,27
Opérateur de saisie de données	4113	59,26
Low		
Taxi	832	31,22
Employés de commerce (par exemple : vendeur)	522	45,26
Ouvrier agricole	613	36,35
Ouvrier dans l'industrie (par exemple : mine)	721	33,3
Ouvrier de la pêche	613	36,35
Ouvrier du textile	826	30
Ouvrier de la manutention	9333	29,48
Employés d'hôtellerie, de restauration et d'entretien ²	913	26,71

¹Par exemple : dans une entreprise de saisie de données

²Par exemple : serveurs, employés d'étage des hôtels, concierge

TABLE 41 – Professions du générateur de position et scores associés

B.5 Questionnaire

QDictionnaire des variables

Ordre : Code variable / Question / Modalités.

Partie A :

Q1.Nom Q1. Iza no anaranao? Quel est votre nom? Champ libre.

Q2.Lieu_naiss

Q2. Aiza ny komina na tanàna nahaterahanao? Quelle est votre commune ou village de naissance? Champ libre.

[Q2_2 à Q8 dans Questionnaire A et B uniquement]

Q2_2.Region_naiss

Q2. Taizany faritra nahaterahanao? Quelle est votre région de naissance?

Q2_2_autre.Region_naiss_autre

Q2[other]. Taiza ny faritra nahaterahanao? Quelle est votre région de naissance? [Autre]

Champ libre.

Q3.Lieu_hab

Q3. Aiza ianao no monina ankehitriny? Où vivez-vous actuellement?

Pas dans questionnaire A.

Q3_autre.Lieu_hab_autre

Q3[other]. Aiza ianao no monina ankehitriny? Où vivez-vous actuellement? [Autre]

Pas dans questionnaire A.

Q4.Type_lieu_hab

Q4. Ianao dia mipetraka amina : Vous habitez dans : Pas dans questionnaire A.

Q5.Tjs_vecu_ici

Questionnaire A Q5. Ianao ve nonina foana tao {Q3.shown}? Avez vous toujours vécu à {Q3.shown}?

[questionnaire A] Q3. Avez vous toujours vécu dans ou à côté d'Antananarivo?

Q6.Lieu_hab_precedent

Questionnaire A Q6. Talohan'ny nahatongavanao teto {Q3.shown}, ianao dia nipetraka tao : Avant d'arriver à {Q3.shown}, vous habitiez dans :

[questionnaire A] Q4. Avant d'arriver à Antananarivo, vous habitiez dans :

Q6_autre.Lieu_hab_precedent_autre

Questionnaire A Q6[other]. Talohan'ny nahatongavanao teto {Q3.shown}, ianao dia nipetraka tao : Avant d'arriver à {Q3.shown}, vous habitiez dans :

[questionnaire A] Q4[other]. Avant d'arriver à Antananarivo, vous habitiez dans : [Autre]

Q7.Region_precedente

Questionnaire A Q7. Tondroy ny faritra nonenanao farany : Indiquez la région de votre dernier lieu de vie :

[questionnaire A] Q5. Indiquez la région de votre dernier lieu de vie :

Q8_1.Raison_venue_famille

[Questionnaire A] Q8[SQ001]. Inona ny antony nandehananao tao {Q3.shown}? Pourquoi êtes-vous venu à {Q3.shown}? [Mba hanaraka ny fianakaviana (Pour suivre ma famille)] [questionnaire A] Q6[SQ001]. Pourquoi êtes-vous venu à Antananarivo? [Pour des raisons familiales]

Q8_2.Raison_venue_educ

Questionnaire A Q8[SQ002]. Inona ny antony nandehananao tao {Q3.shown}? Pourquoi êtes-vous venu à {Q3.shown}? [Mba hanohy fianarana (Pour poursuivre vos études)]

[questionnaire A] Q6[SQ002]. Pourquoi êtes-vous venu à Antananarivo? [Pour continuer votre éducation]

Q8_3.Raison_venue_emp

Questionnaire A Q8[SQ003]. Inona ny antony nandehananao tao {Q3.shown}? Pourquoi êtes-vous venu à {Q3.shown}? [Mba hahita asa (Pour trouver du travail)]

[questionnaire A] Q6[SQ003]. Pourquoi êtes-vous venu à Antananarivo? [Pour trouver un emploi]

Q8_4.Raison_venue_mutat

Pas dans .

[questionnaire A] Q6[SQ004]. Pourquoi êtes-vous venu à Antananarivo ? [Pour des raisons professionnelles (mutation)]

Q8_autre.Raison_venue_autre

Questionnaire A Q8[other]. Inona ny antony nandehanao tao {Q3.shown}? Pourquoi êtes-vous venu à {Q3.shown}? [Autre]

[questionnaire A] Q6[other]. Pourquoi êtes-vous venu à Antananarivo ? [Autre]

[DE NOUVEAU DANS LES 4 QUESTIONNAIRES]

Q9.Genre

[Q7 dans questionnaire A]

Q9. Ianao dia : Vous êtes :

Lehilahy (Un homme)

Vehivavy (Une femme)

Q10.Age

[Q8 dans questionnaire A]

Q10. Ianao dia : Vous avez :

18-24 taona (18-24 ans)

25-34 taona (25-34 ans)

35-44 taona (35-44 ans)

45-55 taona (45-55 ans)

55-64 taona (55-64 ans)

64-75 taona (64-75 ans)

Maherin'ny 75 taona (Plus de 75 ans)

Q11.Situation

[Q9 dans questionnaire A -> à chaque fois retirer 2 pour avoir la question correspondante]

Q11. Ianao dia : Vous êtes :

Mpitovo (Célibataire)

Maty vady (Veuf, veuve)

Miaraka amin'olona (En couple) + Concubin(e)

Manambady (Marié(e))

Nisara-panambadiana (Divorcé(e))

Tsy te hamaly aho (Je ne souhaite pas répondre)

Autre

Q11.Situation_autre

Q11[other]. Ianao dia : Vous êtes : [Autre]

Champ libre.

« Mère célibataire » => « Célibataire »

« Fiancé » => « Marié »

Q12.Conj_niv_emp

Q12. Inona no sokajin'asan'ny vadinao ? Quel est le niveau d'emploi de votre conjoint ?

Tompon'andraikitra ambony (talé), injeniera na mitovy lenta amin'izany (Cadre, ingénieur ou équivalent)

Tompon'andraikitra (sefo) (Responsable, agent de maîtrise)

Mpiasa voahofana manokana (Ouvrier spécialisé)

Mpiasa tsy nahazo fanofanana (Ouvrier sans formation)

Mpiasa tsotra (ohatra manoeuvre) (Agent (par exemple : manoeuvre))

Mpandraharaha (Entrepreneur)

Manao asa-tena (Profession libérale)

Mpiana-draharaha (Stagiaire)

Miasa antranon'olona (Employé de maison)

Mikarakara tokantrano (Père / Mère au foyer)

Autre

Q12.Conj_niv_emp_autre

Q12[other]. Inona no sokajin'asan'ny vadinao? Quel est le niveau d'emploi de votre conjoint? [Autre]

Champ libre.

Q13.Conj_secteur

Q13. Sehatr'asa inona? Dans quel secteur d'activité?

Fambolena sy jono (Agriculture et pêche)

Fitrandrahana (ohatra : harena ankibon'ny tany) (Activités d'extraction (par exemple : mines))

Indostria (ohatra : lamba) (Activités de fabrication (par exemple : textiles))

Fanaovana asa vaventy (Construction)

Varotra (Commerce)

Hotely, fisakafoanana, bar (Hôtels, restaurants, bars)

Fitanterana sy fifandraisan-davitra (Transports et télécommunications)

Informatika sy teknolojia (Activités informatiques (par exemple : développeur))

Fandraharahana ara-bola (ohatra : miasa amin'ny banky) (Activités financières (par exemple : agent d'une banque))

Fandraharana momban'ny trano na tolotr'asa hoan'ireo orinasa (Immobilier, location et services rendus

aux entreprises)

Fahasalamana (Santé)

Fampianarana (Enseignement)

Fitantan-draharaham-panjakana (Administration publique)

Autre

Q13.Conj_secteur_autre

Pas dans questionnaire A

Q13[other]. Sehatr'asa inona? Dans quel secteur d'activité? [Autre]

Champ libre.

« Assurance » => « Activités financières »

« Bijouter » => « Commerce »

« Chargé avant vente » => « Commerce »

« modérateur vidéo francophone » => « Activités informatiques »

« Mpampianatra » => « Enseignement »

« offshoring » => « Commerce »

« Rédactrice web » => « Activités informatiques »

« Sécurité publique » => « Administration publique »

« vente en ligne » => « Commerce »

Q14.Conj_niv_etude

Q14. Inona no fari-pahaizan'ny vadinao? Quel est le niveau d'étude de votre conjoint?

Hatreo amin'ny CEPE (Primaire)

Hatreo amin'ny BEPC (C.E.G)

Hatreo amin'ny Bacc (Lycée)

Sokajy teknika na sekoly manofana arak'asa (Niveau secondaire technique ou « professionnel »)

Ambaratonga ambony (Oniversite) (Supérieur)

Q15.Pere_niv_emp

Q15. Inona no sokajin'asan'ny dadanao? Quel est le niveau d'emploi de votre père?

Tompon'andraikitra ambony (talé), injeniera na mitovy lenta amin'izany (Cadre, ingénieur ou équivalent)

Tompon'andraikitra (sefo) (Responsable, agent de maîtrise)

Mpiasa voahofana manokana (Ouvrier spécialisé)

Mpiasa tsy nahazo fanofanana (Ouvrier sans formation)

Mpiasa tsootra (ohatra manoeuvre) (Agent (par exemple : manoeuvre))

Mpandraharaha (Entrepreneur)

Manao asa-tena (Profession libérale)

Mpiana-draharaha (Stagiaire)

Miasa antranon'olona (Employé de maison)

Mikarakara tokantrano (Père / Mère au foyer)

Autre

Q15.Pere_niv_emp_autre

Q15[other]. Inona no sokajin'asan'ny dadanao? Quel est le niveau d'emploi de votre père? [Autre]

Champ libre.

« Administrateur Civile » => « Cadre, ingénieur ou équivalent »

« Avocat » => « Profession libérale »

« Charpentier »

Q16.Pere_secteur

Q16. Sehatr'asa inona? Dans quel secteur d'activité?

Fambolena sy jono (Agriculture et pêche)

Fitrandrahana (ohatra : harena ankibon'ny tany) (Activités d'extraction (par exemple : mines))

Indostria (ohatra : lamba) (Activités de fabrication (par exemple : textiles))

Fanaovana asa vaventy (Construction)

Varotra (Commerce)

Hotely, fisakafoanana, bar (Hôtels, restaurants, bars)

Fitanterana sy fifandraisan-davitra (Transports et télécommunications)

Informatika sy teknolojia (Activités informatiques (par exemple : développeur))

Fandraharahana ara-bola (ohatra : miasa amin'ny banky) (Activités financières (par exemple : agent d'une banque))

Fandraharana momban'ny trano na tolotr'asa hoan'ireo orinasa (Immobilier, location et services rendus aux entreprises)

Fahasalamana (Santé)

Fampianarana (Enseignement)

Fitantan-draharaham-panjakana (Administration publique)

Autre

Q16.Pere_secteur_autre

[pas dans questionnaire A]

Q16[other]. Sehatr'asa inona? Dans quel secteur d'activité? [Autre]

Champ libre.

« Centre d'Aide au personnes vulnérables » => « Santé »

« Chauffeur » => « Transport et télécommunications »

« Electricien / Réparation électrique » => « Construction »

« Electromécanicien / électromécanique » => « Activités de fabrication »

« Employé de la statistique » => « Administration publique »

« (Ex-)Militaire / Gendarme / Fiarovam-pirenena (sécu nationale) » => « Administration publique »

« Fambolena (Agriculture) » => « Agriculture et pêche »

« informatique » => « Activités informatiques »

« Locataire » => « Immobilier, location et services rendus aux entreprises » (j'imagine qu'il fait louer)

« Mécanique automobile » => « Commerce »

« Meublier » => « Activités de fabrication »

« Ministère de l'intérieur » => « Administration publique »

« Modérateur Francophone » => « Activités informatiques »

« Mpandrafitra (Charpentier) » => « Construction »

« Ouvrage métallique » => « Activités de fabrication »

« Privée » => « Transport et télécommunications » (Q15 : Chauffeur)

« Taxi » => « Transport et télécommunications »

« Télécom » => « Transport et télécommunications »

« Transport privé » => « Transport et télécommunications »

Q17.Pere_niv_etude

Q17. Inona no fari-pahaizan'ny dadanao? Quel est le niveau d'étude de votre père?

Hatreo amin'ny CEPE (Primaire)

Hatreo amin'ny BEPC (C.E.G)

Hatreo amin'ny Bacc (Lycée)

Sokajy teknika na sekoly manofana arak'asa (Niveau secondaire technique ou « professionnel »)

Ambaratonga ambony (Oniversite) (Supérieur)

Q18.Mere_niv_emp

Q18. Inona no sokajin'asan'ny mamanao? Quel est le niveau d'emploi de votre mère?

Tompon'andraikitra ambony (talé), injeniera na mitovy lenta amin'izany (Cadre, ingénieur ou équivalent)

Tompon'andraikitra (sefo) (Responsable, agent de maitrise)

Mpiasa voahofana manokana (Ouvrier spécialisé)

Mpiasa tsy nahazo fanofanana (Ouvrier sans formation)

Mpiasa tsotra (ohatra manoeuvre) (Agent (par exemple : manoeuvre))

Mpandraharaha (Entrepreneur)

Manao asa-tena (Profession libérale)

Mpiana-draharaha (Stagiaire)

Miasa antranon'olona (Employé de maison)

Mikarakara tokantrano (Père / Mère au foyer)

Autre, précisez

Q18.Mere_niv_emp_autre

Q18[other]. Inona no sokajin'asan'ny mamanao? Quel est le niveau d'emploi de votre mère? [Autre]

Champ libre.

Q19.Mere_secteur

Q19. Sehatr'asa inona? Dans quel secteur d'activité?

Fambolena sy jono (Agriculture et pêche)

Fitrandrahana (ohatra : harena ankibon'ny tany) (Activités d'extraction (par exemple : mines))

Indostria (ohatra : lamba) (Activités de fabrication (par exemple : textiles))

Fanaovana asa vaventy (Construction)

Varotra (Commerce)

Hotely, fisakafoanana, bar (Hôtels, restaurants, bars)

Fitanterana sy fifandraisan-davitra (Transports et télécommunications)

Informatika sy teknolojia (Activités informatiques (par exemple : développeur))

Fandraharahana ara-bola (ohatra : miasa amin'ny banky) (Activités financières (par exemple : agent d'une banque))

Fandraharana momban'ny trano na tolotr'asa hoan'ireo orinasa (Immobilier, location et services rendus aux entreprises)

Fahasalamana (Santé)

Fampianarana (Enseignement)

Fitantan-draharaham-panjakana (Administration publique)

Services (ajouté lors du traitement)

Autre, précisez

Q19.Mere_secteur_autre

[pas dans]

Q19[other]. Sehatr'asa inona? Dans quel secteur d'activité? [Autre]

Champ libre.

« Artisanat/ Stylisme » => « Commerce » (=> ventes d'objets artisanaux)

« Coiffeuse / Mpanao volo » = « Services »

« couturière » => « Activités de fabrication »

« Education » => « Enseignement »

« Femme de ménage / Asa an-trano / ASANTRANO » => Services

« Professeur » => « Enseignement »

« salon de beauté » => « Services »

« secrétaire » => « Services » (=> services administratifs)

Q20.Mere_niv_etude

Q20. Inona no fari-pahaizan'ny mamanao? Quel est le niveau d'étude de votre mère?

Hatreo amin'ny CEPE (Primaire)

Hatreo amin'ny BEPC (C.E.G)

Hatreo amin'ny Bacc (Lycée)

Sokajy teknika na sekoly manofana arak'asa (Niveau secondaire technique ou « professionnel »)

Ambaratonga ambony (Oniversite) (Supérieur)

Partie B :

Q21.Foyer_nb

Q21. Firy ianareo no ao anatin'ny tokantrano? Combien êtes-vous dans votre foyer?

[1-?]

Q21_1.Foyer_coloc

Q21A[SQ001]. Iza avy no miara-mipetraka ainao? Qui habite avec vous? [Ny na ireo mpiara-manofa trano amiko (Mon ou mes colocataire(s))]

Oui

Non

Q21_2.Foyer_conj

Q21A[SQ002]. Iza avy no miara-mipetraka ainao? Qui habite avec vous? [Ilay olona miaraka amiko (Mon ou ma conjoint(e) ou partenaire)]

Oui

Non

Q21_3.Foyer_parents

Q21A[SQ003]. Iza avy no miara-mipetraka aminaon'ny? Qui habite avec vous? [Ny ray aman-dreniko (Mes parents)]

Oui

Non

Q21_4.Foyer_conj_parents

Q21A[SQ004]. Iza avy no miara-mipetraka aminaon'ny? Qui habite avec vous? [Ny ray amandrenin'ny ny olona miaraka amiko (Les parents de mon conjoint ou partenaire)]

Oui

Non

Q21A_5.Foyer_enfants

Q21A[SQ005]. Iza avy no miara-mipetraka aminaon'ny? Qui habite avec vous? [Ny zanako (Mes enfants)]

Oui

Non

Q21_6.Foyer_autres_famille

Q21A[SQ006]. Iza avy no miara-mipetraka aminaon'ny? Qui habite avec vous? [Ny olona sansany ao anaty fianakaviako (D'autres membres de ma famille)]

Oui

Non

Q21_autre.Foyer_autres

Q21A[other]. Iza avy no miara-mipetraka aminaon'ny? Qui habite avec vous? [Autre]

Champ libre.

Q22.Foyer_revenus

Q22. Ohatrinona ny totalim-bola miditra ao amin'ny tokantranonao? (Raha kajiana daholo ny an'ny olona rehetra miara-mipetraka amina) Quel est le montant total dont votre foyer dispose par mois? (En prenant en compte les revenus de toutes les personnes qui habitent avec vous)

Q23.Assurance

Q23. Manana fiantohana ara-pahasalamana ve ianao? Avez vous une assurance santé?

[équivalent : Q21..Votre.employeur.vous.paye.t.il.une.assurance.santé?]

Eny, izay ny mpampiasa ahy no mandoa azy (Oui, payée par mon employeur principal)

Eny, izay izaho no mandoa azy (Oui, que je paye)

Tsia (Non)

Eny, izay aloan'olon-kafa : farito (Oui, payée par : autre précisez)

Q23_autre.Assurance_autre

[pas dans questionnaire A]

Q23[other]. Manana fiantohana ara-pahasalamana ve ianao? Avez vous une assurance santé? [Autre]

Champ libre.

Partie C :

Q24.Type_etab

Q24. Karazan-tsekoly inona no nianaranao farany? Quel type d'établissement avez-vous fréquenté pour la dernière fois?

Sekolim-panjankana (Établissement public)

Sekolim-pinoana tsy miankina (Établissement religieux privées)

Sekoly tsootra tsy miankina (Établissement non-religieux privées)

Q25.Niv_educ

Q25. Karazam-pianarana inona no narahinao farany? Quel niveau d'éducation avez-vous suivi pour la dernière fois?

Hatreo amin'ny CEPE (Primaire)

Hatreo amin'ny BEPC (C.E.G)

Hatreo amin'ny Bacc (Lycée)

Sokajy teknika na sekoly manofana arak'asa (Niveau secondaire technique ou « professionnel »)

Ambaratonga ambony (Oniversite) (Supérieur)

Q26.Etudiant

Q26. Mbola mpianatra eny amin'ny oniversite ve ianao? Êtes-vous encore étudiant à l'université?

[équivalent questionnaire A : Q24..Êtes.vous.encore.en.train.d.étudier?]

Eny (Oui)

Tsia (Non)

Q27.Discip_dipl

Q27. Inona no sokajin'ny fianarana nahazahonao diplaoma farany? Dans quelle discipline avez-vous obtenu votre dernier diplôme?

Lalana, toekarena, fitantanana, siansa maha olona (DEGS) (Droit, économie, gestion, sciences humaines)

Haisoratra, siansa momban'ny fiteny, fiteny vahiny (Arts, lettres, sciences du langage, langues)

Matematika sy fizika (Mathématiques, Physique)

Informatika sy teknolojia (Informatique)

Siansa momban'ny natiora, fambolena (Sciences de la nature et de la vie, agronomie)

Fitsaboana, Famasia, Fahasalamana (Médecine-odontologie, pharmacie, santé)

Hafa (farito) (Autre (précisez))

Q27_autre.Discip_dipl_autre

[pas dans questionnaire A]

Q27[other]. Inona no sokajin'ny fianarana nahazahonao diplaoma farany? Dans quelle discipline avez-vous obtenu votre dernier diplôme? [Autre]

Champ libre.

Q28.Formation_pro

Q28. Ivelan'ny sekoly na ny oniversite, ianao ve nahazo fiofanana arak'asa iray na maromaro? En dehors de l'école ou de l'université, avez-vous reçu une ou plusieurs formations professionnelles?

Eny (Oui)

Tsia (Non)

Q28_1.Formation_pro_outilsinfo

Q28B[SQ001]. Mahakasika ny lohahevitra inona? Sur quelles thématiques? [Fitaovana informatika (ohatra : Excel, Word) (Outils informatique (par exemple : Excel, Word))]

Oui

Non

Q28_2.Formation_pro_proginfo

Q28B[SQ002]. Mahakasika ny lohahevitra inona? Sur quelles thématiques? [Fandaharana informatika (Programmation informatique)]

Oui

Non

Q28_3.Formation_pro_lang

Q28B[SQ003]. Mahakasika ny lohahevitra inona? Sur quelles thématiques? [Fiteny vahiny (Langues étrangères)]

Oui

Non

Q28_4.Formation_pro_manag

Q28B[SQ004]. Mahakasika ny lohahevitra inona? Sur quelles thématiques? [Fitantanana (Management)]

Oui

Non

Q28_5.Formation_pro_gestionproj

Q28B[SQ005]. Mahakasika ny lohahevitra inona? Sur quelles thématiques? [Fitantanana tetik'asa (Gestion de projet)]

Oui

Non

Q28_autre.Formation_pro_autre

Q28B[other]. Mahakasika ny lohahevitra inona? Sur quelles thématiques? [Autre]

Champ libre.

Q29.Niv_lang_fr

Q29. Iza amin'ireto fehezan-teny ireto no mahalaza kokoa ny fari-pahaizanao ny teny frantsay? Laquelle de ces phrases décrit le mieux votre niveau de français?

Afaka ampiasaiko ilay fiteny amin'ny trangam-piainana andavan'andro, an-tsoratra sy am-bava (Je peux utiliser cette langue dans des situations de la vie courante, je peux me présenter en français à l'écrit ou à l'oral) => « **Débutant** »

Afaka mamariparitra tranga sy miresaka amin'ny fomba somary mazava, afaka mamaky sy manoratra lahatsoratra fohy aho. (Je peux décrire des événements et communiquer assez couramment, je peux lire et écrire des textes courts.) => « **Intermédiaire** »

Voafahiko tanteraka ilay fiteny, afaka mamaky sy manoratra lahatsoratra lava sy sarotra aho. (Je maîtrise presque parfaitement la langue, je peux lire et écrire des textes longs et compliqués.) => « **Avancé** »

Q30.Formation_lang_fr

Q30. Nahazo fiofanana ve ianao mba hifehezana ny teny frantsay? Avez-vous eu des formations pour maîtriser la langue française?

Eny (Oui)

Tsia (Non)

Partie D :

Q31_1.Appareil_ordi_port

Q31[Q01]. Inona ny fitaovana misy ao aminao? Quels sont les appareils qu'il y a chez vous? [Solosaina afaka entina an-tanana (Un ordinateur portable)]

Oui

Non

Q31_2.Appareil_ordi_bureau

Q31[Q02]. Inona ny fitaovana misy ao aminao? Quels sont les appareils qu'il y a chez vous? [Solosaina fampiasa ao am-birao (Un ordinateur de bureau)]

Oui

Non

Q31_3.Appareil_smartphone

Q31[Q03]. Inona ny fitaovana misy ao aminao? Quels sont les appareils qu'il y a chez vous? [Finday avo lenta (Un smartphone)]

Oui

Non

Q31_4.Appareil_tel_sans_internet

Q31[Q04]. Inona ny fitaovana misy ao aminao? Quels sont les appareils qu'il y a chez vous? [Finday tsy ahazoana aterineto (Un téléphone portable sans accès à internet)]

Oui

Non

Q31_5.Appareil_tablette

Q31[Q05]. Inona ny fitaovana misy ao aminao? Quels sont les appareils qu'il y a chez vous? [Tablety (Une tablette)]

Oui

Non

Q31_6.Appareil_console

Q31[Q06]. Inona ny fitaovana misy ao aminao? Quels sont les appareils qu'il y a chez vous? [Filalaovana lalao sary (Une console de jeux)]

Oui

Non

Q31_7.Appareil_aucun

Q31[Q07]. Inona ny fitaovana misy ao aminao? Quels sont les appareils qu'il y a chez vous? [Tsy misy (Aucun)]

Oui

Non

Q31_autre.Appareil_autre

Q31[other]. Inona ny fitaovana misy ao aminao? Quels sont les appareils qu'il y a chez vous? [Autre]

Champ libre.

Q32.Acces_internet_dehors

Q32. Ivelan'ny birao, amin'ny alalan'ny inona no hahafahanao misitraka amin'ny aterineto? En dehors du bureau, par quel moyen principal accédez-vous à Internet?

Amin'ny alalan'ny tolotra famandrihana avy amin'ny mpamatsy aterineto ("box") (Avec un abonnement à un fournisseur d'accès à Internet (box))

Amin'ny alalan'ny aterineton'ny findaiko (Avec la connexion de mon téléphone portable)

Amin'ny alalan'ny tambazaotra aterineto tsy andoavam-bola, ao amiko (Wifi misokatra) (Avec un réseau gratuit, chez moi (Wifi ouvert))

Amin'ny alalan'ny tambazotra tsy andoavam-bola, tsy ao amiko (Wifi misokatra, ao amina cafe, anya m-pianarana, any amin'ny oniversite, any amin'ny trano fitehirizam-boky) (Avec un réseau gratuit en dehors de chez moi (Wifi ouvert, dans un café, à l'école, à l'université, dans une bibliothèque))

Any amin'ny « cyber-café » (Dans un cyber-café)

Autre, précisez

Q32_autre.Acces_internet_dehors_autre

Q32[other]. Ivelan'ny birao, amin'ny alalan'ny inona no hahafahanao misitraka amin'ny aterineto? En dehors du bureau, par quel moyen principal accédez-vous à Internet? [Autre]

Champ libre.

Q33.Type_reseau_tel

Q33. Manana tolotra sy finday ianao izay ahafahanao misitraka, fara fahatsarany : Vous avez un abonnement et un téléphone qui vous permettent d'accéder, au maximum :

Amin'ny tambazotra 2g (edge) (Au réseau 2g (edge))

Amin'ny tambazotra 3g, 3g+ na H+ (Au réseau 3g, 3g+ ou H+)

Amin'ny tambazotra 4g na 4g+ (Au réseau 4g ou 4g+)

Tsy manana finday aho (Je n'ai pas de téléphone portable)

Q34.Qualite_reseau_tel

Q34. Amin'ny ankapobeany, manao ahoana ny hatsaran'ny connexion ampiasainao : Habituellement, la qualité de votre connexion est :

Tena tsara (tsy tapatapaka, mahery vaika) (Très bonne (pas de coupure, débit constant))

Tsara ihany (tsy dia tapaka, somary milona) (Plutôt bonne (peu de coupures, quelques ralentissements))

Somary ratsy (tapaka matetika, milona be) (Plutôt mauvaise (coupe souvent, beaucoup de ralentissements))

Tena ratsy (tapaka foana, tsy afaka ampiasana amin'ny andro sasany) (Très mauvaise (beaucoup de coupures, certains jours je ne peux pas l'utiliser))

Partie E :

[Pas de Q36 dans FB]

Q36_1.Deja_travail_site_oui

Q36[SQ001]. Efa nahita asa nadraisana karama hita tamin'ny internet ve ianao? Avez vous déjà effectué un travail rémunéré trouvé via un site ou une application? [Eny, hita tamin'ny tambazotra sosialy (vondrona Facebook, whatsapp, telegram, etc) (Oui, trouvé via un réseau social (un groupe facebook, whatsapp, telegram, etc.))]

Oui

Non

Q36_2.Deja_travail_site_non

Q36[SQ003]. Efa nahita asa nadraisana karama hita tamin'ny internet ve ianao? Avez vous déjà effectué un travail rémunéré trouvé via un site ou une application? [Tsia (Non)]

Oui

Non

Q36_autre.Deja_travail_site_oui_autre

Q36[other]. Efa nahita asa nadraisana karama hita tamin'ny internet ve ianao? Avez vous déjà effectué un travail rémunéré trouvé via un site ou une application? [Autre]

Champ libre.

Q35_1.Deja_travail_microworkers

Q35[Q01]. Efa niasa tamin'ireto tranonkala ireto ve ianao? Avez-vous déjà travaillé sur les sites suivants? [Microworkers]

Eny (Oui)

Tsia (Non)

Q35_2.Deja_travail_clickworker

Q35[Q02]. Efa niasa tamin'ireto tranonkala ireto ve ianao? Avez-vous déjà travaillé sur les sites suivants? [Clickworker]

Eny (Oui)

Tsia (Non)

Q35_3.Deja_travail_lionbridge

Q35[Q03]. Efa niasa tamin'ireto tranonkala ireto ve ianao? Avez-vous déjà travaillé sur les sites suivants?
[Lionbridge]

Eny (Oui)

Tsia (Non)

Q35_4.Deja_travail_appen

Q35[Q04]. Efa niasa tamin'ireto tranonkala ireto ve ianao? Avez-vous déjà travaillé sur les sites suivants?
[Appen]

Eny (Oui)

Tsia (Non)

Q35_5.Deja_travail_uhrs

Q35[Q05]. Efa niasa tamin'ireto tranonkala ireto ve ianao? Avez-vous déjà travaillé sur les sites suivants?
[UHRS]

Eny (Oui)

Tsia (Non)

Q35_6.Deja_travail_crowdflower

Q35[Q06]. Efa niasa tamin'ireto tranonkala ireto ve ianao? Avez-vous déjà travaillé sur les sites suivants?
[Crowdflower]

Eny (Oui)

Tsia (Non)

Q35_7.Deja_travail_upwork

Q35[Q07]. Efa niasa tamin'ireto tranonkala ireto ve ianao? Avez-vous déjà travaillé sur les sites suivants?
[Upwork]

Eny (Oui)

Tsia (Non)

Q35_8.Deja_travail_AIEthic

Q35[Q08]. Efa niasa tamin'ireto tranonkala ireto ve ianao? Avez-vous déjà travaillé sur les sites suivants?
[AIEthic]

Eny (Oui)

Tsia (Non)

[Pas de Q37_X dans FB]

Q37_1.Taches_annotation_data

Q37[SQ001]. Ny asa notanterahina dia : Le travail effectué relevait : [Fanasokajiana na fampidirana angona (fanasokajiana sary, fandikana faktiora, na rosia, sns. (De l'annotation ou de la saisie de données (labellisation d'images, retranscription de factures, ou de reçus, etc.)))]

Tsia (Non)

Eny (Oui)

Q37_2.Taches_prod_contenu

Q37[SQ002]. Ny asa notanterahina dia : Le travail effectué relevait : [Famokarana votoaty (fanoratana lahatsoratra, famokarana horonan-tsary, sns. (De la production de contenu (écriture d'articles, production de vidéos, etc.)))]

Tsia (Non)

Eny (Oui)

Q37_3.Taches_crea_freelance

Q37[SQ003]. Ny asa notanterahina dia : Le travail effectué relevait : [Asa famoronana mahaleotena, (famolavolana « logo », afisy) (Un travail créatif en freelance (design de logo, d'affiches))]

Tsia (Non)

Eny (Oui)

Q37_4.Taches_devinfo

Q37[SQ004]. Ny asa notanterahina dia : Le travail effectué relevait : [Asa famolavolana informatika (ohatra : famoronana tranonkala) (Un travail de développement informatique (par exemple : création de sites internet))]

Tsia (Non)

Eny (Oui)

Q37_5.Taches_autre

Q37[other]. Ny asa notanterahina dia : Le travail effectué relevait : [Autre]

Champ libre.

Q37_A.Source_travail_ligne [FB uniquement]

Q37FA.Amin.ny.ankapobeny..aiza.ianao.no.mahita.asa.amin.ny.aterineto.kokoa...En.général..où.trouvez.vous.le.plus.

Q37_A_autre.Source_travail_ligne_autre [FB uniquement]

Q37FA.other...Amin.ny.ankapobeny..aiza.ianao.no.mahita.asa.amin.ny.aterineto.kokoa...En.général..où.trouvez.vous.

Q37_B.Recruteur_co_internet [FB uniquement]

Q37FB..Matetika..rehefa.mahita.asa.amin.ny.internet.ianao...miantoka.ny.fandaniana.aterineto.ho.anao.ny.mpampiasa.

Q37_C.Recruteur_travail_bureau

[pas dans questionnaire A]

Q37C. Matetika, rehefa mahita asa amin'ny aterineto ianao : mangataka anao hiasa any amin'ny birao ny mpampiasa, miaraka amin'ny mpiasa hafa : Le plus souvent, quand vous trouvez du travail sur Internet : Le recruteur vous demande de venir travailler dans un bureau, avec d'autres travailleurs :

Eny (Oui)

Tsia (Non)

Q38_1.Clients_mg

Q38[SQ001]. Amin'ny ankapobeany, olontsotra ve sa orinasa ireo mpanjifa : En général, les clients sont des individus ou des entreprises : [Malagasy (Malgaches)]

Eny (Oui)

Tsia (Non)

Q38_2.Clients_fr

Q38[SQ002]. Amin'ny ankapobeany, olontsotra ve sa orinasa ireo mpanjifa : En général, les clients sont des individus ou des entreprises : [Frantsay (Françaises)]

Eny (Oui)

Tsia (Non)

Q38_3.Clients_us

Q38[SQ003]. Amin'ny ankapobeany, olontsotra ve sa orinasa ireo mpanjifa : En général, les clients sont des individus ou des entreprises : [Amerikanina (Américaines)]

Eny (Oui)

Tsia (Non)

Q38_autre.Clients_autre

Q38[other]. Amin'ny ankapobeany, olontsotra ve sa orinasa ireo mpanjifa : En général, les clients sont des individus ou des entreprises : [Autre]

Champ libre.

Q39_1.Taches_paiement

Q39. Amin'ny ankapobeany, inona no endriky ny karama raisinao? En général, comment êtes-vous payés pour les tâches réalisées?

Fanomezana na sorabola amina mpivarotra amin'ny internet (Carte cadeaux ou crédit sur des sites de vente)

Lelavola alefa ao amin'ny banky (Transfert d'argent vers mon compte bancaire)

Lelavola alefa amin'ny tambazotra (Orange money, Airtel money, MVola, Western union...) (Transfert d'argent sur Internet (Orange money, ...))

Vola ara-teknolojia (Bitcoin,) (Autres monnaies virtuelles (Bitcoin ...))

Hafa (farito) (Autre (précisez))

Q39_2.Taches_paiement_autre

Q39[other]. Amin'ny ankapobeany, inona no endriky ny karama raisinao? En général, comment êtes-vous payés pour les tâches réalisées? [Autre]

Champ libre.

Q40.Nb_pers_entourage_site

Q40. Firy ny isan'ny olona manodidina anao miasa sy mikarama amin'ny internet na ny tambazotra sosialy toan'ny Facebook na Whatsapp? Dans votre entourage, combien de personnes réalisent des tâches rémunérées sur des sites ou des applications, y compris via des groupes facebook ou whatsapp?

[0 – 25]

Q41.Groupe_entraide [FB uniquement]

Q41F..Mpikambana.amina.vondrona.fifanakalozan.kevitra.sy.fifanampiana.ho.an.ny.mampiditra.angona...opérateur.

Eny, amin'ny alalan'ny hafatra eo no ho eo (facebook messenger, whatsapp (Oui, par messagerie instantanée (facebook messenger, whatsapp))

Non (Tsia)

Q41_autre.Groupe_entraide_autre [FB uniquement]

Q41F.other...Mpikambana.amina.vondrona.fifanakalozan.kevitra.sy.fifanampiana.ho.an.ny.mampiditra.angona...opérateur.

Partie F :

@ 20 expériences emplois possibles = Q42[SQ001 à SQ020]

@ 6 infos à chaque fois = Q42[SQ001_X1 à X6]

@ Liste infos :

(1) intitulé du poste

(2) type de contrat

(3) tâche (absent dans l'énoncé du pdf, cf. ci-dessous)

(4) nom entreprise

(5) *salaire*

(6) *date prise de poste*

Q42.1_job

Q42.1_contrat

Q42.1_tache

Q42.1_entreprise

Q42.1_salaire

Q42.1_date

Q42[SQ001_X1]. Inona avy ireo asa nataonao nihotran'ny 6 volana nialohan'ny nahatongavanao tao amin'ny Sous-traitant G? Atomboy amin'ny asanao voalohany dia tohizo araky ny fotoana niasanao ka hatramin'izao asanao ankehitriny izao ato am Sous-traitant G. Mety indrindra raha soratanao mazava tsara ny anaran'ilay asa (ohatra : téléopérateur, ...), ny karazana fifanarahana ara-kasa (ohatra : CDD), ny anaran'ilay orinasa niasana, ny karama noraisina (Ariary) sy ny daty nanombohana niasa tao amin'ilay orinasa. Raha nitana toerana maromaro ianao teo anivon'ny orinasa iray, mariho mitokana ireo toerana ireo Famapiana : Tsindrio « ajouter une ligne » raha hanampy an-dalana hanoratrana izay asa natao mihoatran'ny 6 volana

Quels sont les emplois de plus de 6 mois que vous avez occupés avant de travailler pour Gamed? Commencez par votre premier emploi, puis continuez de manière chronologique jusqu'au poste que vous occupez actuellement. Merci de précisez : l'intitulé du poste (par exemple : Téléopérateur...), le type de contrat (par exemple : CDD), le nom de l'entreprise, le salaire que vous touchiez (en ariary) ainsi que la date de la prise de poste. Si vous avez occupé plusieurs postes au sein d'une même entreprise, indiquez ces différents postes séparément. Aide : le bouton « ajouter une ligne » vous permet d'ajouter un poste, une fois que vous avez indiqué tous les postes de plus de 6 mois que vous avez eu, vous pouvez appuyez sur « Suivant » [Premier emploi][Anaran'ilay asa]

Q43.Autre_pro_actuelle

Q43. Manana asa hafa ve ianao ankoatry ny ao Sous-traitant G? Actuellement, avez-vous une autre activité professionnelle que votre activité principale?

Eny, asa iray (Oui, une activité)

Eny, asa maromaro (Oui, plusieurs activités)

Tsia (Non)

Q43_FB.Freq_plusieurs_jobs [FB uniquement]

Est-ce que ça vous arrive de cumuler plusieurs jobs en même temps ?

Eny, matetika (Oui, souvent)

Eny, indraindray (Oui, parfois)

Tsia (Non)

Q44_FB.Nb_employeurs_3mois

Tao anatin'ny telo volana lasa, mpampiasa firy samy hafa no niasanao, na indray andro fotsiny aza ?

Au cours des trois derniers mois, pour combien d'employeurs différents avez-vous travaillé, ne serait-ce qu'une journée ?

[numérique]

Partie G :

Q45.Titre_emploi_actuel

[pas dans questionnaire A]

Q45. Chez Sous-traitant G, vous êtes actuellement :

Superviseur, responsable

Agent

Hafa (farito)(Autre (précisez))

Q45_autre.Titre_emploi_actuel_autre

[uniquement dans Sous-traitant G]

Q45[other]. Chez Sous-traitant G, vous êtes actuellement : [Autre]

Champ libre.

Q46.Emploi_actuel_duree

Q46. Hafiriana no nanaovanao io asa io, raha volana no isaina ? Depuis combien de mois occupez-vous cet emploi ?

[1 – 29]

Q47.Emploi_actuel_contrat

Q47. Inona no karazana fifanarahana arak'asanao ankehitriny ? Quel est votre contrat ?

CDD

CDI

Asa tsy misy fifanarahana (Travail sans contrat)

Autre

Q47_autre.Emploi_actuel_contrat_autre

Q47[other]. Inona no karazana fifanarahana arak'asanao ankehitriny ? Quel est votre contrat ? [Autre]

Champ libre.

[Q48 absente]

Q49.Heures_travail_semaine

Q49. Adin'ny firy isan-kerinandro ianao no miasa amin'ny ankapobeny ? En général, combien d'heures travaillez-vous par semaine ?

[8 – 40] (Sous-traitant G)

Q50.Rythme_travail_eleve

Q50. Mandeha mafy be ve ianao rehefa miasa : Travaillez-vous à un rythme élevé :

Tsy mbola mihitsy na saika (Jamais ou presque)

Indray mandeha, na indroa isam-bolana (Une ou deux fois par mois)

Indray mandeha, na indroa isan-kerinandro (Une ou deux fois par semaine)

Imbetsaka isan-kerinandro (Plusieurs fois par semaine)

Q51.Nb_semaines_conges

Q51. Firy herinandro ao anatin'ny taona ny fialan-tsasatra anananao? Combien de semaines de congés avez-vous par an?

[0 – 31]

Q52.Salaire

Q52. Ohatrinona ny karama noraisinao tao (amin'ny Ariary)? Quel est le montant de votre salaire (en ariary)?

[FB : En moyenne, quels étaient vos revenus par mois? (en ariary)]

[range à définir après traitement des données]

Q53_1.Phrases_descript_resp_hierarchique

Q53[Q53A]. Mamaritra ny asanao ve ireto fehezanteny manaraka ireto? Les phrases suivantes décrivent-elles votre travail? [Manana olona eo ambany fanaraha-masoko aho (J'ai des personnes sous ma responsabilité hiérarchique)]

Eny (Oui)

Tsia (Non)

Q53_2.Phrases_descript_choix_rythme

Q53[Q53B]. Mamaritra ny asanao ve ireto fehezanteny manaraka ireto? Les phrases suivantes décrivent-elles votre travail? [Afaka mandamina ny ora fiasako aho (Je peux choisir mon rythme de travail)]

Eny (Oui)

Tsia (Non)

Q53_3.Phrases_descript_erreur_qualite

Q53[Q53C]. Mamaritra ny asanao ve ireto fehezanteny manaraka ireto? Les phrases suivantes décrivent-elles votre travail? [Rehefa manao fahadisoana aho dia mety hisy fiantraikany goavana amin'ny kalitaon'ny vokatra na ny tolotra (Lorsque je fais une erreur, elle peut avoir une conséquence grave sur la qualité du produit ou du service)]

Eny (Oui)

Tsia (Non)

Q53_4.Phrases_descript_erreur_cout

Q53[Q53D]. Mamaritra ny asanao ve ireto fehezanteny manaraka ireto? Les phrases suivantes décrivent-elles votre travail? [Rehefa manao fahadisoana, dia mety hisy fiantraikany goavana amin'ny fandaniana ara-bola (Lorsque je fais une erreur, elle peut avoir une conséquence grave sur les coûts financiers)]

Eny (Oui)

Tsia (Non)

Q53_5.Phrases_descript_stress

Q53[Q53E]. Mamaritra ny asanao ve ireto fehezanteny manaraka ireto? Les phrases suivantes décrivent-elles votre travail? [Mety hitondra stress ho ahy ny asako (Mon travail m'expose à du stress)]

Eny (Oui)

Tsia (Non)

Q53_6.Phrases_descript_penible

Q53[Q53F]. Mamaritra ny asanao ve ireto fehezanteny manaraka ireto? Les phrases suivantes décrivent-elles votre travail? [Manery ahy hampiasa vatana be na koa manery ahy hanao fomba fietsika mitondra harerahana lava reny ny asako (Mon travail m'expose à des postures ou à des mouvements pénibles à la longue)]

Eny (Oui)

Tsia (Non)

Q53_7.Phrases_descript_lecture

Q53[Q53G]. Mamaritra ny asanao ve ireto fehezanteny manaraka ireto? Les phrases suivantes décrivent-elles votre travail? [Mitaky ahy hamaky soratra madinika na ratsy printy ny asako (Mon travail m'expose à lire des lettres ou des chiffres de petites tailles, mal imprimées)]

Eny (Oui)

Tsia (Non)

Q53_8.Phrases_descript_nuit

Q53[Q53H]. Mamaritra ny asanao ve ireto fehezanteny manaraka ireto? Les phrases suivantes décrivent-elles votre travail? [Matetika aho no miasa amin'ny alina (Il m'arrive régulièrement de travailler la nuit)]

Eny (Oui)

Tsia (Non)

Q53_9.Phrases_descript_weekend

Q53[Q53I]. Mamaritra ny asanao ve ireto fehezanteny manaraka ireto? Les phrases suivantes décrivent-elles votre travail? [Matetika aho no miasa amin'ny faran'ny herinandro (Il m'arrive régulièrement de travailler le week-end)]

Eny (Oui)

Tsia (Non)

Q53_10.Phrases_descript_equipe

Q53[Q53J]. Mamaritra ny asanao ve ireto fehezanteny manaraka ireto? Les phrases suivantes décrivent-elles votre travail? [Miasa amin'ny ekipa aho (Je travaille en équipe)]

Eny (Oui)

Tsia (Non)

Q53_11.Phrases_descript_autonome

Q53[Q53K]. Mamaritra ny asanao ve ireto fehezanteny manaraka ireto? Les phrases suivantes décrivent-elles votre travail? [Manana fahaleovantena aho amin'ny asako (Je suis autonome dans mon travail)]

Eny (Oui)

Tsia (Non)

Q54_1.Teletravail

Q54A. Miasa tany an-trano ve ianao? (asa lavitra...)? Travaillez-vous depuis chez vous (télétravail...)?

Eny (Oui)

Tsia (Non)

Q54_2.Teletravail_freq

Q54B. Impiry ianao no miasa any an-trano? A quelle fréquence travaillez-vous depuis chez vous?

Iray andro ao anatin'ny herinandro (Un jour par semaine)

Roa andro isan-kerinandro na mihoatra (Deux jours par semaine ou plus)

Q54_2_autre.Teletravail_freq_autre

[uniquement]

Q54B.other....Impiry.iana.no.miasa.any.an.trano...A.quelle.fréquence.travaillez.vous.depuis.chez.vous.....Autre.

Champ libre.

Q55_1.Demande_precision_client

Q55A. Mba mangataka fanazavana mivantana amin'ny mpanome asa anao ve ianao amin'ny mailaka, « slack », na hafa? Vous arrive t-il de demander des précisions directement au client par mail, slack, ou autre?

Eny (Oui)

Tsia (Non)

Q55_2.Demande_precision_client_freq

Q55B. Mangataka fanazavana mivantana amin'ny mpanjifa ny asanao ianao : Vous demandez des précisions directement au client :

Latsaky ny indray mandeha isam-bolana (Moins d'une fois par mois)

Indray mandeha na indroa isam-bolana (Une ou deux fois par mois)

Indray mandeha isan-kerinandro (Une fois par semaine)

Mihoatry ny indray mandeha isan-kerinandro (Plus d'une fois par semaine)

Q56_1.Relations_difficiles_chefs

Q56[Q51A]. Efa nisedra fifandraisana sarotra tamin'ny hafa ve ianao : Vous arrive t-il d'avoir des relations difficiles avec certains de vos : [Sefo (Chefs)]

Eny (Oui)

Tsia (Non)

Q56_2.Relations_difficiles_sub

Q56[Q51B]. Efa nisedra fifandraisana sarotra tamin'ny hafa ve ianao : Vous arrive t-il d'avoir des relations difficiles avec certains de vos : [Ny eo ambany (Vos subordonnés)]

Eny (Oui)

Tsia (Non)

Q56_3.Relations_difficiles_collegues

Q56[Q51C]. Efa nisedra fifandraisana sarotra tamin'ny hafa ve ianao : Vous arrive t-il d'avoir des relations difficiles avec certains de vos : [Ny mpiara-miasa aminao hafa (Vos autres collègues)]

Eny (Oui)

Tsia (Non)

Q56_4.Relations_difficiles_pers_exterieures

Q56[Q51D]. Efa nisedra fifandraisana sarotra tamin'ny hafa ve ianao : Vous arrive t-il d'avoir des relations difficiles avec certains de vos : [Ny vahoaka, ny mpanjifa, ary amin'ny ankapobeny dia ny olona ivelan'ny orinasa (Le public, les clients, et en général les personnes extérieures à l'entreprise)]

Eny (Oui)

Tsia (Non)

Q57.Discussions_travail_proches [Q57 et Q57_X pas dans FB]

Q57A. Tamin'ny herinandro lasa, niresaka tamin'ny olona akaiky anao ve ianao momba ny asanao? La semaine dernière, avez vous parlé avec vos proches de votre travail?

Eny (Oui)

Tsia (Non)

Q57_1.Discussions_travail_proches_fatigue

Q57B[Q57B1]. Raha eny, teo amin'ny lafiny inona? De quels aspects? [Ny harerahana vokatry ny asa (De

la fatigue du travail)]

Eny (Oui)

Tsia (Non)

Q57_2.Discussions_travail_proches_avenir

Q57B[Q57B2]. Raha eny, teo amin'ny lafiny inona? De quels aspects? [Ny hoavin'ny asanao (De l'avenir de votre emploi)]

Eny (Oui)

Tsia (Non)

Q57_3.Discussions_travail_proches_salaire

Q57B[Q57B3]. Raha eny, teo amin'ny lafiny inona? De quels aspects? [Ny karama (Du salaire)]

Eny (Oui)

Tsia (Non)

Q57_4.Discussions_travail_proches_collegues

Q57B[Q57B4]. Raha eny, teo amin'ny lafiny inona? De quels aspects? [Ny fifandraisana amin'ny mpiaramiasa (Des rapports avec vos collègues)]

Eny (Oui)

Tsia (Non)

Q57_5.Discussions_travail_proches_contenu

Q57B[Q57B5]. Raha eny, teo amin'ny lafiny inona? De quels aspects? [Ny votoatin'ny asanao (fahalianana, fahaserotana, vokatra) (Du contenu de votre travail (intérêt, difficultés, résultats))]

Eny (Oui)

Tsia (Non)

Q57_autre.Discussions_travail_proches_autres

Q57B[other]. Raha eny, teo amin'ny lafiny inona? De quels aspects? [Autre]

Champ libre.

Q58.Premier_jour_sem

Q58. Ny andro voalohan'ny herinandro : Le premier jour de la semaine :

Tsy manana moraly ianao (Vous n'avez pas le moral)

Somary faly ianao (Vous êtes plutôt content)

Andro toy ny andro rehetra ity (C'est un jour comme les autres)

Ity no andro sarotra indrindra (C'est le jour le plus dur)

Q59.Incompetence_freq

Q59. Mahatsapa ho tsy ampy fahaiza-manao hanao ny asanao ve ianao ? Avez-vous l'impression de ne pas avoir les compétences pour accomplir votre travail ?

Tsy mbola mihitsy na saika (Jamais ou presque)

Indray mandeha na indroa isam-bolana (Une ou deux fois par mois)

Indray mandeha na indroa isan-kerinandro (Une ou deux fois par semaine)

Imbetsaka isan-kerinandro (Plusieurs fois par semaine)

Q60.Probabilite_perte_emploi [pas dans FB]

Q60. Amin'ny taona ho avy, ny mety hahaverezan'ny asanao ve : Au cours de l'année qui vient, le risque de perdre votre emploi est :

Saika tsy misy (A peu près inexistant)

Iva (Faible)

Avo (Élevé)

Avo dia avo (Très élevé)

Q61.Probabilite_promotion

Q61. Amin'ny taona ho avy, ny mety hahazoana fison-drotana amin'ny asanao ve : Au cours de l'année qui vient, la possibilité d'avoir une promotion au sein de Sous-traitant G est :

Saika tsy misy (A peu près inexistant)

Iva (Faible)

Avo (Élevé)

Avo dia avo (Très élevé)

Q62.Retrouver_emploi_equiv [pas dans FB]

Q62. Raha very asa ianao, mitady asa farafaharatsiny mitovy ve aminao : Si vous perdiez votre emploi chez Sous-traitant G, retrouver un emploi au moins équivalent serait :

Tsotra kely (Très facile) Somary tsotra (Plutôt facile) Somary sarotra (Plutôt difficile) Tsy vita atao (Impossible)

[Q63 absent]

Q64.Conditions_travail [pas dans FB]

Q64. Nanomboka niasanao tao Sous-traitant G ny toetr'asanao dia : Depuis que vous êtes chez Sous-traitant G, vos conditions de travail se sont :

Niharatsy (Dégradées) Tsy niova (Restés identique) Nihatsara (Améliorées) Mbola vao tsy ela ianao no teo ka tsy mbola afaka ny hitsara (Vous n'êtes pas là depuis assez longtemps pour juger)

Partie I :

Q65.Heureux_enfant

Q65. Mety ho faly ve ianao raha toa ka manao asa mitovy aminao ny iray amin'ireo zanakao? Seriez-vous ou auriez-vous été heureux que l'un de vos enfants s'engage dans la même activité que vous?

Eny (Oui) Tsia (Non)

Q66.Appreciation_exterieure

Q66. Araka ny hevitrao, ny asa ataonao ve, raha ivelan'ny sehatry ny asa dia : Selon vous, la profession que vous exercez est-elle, en dehors de votre milieu professionnel :

Somary hitan'ny olona ho tsara (Plutôt bien vue) Ankasitrahan'ny sasany, kianin'ny sasany (Appréciée par les uns, critiquée par les autres) Somary hitan'ny olona ho ratsy (Plutôt mal vue) Tsy fantatry ny olona (Pas connue)

Q67_1.Phrases_ressenti_fierte

Q67[Q61A]. Ireto tolo-kevitra ireto ve mifanaraka amin'ny zavatra hiainanao amin'ny asanao ankehitriny na ny asanao fototra? Les propositions suivantes correspondent-elles à la réalité de votre emploi actuel ou de votre activité principale? [Faly amin'ny asa tsara vita (La fierté du travail bien fait)]

Eny (Oui) Tsia (Non)

Q67_2.Phrases_ressenti_bienpaye

Q67[Q61B]. Ireto tolo-kevitra ireto ve mifanaraka amin'ny zavatra hiainanao amin'ny asanao ankehitriny na ny asanao fototra? Les propositions suivantes correspondent-elles à la réalité de votre emploi actuel ou de votre activité principale? [Matsiaro ho tsara karama (J'ai le sentiment d'être bien payé)]

Eny (Oui) Tsia (Non)

Q67_3.Phrases_ressenti_justevaleur

Q67[Q61C]. Ireto tolo-kevitra ireto ve mifanaraka amin'ny zavatra hiainanao amin'ny asanao ankehitriny na ny asanao fototra? Les propositions suivantes correspondent-elles à la réalité de votre emploi actuel ou de votre activité principale? [Matsiaro ho voahaja amin'ny maha izy ny tena (L'impression d'être reconnu à votre juste valeur)]

Eny (Oui)

Tsia (Non)

Q67_4.Phrases_ressenti_utile

Q67[Q61D]. Ireto tolo-kevitra ireto ve mifanaraka amin'ny zavatra hiainanao amin'ny asanao ankehitriny na ny asanao fototra? Les propositions suivantes correspondent-elles à la réalité de votre emploi actuel ou de votre activité principale? [Mahatsiaro ho manao asa ilain'ny hafa (L'impression de faire quelque chose d'utile aux autres)]

Eny (Oui)

Tsia (Non)

Q67_5.Phrases_ressenti_exploitation

Q67[Q61E]. Ireto tolo-kevitra ireto ve mifanaraka amin'ny zavatra hiainanao amin'ny asanao ankehitriny na ny asanao fototra? Les propositions suivantes correspondent-elles à la réalité de votre emploi actuel ou

de votre activité principale? [Mahatsiaro ho voambaka (Le sentiment d'être exploité)]

Eny (Oui)

Tsia (Non)

Q67_6.Phrases_ressenti_ennui

Q67[Q61F]. Ireto tolo-kevitra ireto ve mifanaraka amin'ny zavatra hiainanao amin'ny asanao ankehitriny na ny asanao fototra? Les propositions suivantes correspondent-elles à la réalité de votre emploi actuel ou de votre activité principale? [Mahatsiaro ho leo (L'ennui)]

Eny (Oui)

Tsia (Non)

Q67_7.Phrases_ressenti_interchangeable

Q67[Q61G]. Ireto tolo-kevitra ireto ve mifanaraka amin'ny zavatra hiainanao amin'ny asanao ankehitriny na ny asanao fototra? Les propositions suivantes correspondent-elles à la réalité de votre emploi actuel ou de votre activité principale? [Mahatsiaro hoe ny asa hatao dia asa vitan'ny hafa ihany (L'impression que ce que vous faites, n'importe qui pourrait le faire)]

Eny (Oui)

Tsia (Non)

Q68.Ressenti_general

Q68. Raha ny farany, ao anatin'ny asanao, inona no manandanja kokoa? Finalement, dans votre travail, qu'est-ce qui l'emporte?

Ny anton'ny fahafaham-po (Les motifs de satisfaction)

Ny anton'ny tsy fahafaham-po (Les motifs d'insatisfaction)

Mifandanja eo ihany ireo (Les uns et les autres s'équilibrent à peu près)

[A PARTIR D'ICI : DE COTE POUR L'INSTANT]

Partie J :

Q69_[1 : 23] ou [A : W]

Mahafantatra olona ve ianao miasa amin'ireto sehatr'asa ireto? Olona fantatrao anarana feno ary olona mifandray sy miresaka amina ary tena mifankazatra (miresaka fa tsy hoe mifampiarahaba fotsiny). Raha tsy miasa intsony ilay olona dia soraty ny asany farany fantatrao. Eritrereto tsara ilay olona isaky ny soka-jin'asa anaty ny lisitra dia tsipio amin'ny toerana mifanitsy aminy amin'ny tabilao. Raha tsy mahafantatra olona ianao dia tsipio eo amin'ny hoe « Tsia »

Connaissez-vous quelqu'un dans chacune de ces professions? Il s'agit de personnes dont vous connaissez le nom et prénom, et avec qui vous pourriez parler ou échanger de manière approfondie (en disant plus que bonjour ou bonsoir!) si vous les rencontriez par hasard dans la rue. Pour les personnes qui n'exercent plus d'activité professionnelle, inscrivez la dernière profession. Pensez à une personne pour chaque profession dans la liste, et cochez les cases appropriées. Si vous ne connaissez personne, cochez la case « Je ne connais personne dans cette profession »

Mpitsabo (Médecin)

Mpampiantra (Instituteur)

Tompon'orinasa (Chef d'entreprise)

Mpampianatra eny am anjery manontolo (Professeur à l'université)

Mpikirakira solosaina (Développeur logiciel ou web)

Data analyst, data scientist

Mpanolo-tsaina IT (Consultant informatique)

Mpiandraikitra hotely na toerampisakafoana (Manager dans un hôtel ou restaurant)

Mpiandraikitra amin'ny fambolena (Manager dans l'agriculture)

Mpiandraikitra amin'ny jono (Manager dans la pêche)

Mpiandraikitra amin'ny indostria (Manager dans l'industrie)

Mpiandraikitra mifandray amin'ny orinasa (Manager dans les services)

Misehatra amin'ny varotra (Commerçant)

Miasa amin'ny fandikana lahatsoratra amin'ny solosaina (Opérateur de saisie de données)

Fiarakaretsaka (Taxi)

Mpivarotra (Employé de commerce, par exemple : vendeur)

Mpamboly (Ouvrier agricole)

Mpiasa amin'indostria (Ouvrier dans l'industrie, par exemple : mine)

Mpanjono (Ouvrier de la pêche)

Mpanjaitra (Ouvrier du textile)

Miasa am heritsandry (Ouvrier de la manutention)

Polisy (Policier)

Miasa am hotely, trano fisakafoanana toy ny fanaovana asa fanadiovana (Employé d'hôtellerie, de restauration et d'entretien, dont serveurs, employés d'étage des hôtels concierge)

Q69_1.Niveau_relation [x 23]

Q69_2.Genre_individu [x 23]

Partie K :

Q70_[1 : 15] ou [A : O].Collegues

Iza avy ny mpiara miasa mifankahita ainao ivelan'ny toerampiasana ? (miara misotro, miaraka miala voly)

Qui sont les collègues que vous voyez en dehors du travail ? (boire un verre, aller au restaurant)

Questionnaire A 1 à 116

Q70[SQ001_SQ001]. Iza avy ny mpiara miasa mifankahita ainao ivelan'ny toeram-piasana ? (miara misotro, miaraka miala voly) Qui sont les collègues que vous voyez en dehors du travail ? (boire un verre, aller au restaurant) Si vous ne voyez aucun de vos collègues en dehors du travail, cliquez directement sur suivant. [AH-SAM Ranarivelo Francki Ronaldo][[]]

[] 1 à 56

Q64[SQ056_SQ001]. Iza avy ny mpiara miasa mifankahita ainao ivelan'ny toeram-piasana (miara misotro, miaraka miala voly). Si vous ne voyez aucun de vos collègues en dehors du travail, cliquez directement sur suivant. [RAMAROSANDRATANA Rado][[]]

Partie L :

Q70.Niveau_relation_collegues [pour les 15]

Q70[SQ001_SQ001]. Iza avy ny mpiara miasa mifankahita aminao ivelan'ny toeram-piasana? (miara misotro, miaraka miala voly) Qui sont les collègues que vous voyez en dehors du travail? (boire un verre, aller au restaurant) Ny bokotra « ajouter » dia hafahanao manampy anarana, rehefa fenonao ny anarana rehetra dia tsindrio ny bokotra « Suivant ». Raha tsy misy anarana ho soratana dia tsindrio avy hatranay ny « Suivant » Le bouton « ajouter » vous permet d'ajouter un nom , une fois que vous avez indiqué tous les noms, vous pouvez appuyez sur « Suivant ». Si vous n'avez pas de nom à ajouter, cliquez directement sur « Suivant ». [1][Noms]

Q70_B. [QUE DANS]

Q70B[SQ001_SQ001]. Si vous voulez ajouter des noms qui ne sont pas dans la liste, vous pouvez le faire ici : Fanampiana : ny bokotra « hanampy an-dalana » dia hafahanao manampy anarana, rehefa fenonao ny anarana rehetra dia tsindrio ny bokotra « Manaraka ». Raha tsy misy anarana ho soratana dia tsindrio avy hatranay ny « Manaraka » Le bouton « ajouter une ligne » vous permet d'ajouter un nom , une fois que vous avez indiqué tous les noms, vous pouvez appuyez sur « Suivant ». Si vous n'avez pas de nom à ajouter, cliquez directement sur “Suivant”. [1][Noms]

Q70_C.

Q70C[SQ001]. Ho anao, io olona io ve : Pour vous, cette personne est : [Q70_SQ001_SQ001]

Q70_D.

Q70D[SQ001_SQ001]. Iza amin'ireto no mifankahita ivelan'ny asa? Parmi ceux-ci, lesquels se voient aussi entre eux en dehors du travail? tsipio eo amin'ny fifanojoan'ny anarana anakiroa hafahanao manondro raha mpinama izy ireo Cochez les cases à l'intersection de deux noms pour indiquer qu'ils sont amis Hiverimberina matetika ireto fanintaniana ireto ary dia tsipio ny sela mifandraika aminy. La même information vous sera demandée plusieurs fois, cochez les cases à chaque fois. Torohevitra : araho andalana isakiny andalana dia hamarino ny tsanganana mifandraika aminy, dia tohizo aminy andalana manaraka. Aide : Progressez ligne par ligne, commencez par la première, cochez les colonnes qui s'appliquent, puis passez à la seconde ligne, etc. Raha tsy misy mpinamana ireo namanao dia tsindrio ny "Suivant" Si aucun de vos amis ne sont amis entre-eux, vous pouvez directement cliquer sur "Suivant". [Q70_SQ001_SQ001][Q70_SQ001_SQ001]{is_empty}

Partie M :

Q71.

Q71[SQ001_SQ001]. Raha ny telo volana farany, iza no efa nakanao hevitra na fanampiana mba nahafa-

hanao nanao « annotation » ? Sur les trois dernier mois, à qui avez-vous demandé un conseil ou de l'aide pour réaliser une annotation ? Mety hoe : naka fanazana tamin'ny tompon'andraikitra na toro hevitra fohy tamin'ny mpiara miasa, etc. Cela peut être : demander une clarification à un manager, un conseil bref à un collègue, etc. Ny bokotra « ajouter » dia hafahanao manampy anarana, rehefa fenonao ny anarana rehetra dia tsindrio ny bokotra « Suivant ». Raha tsy misy anarana ho soratana dia tsindrio avy hatranay ny « Suivant » Le bouton « ajouter » vous permet d'ajouter un nom , une fois que vous avez indiqué tous les noms, vous pouvez appuyez sur « Suivant ». Si vous n'avez pas de nom à ajouter, cliquez directement sur « Suivant ». [1][Noms]

Partie N :

Q71_B.

Parmi ceux-ci, lesquels ont travaillé ensemble sur un même projet au cours des trois derniers mois ?

Q71B[SQ001_SQ001]. Aminy ireto olona ireto, iza avy no efa niara niasa tao anatiny tetikasa hiraisana taminy 3 volana lasa ? Parmi ceux-ci, lesquels ont travaillé ensemble sur un même projet au cours des trois derniers mois ? Hamarino ny sela manelanelana ny anarana anakiroa ahafahana manamarina fa niara niasa taminy tetikasa hiombonana izy ireo. cochez les cases à l'intersection de deux noms pour indiquer qu'ils ont travaillé dans la même équipe, sur un même projet. Hiverimberina matetika ireto fanintaniana ireto ary dia tsipio ny sela mifandraika aminy. La même information vous sera demandée plusieurs fois, cochez les cases à chaque fois. Torohevitra : araho andalana isakiny andalana dia hamarino ny tsanganana mifandraika aminy, dia tohizo aminy andalana manaraka. Aide : Progressez ligne par ligne, commencez par la première, cochez les colonnes qui s'appliquent, puis passez à la seconde ligne, etc. Raha tsy misy niara niasa ihany koa dia tsindrio ny "Suivant" Si aucun d'entre eux n'ont travaillé ensemble, vous pouvez directement cliquer sur "Suivant". [Q71_SQ001_SQ001][Q71_SQ001_SQ001]{is_empty(Q71_SQ001_SQ001)}

Q71_C.

Q71C[SQ001]. Raha ny telo volana farany, nangataka hevitra na fanampiana tamin'io olona io ve ianao : Sur les trois derniers mois, Vous avez demandé de l'aide à cette personne : [Q71_SQ001_SQ001]

Q71_D.

Q71D[SQ001]. Io olona i ove : Cette personne est : [Q71_SQ001_SQ001]

Partie O :

Q72.

Sur les trois derniers mois, quelqu'un vous a-t-il demandé un conseil ou de l'aide pour réaliser son travail?

Q72[SQ001_SQ001]. Raha ny telo volana farany, nisy olona ve naka hevitra taminao na nangataka fanampiana mba hanantanterahany ny asany? Sur les trois derniers mois, quelqu'un vous a-t-il demandé un conseil ou de l'aide pour réaliser son travail? Mety mikasika fisalasalana amina « label » na mety izay torohevitra mifandray amin'ny « annotation ». Cela peut concerner : une incertitude sur un label, ou tout autre conseil lié au travail d'annotation. Ny bokotra « ajouter » dia hafahanao manampy anarana, rehefa fenonao ny anarana rehetra dia tsindrio ny bokotra « Suivant ». Raha tsy misy anarana ho soratana dia tsindrio avy hatranay ny « Suivant » Le bouton « ajouter » vous permet d'ajouter un nom , une fois que vous avez indiqué tous les noms, vous pouvez appuyez sur « Suivant ». Si vous n'avez pas de nom à ajouter, cliquez directement sur « Suivant ». [1][Noms]

Partie P :

Q72_B.

Parmi ceux-ci, lesquels ont travaillé ensemble sur un même projet au cours des trois derniers mois?

Q72B[SQ001_SQ001]. Aminy ireto olona ireto, iza avy no efa niara niasa tao anatin'ny tetikasa hiraisana taminy 3 volana lasa? Parmi ceux-ci, lesquels ont travaillé ensemble sur un même projet au cours des trois derniers mois? Hamarino ny sela manelanelana ny anarana anakiroa ahafahana manamarina fa niara niasa taminy tetikasa hiombonana izy ireo. cochez les cases à l'intersection de deux noms pour indiquer qu'ils ont travaillé dans la même équipe, sur un même projet. Hiverimberina matetika ireto fanintaniana ireto ary dia tsipio ny sela mifandraika aminy. La même information vous sera demandée plusieurs fois, cochez les cases à chaque fois. Torohevitra : araho andalana isakiny andalana dia hamarino ny tsanganana mifandraika aminy, dia tohizo aminy andalana manaraka. Aide : Progressez ligne par ligne, commencez par la première, cochez les colonnes qui s'appliquent, puis passez à la seconde ligne, etc. Raha tsy misy niara niasa ihany koa dia tsindrio ny "Suivant" Si aucun d'entre eux n'ont travaillé ensemble, vous pouvez directement cliquer sur "Suivant". [Q72_SQ001_SQ001][Q72_SQ001_SQ001]{is_empty(Q72_SQ001_SQ001)}

Q72_C.

Q72C[SQ001]. Raha ny telo volana farany, nangataka hevitra na fanampiana taminao ve io olona io : Sur les trois derniers mois, cette personne vous a-t-elle demandé de l'aide : [Q72_SQ001_SQ001]

Q72_D.

Q72D[SQ001]. Io olona i ove : Cette personne est : [Q72_SQ001_SQ001]

C Analyse des données quantitatives

C.1 Analyses descriptives

Variable	N = 293
Genre	
Un homme	196 (66,7%)
Une femme	98 (33,3%)
Age	
18-24 ans	79 (26,9%)
25-34 ans	176 (59,9%)
35-44 ans	33 (11,2%)
45-55 ans	6 (2%)
Statut matrimonial	
Marié(e)	133 (45,2%)
Célibataire	108 (36,7%)
En couple	38 (12,9%)
Ne souhaite pas répondre	10 (3,4%)
Divorcé(e)	2 (0,7%)
Veuf, veuve	2 (0,7%)
Autre	1 (0,3%)

¹ n (%)

TABLE 42 – Variables socio-démographiques

Variable	N = 170
CSP du conjoint	
Ouvrier spécialisé	30 (17,2%)
Profession libérale	28 (17%)
Agent, manoeuvre	22 (12,6%)
Père / Mère au foyer	18 (10,3%)
Responsable, agent de maîtrise	20 (11,5%)
Stagiaire	8 (4,6%)
Cadre, ingénieur ou équivalent	7 (4%)
Entrepreneur	7 (4%)
Autre	32 (18,4%)

¹ n (%)

TABLE 43 – Origine sociale du conjoint

Variable	N = 170
CSP du conjoint	
Secteur d'activité du conjoint	
Activités informatiques	46 (27,1%)
Commerce	33 (19,4%)
Enseignement	12 (7,1%)
Santé	10 (5,9%)
Transport et télécommunications	8 (4,7%)
Activités financières	7 (4,1%)
Administration publique	6 (3,5%)
Agriculture et pêche	4 (2,4%)
Hôtels, restaurants, bars	4 (2,4%)
Activités de fabrication	4 (2,5%)
Activités d'extraction	3 (1,8%)
Construction	2 (1,2%)
Immobilier, location et services rendus aux entreprises	2 (1,2%)
Autre	29 (17,1%)
N.A	4

¹ n (%)

TABLE 44 – Secteur d'activité du conjoint

Variable	N = 279	
Salaire	Ariary	Euros
Minimum	100 000	23,97
Maximum	17 000 000	4082,2
Moyenne	1 419 938	340,97
Médiane	700 000	168,09
Manquant	38	

TABLE 45 – Revenus du foyer par mois

Variable	N = 268
Type d'établissement	
Établissement non-religieux privé	119 (44%)
Établissement public	99 (37%)
Établissement religieux privé	50 (19%)

¹ n (%)

TABLE 46 – Type d'établissement fréquenté

Variable	N = 268
Nombre d'étudiant	
Non	217 (81%)
Oui	51 (19%)
Niveau d'éducation	
Supérieur	200 (75%)
Lycée général	37 (14%)
Lycée pro	25 (9.3%)
Collège	4 (1.5%)
Primaire	2 (0.7%)

¹ n (%)

TABLE 47 – Niveau d'études

Variable	N = 206
Outils informatiques	
Oui	122 (59%)
Non	84 (41%)
Programmation informatique	
Non	162 (79%)
Oui	44 (21%)
Langues étrangères	
Non	137 (67%)
Oui	69 (33%)
Management	
Non	174 (84%)
Oui	32 (16%)
Gestion de projet	
Non	173 (84%)
Oui	33 (16%)

¹ n (%)

TABLE 48 – Formation professionnelle

Variable	N = 268
Niveau langue française	
Intermédiaire	102 (38%)
Avancé	93 (35%)
Débutant	73 (27%)
Formation au français	
Non	189 (71%)
Oui	79 (29%)

¹ n (%)

TABLE 49 – Niveau de langue en français

Variable	N = 279
Niveau d'étude du père	
Lycée général	92 (33%)
Supérieur	80 (29%)
Collège	47 (17%)
Lycée pro	30 (11%)
Primaire	28 (10%)
N.A	2

¹ n (%)

TABLE 50 – Origine sociale du père

Variable	N = 279
Niveau d'étude de la mère	
Lycée général	97 (35%)
Collège	80 (29%)
Supérieur	57 (21%)
Lycée pro	23 (8.3%)
Primaire	21 (7.6%)
N.A	1

¹ n (%)

TABLE 51 – Origine sociale de la mère

C.2 Tri croisé pour l'analyse du capital social, économique et culturel des annotateurs

	Homme	Femme ¹	Total
Classe 1	28,9	71,1	100
Classe 2	23,9	76,1	100
Classe 3	42,9	57,1	100
Ensemble	29,8	70,2	100

¹Réponses en pourcentage, Fisher.test = 0.04

TABLE 52 – Genre par la classe

	18-24 ans	25-34 ans ¹	<35 ans	Total
Classe 1	18,9	75,6	5,6	100
Classe 2	43,1	55	1,8	100
Classe 3	5,4	66,1	28,6	100
Ensemble	26,3	64,7	9	100

¹Réponses en pourcentage, Fisher.test = 1,04e-11

TABLE 53 – Âge par la classe

	Célibataire	En couple ¹	Marié	Autre	Total
Classe 1	6,7	17,8	72,2	3,3	100
Classe 2	76,1	14,7	4,6	4,6	100
Classe 3	7,1	7,1	76,8	8,9	100
Ensemble	36,5	14,1	44,3	5,1	100

¹Réponses en pourcentage, Fisher.test = 2,2e-16

TABLE 54 – Statut matrimonial par la classe

	Non	Oui ¹	Total
Classe 1	90	10	100
Classe 2	32,1	67,9	100
Classe 3	69,6	30,4	100
Ensemble	60,8	39,2	100

¹Réponses en pourcentage, Fisher.test = 2,738e-16

TABLE 55 – Habiter avec ses parents par la classe

	Non	Oui ¹	Total
Classe 1	37,8	62,2	100
Classe 2	100	0	100
Classe 3	28,6	71,4	100
Ensemble	62,4	37,6	100

¹Réponses en pourcentage, Fisher.test = 2,2e-16

TABLE 56 – Habiter avec ses enfants par la classe

	Lycée ou avant	Supérieur	Total
Classe 1	42,2	57,8	100
Classe 2	15,6	84,4	100
Classe 3	8,9	91,1	100
Ensemble	23,5	76,5	100

¹Réponses en pourcentage, Fisher.test = 8,632e-7

TABLE 57 – Études par la classe

	Non	Oui	Total
Classe 1	88,9	11,1	100
Classe 2	64,2	35,8	100
Classe 3	94,6	5,4	100
Ensemble	79,6	20,4	100

¹Réponses en pourcentage, Fisher.test = 2,738e-16

TABLE 58 – Encore étudiant par la classe

	[26,71 - 49,54)	[49,54 - 52,46)	[52,46 - 56,16)	[56,16 - 78,57)	Total
Classe 1	58,9	17,8	8,9	14,4	100
Classe 2	31,2	26,6	20,2	22	100
Classe 3	7,1	19,6	57,1	16,1	100
Ensemble	35,7	22	24,3	18	100

¹Réponses en pourcentage, Fisher.test = 4,273e-12

TABLE 59 – Position sociale par la classe

	100-500	500-750	750-1325	1325-1000000	Total
Classe 1	34,4	36,7	22,2	6,7	100
Classe 2	15,6	31,2	30,3	22,9	100
Classe 3	10,7	8,9	12,5	67,9	100
Ensemble	21,2	28,2	23,5	27,1	100

¹Réponses en pourcentage, Fisher.test = 2,38e-14

TABLE 60 – Revenus du foyer par la classe

C.3 Analyse des conditions de travail

	Oui payé par mon employeur	Oui que je paye	Non	Autre	Total
Sous-traitant G	57,4	21,3	19,1	2,1	100
Sous-traitant B	85,3	0	13,8	0	100
Sous-traitant A	86,9	3,6	8,3	1,2	100
Ensemble	80,6	5,3	13	1,2	100

¹Réponses en pourcentage, Fisher.test = 3,027e-06

TABLE 61 – Accès à une assurance employeur par entreprise

Variable	N = 262
Choix du rythme	
Oui	153 (59,3%)
Non	105 (40,7%)
Rythme de travail intense	
Plusieurs fois par semaine	107 (41,5%)
Une ou deux fois par semaine	88 (34,1%)
Une ou deux fois par mois	32 (12,4%)
Jamais ou presque	31 (12%)
N.A	36

TABLE 62 – Rythme de travail

Personnes sous sa responsabilité	
Non	196 (76%)
Oui	62 (24%)
Mon travail a des conséquences sur la qualité	
Oui	240 (93%)
Non	18 (7%)
Mon travail a des conséquences sur les coûts	
Oui	171 (66,3%)
Non	87 (33,7%)
Mon travail m'expose à du stress	
Non	142 (55%)
Oui	116 (45%)
Mon travail m'expose à des postures pénibles	
Non	208 (80,6%)
Oui	50 (20,4%)
Mon travail m'expose à des petits caractères	
Non	195 (75,6%)
Oui	63 (24,4%)
Travail de nuit	
Non	189 (73,3%)
Oui	69 (26,7%)
Travail le week-end	
Non	160 (62%)
Oui	98 (38%)
Je travaille en équipe	
Oui	246 (95,3%)
Non	12 (4,7%)
Je suis autonome	
Oui	197 (76,4%)
Non	61 (23,6%)

¹ n (%)

TABLE 63 – Conditions de travail

	Ne peut pas choisir son rythme	Peut choisir son rythme	Total
Sous-traitant G	55,3	44,7	100
Sous-traitant B	21,5	78,5	100
Sous-traitant A	59	41	100
Ensemble	42,2	57,8	100

¹Réponses en pourcentage, Fisher.test = 3,787e-07

TABLE 64 – Choix du rythme selon l'entreprise

	Ne travaillent pas la nuit	Travaillent de nuit	Total
Sous-traitant G	97,9	2,1	100
Sous-traitant B	84,9	15,1	100
Sous-traitant A	50	50	100
Ensemble	75,2	24,8	100

¹Réponses en pourcentage, Fisher.test = 8,689e-11

	Ne travaillent pas le week-end	Travaillent le week-end	Total
Sous-traitant G	91,5	8,5	100
Sous-traitant B	77,4	22,6	100
Sous-traitant A	34,6	65,4	100
Ensemble	65,1	34,9	100

¹Réponses en pourcentage, Fisher.test = 2,062e-12

TABLE 65 – Travail la nuit et le week-end selon l'entreprise

	Non	Oui	Total
Sous-traitant G	61,7	38,3	100
Sous-traitant B	52,6	47,4	100
Sous-traitant A	73,1	26,9	100
Ensemble	61,1	38,9	100

¹Réponses en pourcentage, Fisher.test = 0,01593

TABLE 66 – Présence d'enfants dans le foyer selon l'entreprise

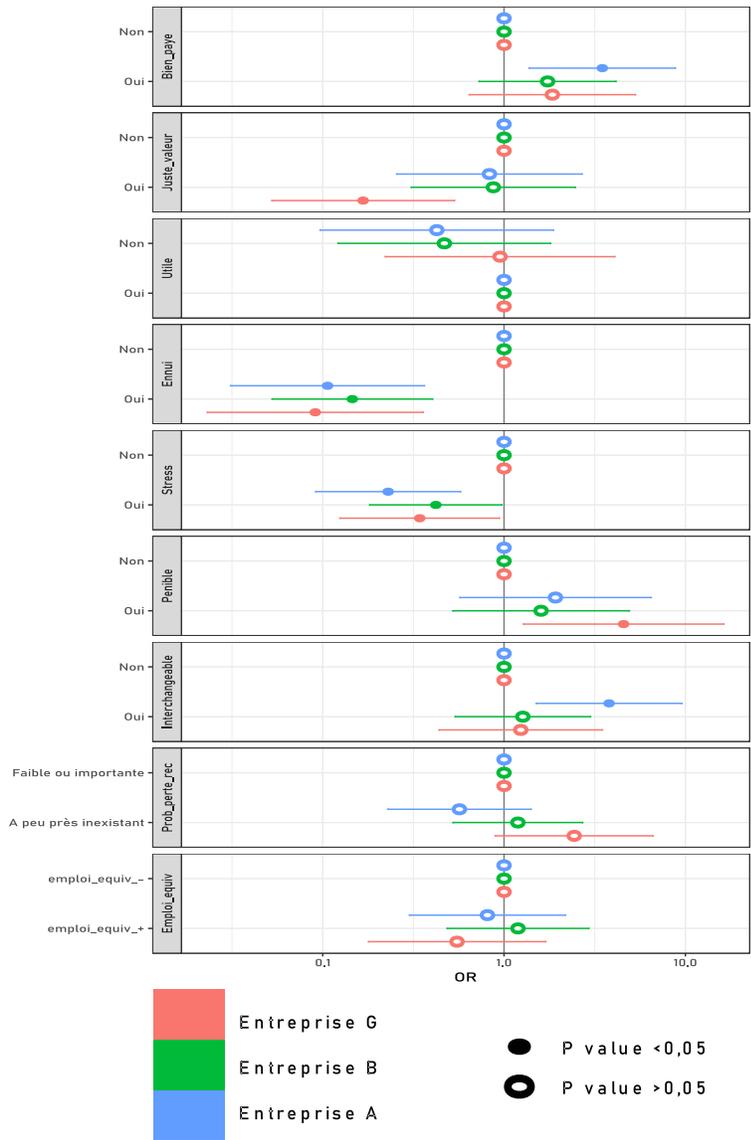


FIGURE 52 – Analyse des variables prédisant l'appartenance aux entreprises de sous-traitance

C.4 Analyse des revenus du foyer

Part salaire revenu foyer		N = 279
poids_salaire		
Minimum		4
Maximum		100
Moyenne		60
Médiane		60
Manquant		97

TABLE 67 – Part du salaire dans le revenu du foyer

Variable			N = 279
Salaire	Ariary	Euros	
Minimum	100 000	23,97	
Maximum	17 000 000	4082,2	
Moyenne	1 419 938	340,97	
Médiane	700 000	168,09	
Manquant	38		

TABLE 68 – Revenus du foyer par mois

C.5 Analyse des perspectives professionnelles des annotateurs

Variable	N = 296
Probabilité promotion	
A peu près inexistant	39 (15,1%)
Élevé	103 (39,9%)
Très élevé	18 (7%)
Faible	98 (38%)
Unknown	37
Probabilité retrouver un emploi équivalent	
Plutôt difficile	147 (67,4%)
Plutôt facile	56 (25,7%)
Très facile	11 (3,7%)
Impossible	4 (1,4%)
Unknown	77

¹ n (%)

TABLE 69 – Probabilité d’avoir des évolutions professionnelles dans le secteur

Variable	N = 279 ¹
Appréciation extérieure	
Plutôt bien vue	107 (41,8%)
Appréciée par les uns, critiquée par les autres	62 (24,2%)
Plutôt mal vue	2 (0,8%)
Pas connue	85 (33,2%)
Manquant	39

¹n (%)

Variable	N = 279 ¹
Enfant même métier	
Non	130 (50,8%)
Oui	126 (49,2%)
Manquant	38

Table des figures

1	Nombre de publications sur le micro-travail par an. L'année 2020 n'est pas complète, notre collecte de données s'étant arrêtée en juillet 2020	52
2	Nombre de publications par concept mobilisé	53
3	Nombre de publications par plateformes étudiées	53
4	Nombre de publications par faces de marché étudiées	54
5	Carte des pays enquêtés dans les études sur le micro-travail. Lecture : plus un pays a été étudié dans la littérature sur le <i>micro-travail</i> , plus il est affiché en sombre sur la carte.	55
6	Cartographie des concepts présents dans littérature scientifique lié au micro-travail. Lecture : les concepts sont classés selon qu'ils évoquent davantage les plateformes ou l'IA	70
7	Modes d'externalisation du travail des données	71
8	Schéma général de notre méthodologie	77
9	Nombre de documents par année pour le corpus d'articles scientifiques "Mturk". Lecture : ce corpus correspond à une collecte d'articles utilisant la plateforme Amazon Mechanical Turk quelque soit la discipline scientifique	98
10	Nombre de documents par discipline scientifique	99
11	Capture d'écran du logiciel d'annotation de la caisse de cantine développé par Automatik	149
12	Capture d'écran du logiciel de gestion de l'annotation de l'entreprise Automatik	150
13	Capture d'écran du guide d'annotation d'un projet d'IA appliqué aux vêtements	152
14	Capture d'écran du guide d'annotation d'un projet d'IA appliqué aux vêtements	153
15	Capture d'écran d'un échange entre des annotateurs et des chefs de projet de l'entreprise Automatik	160
16	Capture d'écran d'un échange entre des annotateurs et des chefs de projet de l'entreprise Automatik	161
17	Typologie des entreprises d'intelligence artificielle (source : desk research / Crunchbase.com)	172
18	Localisation des entreprises d'IA (source : desk research / Crunchbase.com)	175
19	Localisation des entreprises de travail des données (source : desk research / Crunchbase.com)	176
20	Chaîne de valeur BPO et plateforme ouverte d'une firme focale de l'IA	187
21	Chaîne de valeur BPO et plateforme ouverte d'une firme focale de l'IA	191
22	Chaîne de valeur BPO intégré à une plateforme et plateforme ouverte d'une firme focale de l'IA	197
23	Les règles d'annotations des armoires de fibre optique (Slack de Automatik)	286

24	Les règles d'annotations des armoires de fibre optique (Slack de Automatik)	286
25	Exemple d'images satellites à annoter (Potter, 2020)	287
26	Exemple d'annotation par point	289
27	Exemple d'annotation pour le développement d'une voiture autonome	290
28	Exemple d'une annotation par zone de l'image	291
29	L'annotation d'un plateau repas pour une caisse de cantine automatique	298
30	Exemple d'un projet d'annotation d'une image satellite	302
31	Exemple d'un projet d'annotation pour la classification d'avions	303
32	Data work production chain. Source : Authors' elaboration. Source : Le Ludec, Cornet et Casilli, 2023	318
33	Exportations annuelles en million de dollars des principaux pays exportateurs de services dans les pays d'Afrique francophone (hors Maroc, source : OMC)	330
34	Répartition par année de création des entreprises franches encore actives dans l'externalisation de services	334
35	Répartition par types d'activité des entreprises franches encore actives dans l'externalisation de services	335
36	Répartition par nationalité des propriétaires des entreprises franches encore actives dans l'externalisation de services	335
37	L'externalisation de la saisie de données aux États-Unis dans les années 70 (source : Computerworld the Newsweekly for the Computer Community 1974-07-17 Vol 8 Iss 29)	350
38	Distribution des salaires des annotateurs de données. Lecture : les valeurs sont exprimées en ariary (1 euros = 5000 ariary) et les points rouges sont des "outliers" du fait de leur niveau de salaire	376
39	Distribution de la position sociale des annotateurs de données. Lecture : les valeurs correspondent à la moyenne des scores associés aux professions indiquées dans le générateur de position sociale (voir la méthodologie)	377
40	Contribution des variables de notre questionnaire à l'ACM. Lecture : la position des variables sur les axes déterminent indiquent les variables déterminantes dans la composition des axes	381
41	Classification des annotateurs en trois groupes homogènes. Lecture : la position des individus est déterminée par les variables qui composent les axes	381
42	Le nombre d'emplois occupés par les annotateurs de données	388
43	Comparaison de l'ancienneté des annotateurs sur le marché du travail entre les trois entreprises étudiées à travers notre questionnaire	389

44	Typologie des trajectoires professionnelles des travailleurs des données. Lecture : chaque classe est constituée d'individus homogènes dans leur trajectoire professionnelle et est déterminée par la fréquence d'individus dans chaque "état"	390
45	Distribution de la durée dans l'emploi actuel	395
46	Comparaison des salaires entre les trois entreprises étudiées à travers notre questionnaire. Lecture : 1 euros correspond à 5000 ariary au cours du 15/11/2023	405
47	Trois formes d'intégration des annotateurs sur le marché du travail	410
48	Distribution des salaires des annotateurs de données	415
49	Comparaison des salaires entre les trois entreprises étudiées à travers notre questionnaire	416
50	Variable explicative du fait que les annotateurs soient heureux que leurs enfants fassent le même métier qu'eux	429
51	Mail de réponse à une prise de contact pour diffuser le questionnaire	496
52	Analyse des variables prédisant l'appartenance aux entreprises de sous-traitance	555

Liste des tableaux

1	Description des entreprises d'IA étudiées	83
2	Calendrier et modalités de la collecte de données principale	88
3	Mode de production des données des entreprises d'IA étudiées	119
4	Classification des sous-traitants de Google pour les données	178
5	Tableau de synthèse des modes d'externalisation de l'annotation de données	180
6	Tableau de synthèse des modes d'externalisation de l'entreprise Appen	198
7	Typologie des fournisseurs de données	203
8	Typologie des entreprises d'IA selon leur besoin en matière d'annotation (d'après l'auteur)	209
9	Appartenance à une entreprise de notre échantillon croisé avec les demandes de précision aux clients	307
10	Appartenance à une entreprise de notre échantillon croisé avec le sentiment d'autonomie	307
11	Niveau d'études selon l'ancienneté sur le marché du travail	355
12	Niveau de français selon la formation professionnelle aux langues. Lecture : pour des raisons de taille de l'échantillon, nous avons regroupé les "intermédiaire" et les "débutant" dans une modalité "pas avancé"	362
13	Niveau de français selon le fait de posséder un smartphone	363
14	Niveau de français selon le niveau d'études	364
15	Modèle de régression ordinale pour tester le niveau de français selon plusieurs variables	365
16	Composition du foyer	375

17	Métriques du générateur de position	378
18	Descriptifs des variables utilisées pour l'ACP	380
19	Recodage des variables du module parcours professionnel	387
20	Conditions d'emploi dans les entreprises de sous-traitance malgaches	399
21	Structure des coûts de l'entreprise A (en euros) : comparaison mode informel contre mode forme	401
22	Conditions d'emploi	402
23	Rapport des travailleurs au salaire	406
24	Interchangeable selon l'entreprise	406
25	Probabilité de perdre son emploi	407
26	Autonomie et contacts avec clients	408
27	Probabilité de perdre son emploi	411
28	Salaire des annotateurs en fonction de leur position hiérarchique	417
29	Sentiment d'être bien payé selon la taille du foyer	418
30	Probabilité de promotion selon l'entreprise	426
31	Les perspectives professionnelles des annotateurs de données	430
32	Trois groupes de travailleurs des données de la classe moyenne urbaine	439
33	Trois intégrations des travailleurs des données dans le marché du travail malgache	441
34	Trois intégrations des travailleurs des données dans le marché du travail	442
35	Description des personnes enquêtées au sein des entreprises d'IA	491
36	Observations	492
37	Description des personnes enquêtées au sein des entreprises d'annotation de données	492
38	Codage pour les travailleurs des données	493
39	Codage pour les dirigeants des entreprises d'annotation et les salariés d'entreprises d'IA	494
40	Codage pour les salariés des entreprises d'IA	495
41	Professions du générateur de position et scores associés	499
42	Variables socio-démographiques	544
43	Origine sociale du conjoint	545
44	Secteur d'activité du conjoint	546
45	Revenus du foyer par mois	546
46	Type d'établissement fréquenté	547
47	Niveau d'études	547
48	Formation professionnelle	548
49	Niveau de langue en français	548
50	Origine sociale du père	549

51	Origine sociale de la mère	549
52	Genre par la classe	549
53	Âge par la classe	550
54	Statut matrimonial par la classe	550
55	Habiter avec ses parents par la classe	550
56	Habiter avec ses enfants par la classe	550
57	Études par la classe	551
58	Encore étudiant par la classe	551
59	Position sociale par la classe	551
60	Revenus du foyer par la classe	551
61	Accès à une assurance employeur par entreprise	552
62	Rythme de travail	552
63	Conditions de travail	553
64	Choix du rythme selon l'entreprise	554
65	Travail la nuit et le week-end selon l'entreprise	554
66	Présence d'enfants dans le foyer selon l'entreprise	554
67	Part du salaire dans le revenu du foyer	556
68	Revenus du foyer par mois	556
69	Probabilité d'avoir des évolutions professionnelles dans le secteur	557

Titre : Des humains derrière l'intelligence artificielle. La sous-traitance du travail de la donnée entre la France et Madagascar

Mots-clés : Intelligence artificielle, sous-traitance, annotation de données, développement, Sud Global, travail des données

Résumé : Le développement effréné de l'intelligence artificielle suscite de plus en plus de craintes quant à ses impacts sur le travail et sur la société. Face aux inquiétudes liées à la fin du travail, nous prolongeons les perspectives de recherche ouvertes par la littérature en montrant que, derrière l'explosion récente des projets d'IA commerciales des pays du nord, on retrouve un nombre croissant de travailleurs de la donnée situés dans des pays du Sud. Notre thèse porte ainsi sur une des réalités les plus concrètes de l'IA : son rôle dans la division internationale du travail. Nous nous intéressons plus spécifiquement aux modèles développés dans un contexte commercial en nous demandant comment sont-ils fabriqués et quelles sont leurs conséquences sur le travail ?

Nous nous appuyons sur des données originales provenant d'une enquête par questionnaire et par entretiens réalisée auprès de l'ensemble des acteurs de la chaîne de production des SIA français. En France d'abord, nous avons interrogés des concepteurs de modèles et des utilisateurs de systèmes d'IA. A Madagascar ensuite, où nous avons questionnés des travailleurs des données et des dirigeants

d'entreprises de sous-traitance spécialisée dans cette activité. Notre thèse articule alors quatre niveaux d'analyses. L'étude de l'organisation du *travail de l'IA* sous l'angle des mécanismes de sous-traitance du secteur de l'IA. Nous montrons que, même s'ils opèrent à distance, les travailleurs sont fortement intégrés au processus de production, ce qui s'explique par la nécessaire maîtrise de la qualité des données et donc des modèles d'IA. L'analyse des *conséquences de l'IA sur le travail* de part l'examen des activités externalisées essentielles à son développement. Cette partie nous permet de montrer qu'il s'agit d'un travail qualifié qui s'inscrit dans un secteur offrant de nombreuses activités "*complémentaires*" qui suppléent le fonctionnement des modèles d'IA. Nous montrons également que cette industrie opère dans un contexte post-coloniale qui laisse apparaître un "*extractivisme de données*" synonyme de la mise en place d'une véritable "*chaîne de précarité globale*". Nous contribuons finalement à la sociologie de l'IA, en analysant *ce que le travail de l'IA explique du fonctionnement de l'intelligence artificielle*, notamment en matière de "biais".

Title : Humans behind artificial intelligence. Outsourcing data processing between France and Madagascar.

Keywords : Artificial Intelligence, Outsourcing, Data annotation, Global South, Offshoring, Data work

Abstract : The unrestrained development of artificial intelligence is giving rise to growing fears about its impact on work and society. In the context of concerns about the end of work, we expand the research perspectives opened up by the literature to show that, behind the recent explosion of commercial AI projects in Northern countries, we find a growing number of data workers located in the Global South. Our work thus focuses on one of the most concrete realities of AI : its role in the international division of labour. We specifically look at AI models developed in a commercial context, asking how they are developed and what their consequences are for work.

Our research is grounded in original data from questionnaire and interview with all actors in the French AI production chain : in France, we interviewed model developers and users of AI systems; and in Madagascar, we interviewed data workers and managers of subcontracting firms specialising in this activity. Our thesis is then based on four levels of analysis.

A study of the work organisation of AI from the perspec-

tive of subcontracting mechanisms in the field of AI. We show that, although they operate remotely, workers are strongly integrated into the production process, which is explained through the need of monitoring data quality and therefore the performance of AI models. An analysis of the consequences of AI on work by analysing outsourced activities that are essential for its development. This part allows us to show that these are qualified jobs belonging to a sector that offers many 'complementary' activities that complement the functioning of AI models. We also show that this industry operates in a post-colonial context that gives rise to a 'data extractivism' that gives rise to a 'data extractivism' that is synonymous with AI models' development. We also show that this industry operates in a post-colonial context, giving rise to a 'data extractivism' that is synonymous with the establishment of a veritable 'chain of global precarity'. Finally, we contribute to the sociology of AI by analysing what the work of AI explains about the functioning of artificial intelligence, especially in terms of 'bias'.