



HAL
open science

Le travail et l'emploi à l'épreuve de l'IA : État des lieux et analyse critique de la littérature

Flore Barcellini, Tamari Gamkrelidze, Nathalie Greenan, Annie Jolivet,
Moustapha Zouinar

► To cite this version:

Flore Barcellini, Tamari Gamkrelidze, Nathalie Greenan, Annie Jolivet, Moustapha Zouinar. Le travail et l'emploi à l'épreuve de l'IA : État des lieux et analyse critique de la littérature. IRES. 2024. hal-04722240

HAL Id: hal-04722240

<https://hal.science/hal-04722240v1>

Submitted on 4 Oct 2024

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.

le **cnam**
crtcd

le **cnam**
ceet

FO
la force syndicale

iRES
Institut de Recherches
Économiques et Sociales

Le travail et l'emploi à l'épreuve de l'IA : Etat des lieux et analyse critique de la littérature

Flore Barcellini, CRTD - Le Cnam

Tamari Gamkrelidze, CRTD - Le Cnam

Nathalie Greenan, LIRSA - CEET- Le Cnam

Annie Jolivet, CRTD - CEET- Le Cnam, chercheure associée à
l'IRES

Moustafa Zouinar, CRTD - Le Cnam et Orange Labs

Mars 2024

Rapport de recherche financé par l'Agence d'objectifs de l'Institut
de Recherches Économiques et Sociales dans le cadre d'une
convention avec la CGT - Force Ouvrière

Sommaire

1. INTRODUCTION GENERALE.....	6
1.1. DES EVOLUTIONS ACTUELLES DE L'IA	6
1.2. OBJECTIF DU RAPPORT ET PERSPECTIVES ADOPTÉES.....	8
1.3. CONTRIBUTIONS DU RAPPORT	10
2. INTELLIGENCE ARTIFICIELLE : DE SES POTENTIALITES A SES LIMITES	12
2.1. DEFINITIONS ET APPROCHES DE L'INTELLIGENCE ARTIFICIELLE	12
2.1.1. <i>Définitions</i>	12
2.1.2. <i>Approches de l'IA : approche symbolique et approche connexionniste</i>	13
2.2. LES GRANDES FONCTIONNALITES DE L'IA	15
2.3. LIMITES DE L'IA ET REVERS DES PROMESSES QU'ELLE SUSCITE	16
2.3.1. <i>Opacité, incertitude, qualité des données et biais</i>	16
2.3.2. <i>La face cachée des promesses de l'IA : vers plus automatisation et de subordination algorithmique ?</i>	18
A RETENIR	21
3. TRANSFORMATIONS DE L'EMPLOI LIEES A L'IA.....	22
3.1. LE PROGRES TECHNIQUE BIAISE AU FIL DES REVOLUTIONS TECHNOLOGIQUES	22
3.1.1. <i>L'abandon de l'hypothèse d'un progrès technique biaisé en faveur du capital</i>	23
3.1.2. <i>L'ordinateur personnel biaisé selon la qualification et selon l'âge</i>	24
3.1.3. <i>Les TIC biaisées selon le caractère routinier des tâches et la polarisation des emplois</i>	24
3.1.4. <i>L'IA et la robotique repoussent la limite des tâches automatisable</i>	25
3.2. LE DETERMINISME TECHNOLOGIQUE AU CŒUR DES APPROCHES ECONOMIQUES DES CONSEQUENCES DE L'IA SUR L'EMPLOI 26	
3.3. REPOSITIONNER L'APPROCHE ECONOMIQUE A LA SOURCE DE LA TECHNOLOGIE ET DE SES USAGES : LE ROLE PROTECTEUR POUR L'EMPLOI DE LA CAPACITE D'APPRENTISSAGE DES ORGANISATIONS	28
A RETENIR	32
4. LES EFFETS SUR LE TRAVAIL ET SUR L'EMPLOI : QUELLES ARTICULATIONS ?	33
4.1. ANALYSER A LA FOIS LES EFFETS SUR LE TRAVAIL ET LES EFFETS SUR L'EMPLOI	33

4.1.1.	<i>Un exemple d'analyse monographique</i>	33
4.1.2.	<i>Un exemple d'études de cas</i>	35
4.1.3.	<i>Un exemple d'analyse prospective</i>	36
4.1.4.	<i>Des exemples récents d'effets sectoriels voire systémiques sur le travail et sur l'emploi</i>	38
4.2.	REGULATION DU DEVELOPPEMENT ET DES USAGES DE L'IA.....	41
4.2.1.	<i>Réglementer ou pas, avec quels instruments ?</i>	41
4.2.2.	<i>Dialogue social et accords collectifs</i>	45
	A RETENIR	47
5.	USAGES DE L'IA ET TRANSFORMATIONS DU TRAVAIL : DE LA PROJECTION AUX USAGES EN SITUATION REELLE DE TRAVAIL	48
5.1.	UTILITE, VALEUR ET RISQUES PERÇUS POUR LA PERFORMANCE DU TRAVAIL REALISE AVEC UNE IA.....	49
5.1.1.	<i>Un point de vue nuancé sur les bénéfices-risques de l'IA et une valeur a priori accordée aux tâches automatisables discutables</i>	49
5.1.2.	<i>Du besoin de fiabilité et de contrôle de la qualité des « décisions » des SIA</i>	50
5.1.3.	<i>L'IA au service de la fiabilité des actions</i>	52
5.1.4.	<i>Conséquences pour la santé et le sens du travail</i>	53
	Zoom sur les points de vue des acteurs sur l'IA dans le domaine juridique (Gamkrelidze, 2022)	54
	A Retenir	57
5.2.	L'IA A L'ÉPREUVE DES EXPERIMENTATIONS : UNE PERFORMANCE ACCRUE DANS LES TACHES REALISEES A L'AIDE DE L'IA, MAIS UNE PERFORMANCE A PONDERER.....	58
5.2.1.	<i>Des diagnostics d'imagerie accrus dépendants du niveau d'expertise et des caractéristiques techniques de l'IA</i>	58
5.2.2.	<i>L'IA générative : une source d'amélioration de la performance au travail ?</i>	59
5.2.3.	<i>Limites des résultats expérimentaux</i>	61
	A Retenir	62
5.3.	USAGES DE L'IA EN SITUATION DE TRAVAIL REEL : CONSEQUENCES SUR LES PROFESSIONNELS ET LEURS ACTIVITES.....	62
5.3.1.	<i>Vers une vision élargie de la valeur et de la performance d'une activité</i>	63
5.3.2.	<i>Des SIA augmentant ou entravant l'apprentissage et le développement de l'expertise</i>	64
5.3.3.	<i>Dualité entre opacité/explicabilité des systèmes, fiabilité et pertinence de l'action et</i>	

<i>dépendance organisationnelle des usages de l'IA</i>	64
5.3.4. <i>Le cas des chatbots : entre intensification du travail ? et ses entraves à l'interaction directe entre « clients » et experts</i>	67
Zoom sur les usages de SIA et ses conséquences dans le secteur de la radiologie (Gamkrelidze, 2022) .	69
5.3.5. <i>Vers une vision élargie et située de la performance</i>	72
A Retenir	74
5.4. LA CONDUITE DES TRANSFORMATIONS DU TRAVAIL PAR L'IA : UN IMPENSE DES DISCOURS SUR L'IA ET DES DEMARCHES DE CONDUITES DE PROJETS	74
5.4.1. <i>Des principes éthiques et des démarches centrées « humain » pour encadrer la conception et l'intégration de systèmes d'IA</i>	75
5.4.2. <i>Des intentions de transformations du travail liés à l'IA renvoyant à des logiques partielles et un impensé de la conduite de projets</i>	77
5.4.3. <i>Vers des démarches de conduite de projet centrées travail qui donnent une valeur à l'expérience et l'activité des professionnels</i>	79
A Retenir	81
6. RECOMMANDATIONS ET PERSPECTIVES	82
6.1. LES 4 PILIERS DU DEVELOPPEMENT D'USAGES SOUTENABLES DE L'IA DANS LA SPHERE PROFESSIONNELLE.....	82
6.2. PERSPECTIVES	85
6.3. PRECONISATIONS.....	86
REMERCIEMENTS	89
BIBLIOGRAPHIE	89
ANNEXES	102
ANNEXE 1 METHODOLOGIE DE RECHERCHE ET D'ANALYSE SYSTEMATIQUE DES ETUDES EMPIRIQUES.....	102
ANNEXE 2. METHODOLOGIES DE CONCEPTION D'IA « CENTREES HUMAIN »	110

1. Introduction générale

1.1. Des évolutions actuelles de l'IA

Des avancées technologiques dans le domaine de l'apprentissage dit automatique ou machine (*Machine Learning*) et l'élargissement des applications de cette technique ont suscité un regain d'intérêt inédit pour l'Intelligence Artificielle (IA) dans les champs politique et socio-économique. L'augmentation de la puissance de calcul des ordinateurs associée à la disponibilité de données massives (*Big Data*) et au perfectionnement des algorithmes ont en effet augmenté de façon significative les performances de l'IA. Au-delà de ces avancées, l'IA est également promue par un écosystème composé de startups, de cabinets de conseil, d'entreprises, d'institutions et de centres de recherche privés ou publics (Gamkrelidze, 2022) qui diffusent des discours autour de l'IA annonçant une nouvelle ère d'innovations technologiques et servicielles. D'aucuns estiment que ces innovations entraîneront des conséquences profondes sur le travail et l'emploi¹. Mais, nous le verrons, il existe en fait peu d'études empiriques qui permettent réellement de décrire, et encore moins de prédire de manière fiable ces conséquences. Au vu de ces discours et des enjeux de compétitivité et d'emploi associés à l'IA, plusieurs pays ont mis en place des plans stratégiques visant à soutenir et encadrer le développement de l'IA (États-Unis, Chine, France, etc.). Pour ce qui concerne la France, en 2018, Cédric Villani a publié un rapport conseillant le gouvernement français et l'Europe sur le développement de la stratégie française et européenne en IA. Depuis, le Programme National de Recherche en IA (PNRIA) a été lancé en France et de nombreuses actions de recherche ont été entreprises.

Cependant, il convient de rappeler que les débats autour des promesses et craintes associées à l'IA ne sont pas nouveaux et ont commencé à partir des années 1960-1970 avec le développement des premiers Systèmes d'Intelligence Artificielle (SIA), en particulier des systèmes dit « experts » (cf. Section 2). Basés sur de l'IA symbolique, ces systèmes étaient censés assister des travailleurs, par exemple des médecins dans le diagnostic médical ou des techniciens dans l'identification de pannes de machines. La promesse annoncée était déjà d'améliorer la performance des travailleurs dans ce type de tâches. De même, la question de l'effet de la diffusion des nouvelles technologies sur l'emploi n'est pas nouvelle et les économistes ont cherché à modéliser les effets du progrès technique sur l'emploi dès le début de l'ère industrielle (cf. section 3). Cependant, dans cette première période de l'IA, les craintes d'une diffusion massive de celle-ci au travail et de son effet sur l'emploi restaient marginales, du fait notamment des échecs du déploiement des systèmes experts dans les milieux de travail.

¹ Les notions de travail et d'emploi semblent être interchangeables pour la plupart des acteurs économiques (Guérin et al., 2021). Nous considérons ici le travail humain comme « *l'activité de production, de biens ou de services et l'ensemble des conditions de cet exercice* » (Guérin et al., 2021) (p. 38). Il englobe des résultats visés et produits, des tâches et leur organisation, un milieu, des acteurs du processus et des ressources. Tandis que l'emploi renvoie à « *une convention qui définit le cadre contractuel et institutionnel dans lequel le travail est réalisé* » (Guérin et al., 2021) p. 38). L'emploi se caractérise par un statut, une rémunération et des droits. Le cadre organisationnel est donc inhérent à l'emploi, c'est ce cadre qui définit les droits, les obligations et les conditions d'exercice du travail humain.

Aujourd'hui, les SIA sont plus performants dans différents domaines (traitement automatique du langage, traitement d'images, génération de contenus) et ils sont toujours envisagés comme un moyen d'amélioration de la performance dans différents secteurs professionnels (médical, juridique, industriel, ressources humaines, information et communication etc.) et d'automatisation des tâches répétitives, pour « libérer l'humain » des tâches pénibles (c'est-à-dire à forte exigences physiques, répétitives ou dangereuses) et/ou d'augmenter la performance et la productivité en permettant aux travailleurs de se concentrer sur des tâches dites « à plus forte valeur ajoutée » (Brynjolfsson & McAfee, 2014). La récente apparition de l'IA dite « générative » - ou plus largement les modèles dits de fondation² - suscitent quant à elle de plus fortes préoccupations concernant la transformation de l'emploi, des métiers et du travail. Les débats antérieurs sont donc aujourd'hui renouvelés par les questions à la fois d'ordre politique (gouvernance du développement de l'IA, stratégies nationales), socio-économique (effets sur l'emploi, création vs destruction d'emploi, etc.), éthique (conséquences humaines néfastes de l'IA) et juridique (réglementation, questions de responsabilité, etc.).

Dans ce contexte, trois grandes craintes émergent. Elles concernent principalement les conséquences de l'introduction de SIA sur l'emploi, et non sur le travail alors qu'il existe des risques réels de dégradation de la qualité et du sens du travail comme nous le verrons plus loin :

- La première crainte concerne un accroissement de l'automatisation qui pourrait conduire au remplacement massif de l'humain dans un grand nombre de secteurs professionnels, (Frey & Osborne, 2013).
- La deuxième concerne le développement d'emplois précaires pour les travailleurs de plateforme ou du clic (Casilli, 2019) ;
- La troisième qui tend plus vers les questions de travail concerne un contrôle accru des salariés par le management algorithmique (Zuboff, 2019 et infra), c'est-à-dire la gestion d'une organisation et de ses travailleurs et travailleuses via des algorithmes informatiques.

Cependant, les usages de l'IA et leurs effets potentiels sont souvent pensés avec une compréhension pauvre ou biaisée :

- Des relations entre développement technologique et emploi ;
- du travail et des ressources nécessaires aux travailleurs pour agir de manière performante et en santé (les connaissances et raisonnements mobilisés, le caractère central des collectifs de travail, l'entraide et la coopération, les marges de manœuvre disponibles dans le travail, le caractère développemental du travail, l'engagement du corps, les valeurs engagées au travail, le rôle de l'expérience), et sans réel questionnement sur ce qui donne de la valeur et du sens au travail (voir par ex. Coutrot et Perez, 2022).

² Modèle [d'intelligence artificielle](#) de grande taille, entraîné sur une grande quantité de données non étiquetées, qui peut être adapté à un large éventail de tâches (par exemple, ChatGPT, Dall-E)

1.2. Objectif du rapport et perspectives adoptées

L'objectif de ce rapport est de proposer un examen critique d'une part des promesses, des craintes et des discours « futurologistes » ou spéculatifs associées à l'introduction de l'IA dans le monde du travail, et d'autre part des travaux de recherche s'intéressant à ses conséquences sur l'emploi et le travail³. Pour ce faire, nous nous appuyons sur :

- **une approche critique des modèles économiques dominants** et des propositions alternatives permettant de penser les évolutions de l'emploi en lien avec l'introduction de l'IA dans la sphère professionnelle. Nous soutenons notamment que ces modèles s'appuient sur une vision réductrice de la façon de penser les effets du progrès techniques sur l'emploi et qu'ils sont fondées sur des données et des technologies émergentes antérieures.
- **une approche critique des promesses de performance accrue des activités humaines au travail** sous-tendues par les discours des promoteurs de l'IA. Cette critique s'appuie notamment sur des connaissances établies depuis longtemps en ergonomie sur les conséquences de l'automatisation et l'utilisation des technologies sur les activités concrètes (Bainbridge, 1983 ; Rabardel, 1995 ; Rabardel et Daniellou, 2005), sur les organisations du travail (de Terssac et Reynaud, 1992), la qualité du travail, la santé au travail (Volkoff, 2005) et le sens du travail (Perez et Coutrot, 2022). Elle s'appuie également sur un examen lui aussi critique des modalités de production de connaissances sur les potentialités de l'IA en lien avec une augmentation des capacités de l'humain et de ses performances cognitives. Nous verrons notamment qu'il n'existe que très peu d'études empiriques qui décrivent l'expérience des travailleurs engagés dans des activités dans des situations réelles de travail, et non purement expérimentales.

Ces approches sont héritières de perspectives qui permettent de penser les relations entre technologies et transformations du travail et de l'emploi, dont nous rappelons ici les fondements :

- **une perspective non déterministe des technologies** qui rejette le « **fatalisme technologique** », en rappelant que les usages organisationnels de la technique et les conséquences de celles-ci ne sont pas entièrement déterminés par ses caractéristiques techniques intrinsèques. Ces usages sont le produit des décisions et des activités d'acteurs divers (concepteurs, pouvoirs publics, décideurs d'entreprises ou d'institutions, travailleurs) qui peuvent en retour façonner la technique. Les conséquences de telle ou telle technologie résultent de ce que ces acteurs font avec elle : comme ils se l'approprient, comment ils l'utilisent, comment ils la mettent en place.
- **une perspective historico-culturelle** qui permet de penser le développement technologique dans le temps, et non comme une rupture faisant fi des expériences passées et/ou des autres technologies et changements organisationnels passés ou en cours. Cette perspective appelle

³ Sont exclus du champ de ce rapport la question du « travail du clic » ou des plateformes déjà traitée par ailleurs (voir p.ex. Casilli), ainsi que les effets « grand public » de l'introduction des IA génératives qui restent relativement récentes et qui n'obéissent pas totalement aux mêmes motifs d'introduction et d'usages : dans la sphère privée on peut choisir d'utiliser ou non les services d'une IA ce qui n'est pas le cas dans les entreprises ou institutions où la technologie peut être imposée.

notamment des réflexions sur la façon dont on construit l'usage futur potentiel d'une technologie intégrée dans une situation de travail en ne faisant pas fi du passé (par exemple de l'histoire et de la culture d'une entreprise, des savoir-faire existants qui pourront être remobilisés), c'est-à-dire sur la façon dont sont conduits les projets de transformation des situations de travail.

- **une perspective située donnant une valeur aux activités et expériences concrètes** d'usage intégré dans une activité de travail. Ceci est important car les activités humaines en situation sont encore pensées de manière mécaniste ou comportementaliste : agir au travail se résumerait pour beaucoup à appliquer des procédures pré-pensées (qui font partie de ce que les ergonomes appellent la *tâche*, renvoyant à ce qui est prescrit par les organisateurs du travail) sans mobilisation d'une intelligence pratique en situation. Or, comme l'ergonomie l'a bien montré depuis longtemps, l'action au travail n'est pas qu'une simple exécution de tâches ou de procédures : il l'activité doit être soigneusement distinguée de la tâche (le prescrit). En effet, agir au travail relève de processus bien plus complexes qui engagent les valeurs, l'expérience, les connaissances des travailleurs. Cette action s'inscrit par ailleurs toujours dans un collectif. Elle se concrétise dans des choix, des arbitrages, des ajustements dans l'action en fonction des situations rencontrées (ce que les ergonomes appellent *l'activité*). Ces arbitrages sont plus ou moins aisés et plus ou moins coûteux - notamment pour la santé - en fonction des marges de manœuvre qui sont laissées aux individus et au collectif dans des situations de travail parfois contraintes par des choix organisationnels aveugles aux exigences réelles d'un travail performant et préservant la santé (connaissances et savoir-faire construits au fil de l'expérience et souvent incorporés, besoin de coopération, latitude de prise de décisions et d'actions nécessaires...). Ceci est un point important qui amène à se questionner sur la valeur attribuée a priori aux tâches (et non aux activités).
- **Une perspective systémique** soulignant qu'il convient de penser les organisations comme des systèmes sociotechniques dont les composantes sont liées et toute modification d'une composante peut entraîner des modifications dans les autres, et des relations entre elles. Or les effets de l'introduction d'une technologie sur les autres composantes d'un système de travail sont souvent impensés. Les décisions d'introduction d'une technologie relèvent le plus souvent d'une forme de « solutionnisme technologique » (Morozov, 2013) - la croyance que la technologie est toujours la bonne façon de résoudre les problèmes humains ou organisationnels - souvent lié à une rationalité purement financière sans réelle analyse systémique des « problèmes » que l'on cherche à résoudre et de ses conséquences systémiques. Il est montré de longue date que l'introduction et l'usage de dispositifs techniques peut conduire à des transformations profondes non seulement des activités mais aussi de l'organisation du travail (procédures, horaires, modalités de coopération...) (Norman, 1993; Woods, 1996) avec des effets potentiellement néfastes sur la qualité et la fiabilité du travail ou encore la santé au travail (physique et psychologique). Ces résultats établis doivent aussi être pris en compte dans les réflexions sur les SIA actuels.
- **une perspective empirique**, qui marque la nécessité d'analyser les transformations du travail sur la base d'études empiriques pour dépasser les réflexions spéculatives. Cependant, il est

important de noter que les études qui documentent les transformations du travail liées à l'IA commencent à apparaître mais restent peu nombreuses soit parce qu'il serait difficile pour les chercheurs et chercheuses d'accéder à des terrains de recherche en situation réelle de travail, soit parce que les projets de conception de SIA n'aboutiraient pas et/ou resteraient à des stades expérimentaux. Nous y reviendrons en conclusion car ceci appelle le développement d'observatoires de ce que l'on peut voir comme les expérimentations en cours de l'introduction de l'IA dans le monde du travail.

Un premier corolaire de ces perspectives est que nous considérons qu'il n'existe pas de propriétés universelles liées à l'usage de l'IA au travail : le travail avec une IA ne serait pas plus performant, augmenté ou soulagé quelles que soient les situations, le motif de son introduction ou encore la manière dont ce projet d'introduction a été conduit ; et inversement pour les risques de dégradation de la qualité du travail ou de subordination. Un second corolaire relève de l'importance de l'encadrement des choix stratégiques et organisationnels liés à l'introduction de l'IA par des processus de régulation (dialogue social, conduites de projet...). Nous reviendrons sur ces deux points dans la suite.

1.3. Contributions du rapport

Dans une première partie de ce rapport, nous proposons un cadrage de ce que l'on appelle Intelligence Artificielle et de ses approches (symbolique ou connexionniste), ainsi que des cinq grandes fonctionnalités techniques des SIA (recherche, reconnaissance et analyse d'informations ; prédiction et diagnostic, recommandation, génération de contenu nouveau, exécution autonome d'action), leurs limites techniques (opacité, incertitude sur l'évolution dans le temps des systèmes, questions sur les données utilisées pour l'apprentissage) et le revers de leurs promesses (mirage de la collaboration humain-technologie, risque accru d'automatisation, management algorithmique).

Les craintes sur l'emploi étant une préoccupation majeure liée à l'introduction des SIA, nous présentons dès la deuxième partie l'analyse critique des modèles économiques décrivant les conséquences de l'IA sur l'emploi. Nous soulignons en particulier : (1) leur fondement techno-déterministe, basé sur des qualifications ou des tâches pensées de manière extrinsèque aux expériences et activités concrètes – et (2) l'absence de prise en compte de la capacité d'apprentissage des organisations comme un facteur à la fois d'innovation et protecteur pour l'emploi. Ceci prône pour le développement d'une approche économique alternative basée sur la capacité d'apprentissage des organisations. Cette analyse critique appelle le besoin d'identifier de manière empirique les effets de l'IA sur le travail et l'emploi, notamment (1) les liens entre projets d'introduction d'IA, transformations conjointes du travail et de l'emploi et leur cadrage par des processus de régulations ; et (2) les transformations du travail concrètes et leurs conduites.

Dans une troisième partie, nous poursuivons donc l'exploration des enjeux d'emploi liés à l'IA, mais cette fois-ci en tentant de les articuler aux questions de travail pour arriver à une analyse critique des régulations cadrant ces transformations. Cette analyse montre que les travaux qui permettent d'articuler

conjointement les effets sur l'emploi et les effets sur le travail sont relativement peu nombreux. Une partie des effets sur l'emploi est peu visible (délocalisation, sous-traitance, départs naturels, réduction des effectifs déjà engagée). Les travaux sur la régulation de l'IA soulignent la nécessité de penser une régulation d'ensemble pour pallier les incohérences et les lacunes des instruments juridiques actuels. La possibilité de réguler les conséquences des SIA sur le travail et l'emploi dépend beaucoup du modèle de relations professionnelles. De grandes différences existent entre les pays.

Enfin, dans la dernière partie nous tentons de nous approcher des conséquences de l'IA sur les activités concrètes, en examinant les usages - projetés ou réels des SIA – et leurs conséquences concrètes sur la base de différentes études empiriques : le recueil du point de vue *a priori* des utilisateurs potentiels de l'IA ; des études expérimentales mettant en avant les performances accrues de travail avec une IA dans certains contextes, qui restent néanmoins expérimentaux, et des usages en situations réelles. Nous montrons ainsi que les transformations de l'IA sur le travail sont souvent en décalage avec les promesses de performance attendue et que le travail, sa valeur réelle et sa dimension systémique, sont souvent des impensés de ces transformations, de même que la conduite des projets d'intégration d'IA.

Ces contributions sont synthétisées à la fin de chaque section d'intérêt dans des encarts « A retenir », nous illustrons également certains résultats emblématiques dans deux secteurs (radiologie et juridique) par des « Zoom sur... » tirés du travail réalisé par Tamari Gamkrelidze dans le cadre de sa thèse de doctorat (Gamkrelidze, 2022).

Nous concluons sur des recommandations décrivant 4 piliers pour le développement d'usages soutenables de l'IA : dialogue social, développement de la capacité d'apprentissage des organisations, conduites de projets participatives et centrées travail, et expérimentations in situ pour comprendre et documenter les évolutions rapides liens, IA, travail, emploi.

2. Intelligence artificielle : de ses potentialités à ses limites

Nous revenons ici sur les définitions de l'IA ainsi que les deux approches – symbolique et connexionniste – à son fondement. Ceci nous permet de porter un point de vue nuancé sur (1) la notion d'intelligence artificielle alors même que la définition d'intelligence humaine fait l'objet de débat ; (2) ce qui renvoie à l'intelligence des machines dans l'approche connexionniste – en fait la construction de modèles probabilistes sur la base de données souvent structurées par des humains. Nous rappelons ensuite brièvement les cinq grandes tâches ou fonctionnalités associées aux SIA (recherche, reconnaissance et analyse d'informations, diagnostic et prédiction, recommandation, exécution d'actions autonomes, génération de contenu). Enfin, nous discutons des promesses associées à l'IA et des revers qu'elle suscite en lien avec le fonctionnement de ces systèmes (opacité, incertitude sur l'évolution dans le temps des systèmes, questions sur les données utilisées pour l'apprentissage) et la face cachée des promesses de l'IA (mirage de la collaboration humain-technologie, automatisation, management algorithmique).

2.1. Définitions et approches de l'Intelligence Artificielle⁴

2.1.1. Définitions

Depuis son émergence, l'IA reste difficile à définir et circonscrire. A l'origine, lors de la conférence de Darmouth en 1956 qui avait réuni les chercheurs qui l'ont formalisée, l'IA était associée à l'hypothèse selon laquelle il pouvait être possible de reproduire « sur ordinateurs » les facultés cognitives humaines (p.ex. perception et représentation de connaissances, apprentissage, raisonnement, interaction et adaptation). Aujourd'hui, l'Intelligence Artificielle peut être vue comme un domaine technoscientifique, à la fois :

- une discipline à la croisée des mathématiques, de l'informatique et des sciences cognitives,
- et un ensemble d'algorithmes, de systèmes informatiques, de machines et plus largement de technologies sous différentes formes (logiciels, robotiques etc.) qui s'inspirent de ou visent à imiter des facultés cognitives humaines comme la perception, la production et la compréhension du langage naturel, la représentation des connaissances, ou encore le raisonnement (Gamkrelidze et al., 2021).

⁴ Cette section est adaptée de la thèse de Tamari Gamkrelidze (2022).

McCarthy (2007), considéré comme « l'inventeur » du terme « Intelligence Artificielle », définit ainsi l'IA comme la science et l'ingénierie de conception des machines « intelligentes » et plus spécifiquement des logiciels informatiques « intelligents ».

La majorité des définitions existantes reprennent ces éléments alors même que la définition de l'intelligence humaine fait l'objet d'un débat (Legg & Hutter, 2007) et que la pertinence du terme « intelligence artificielle » est régulièrement remis en question (De Ganay & Gillot, 2017; Julia, 2019).

2.1.2. *Approches de l'IA : approche symbolique et approche connexionniste*

Depuis les années 1950, les recherches sur l'IA ont donné lieu à des avancées technologiques liées principalement à l'automatisation de la production et la compréhension du langage naturel, de la représentation des connaissances humaines, du raisonnement logique (qui n'est qu'une forme des raisonnements humains), l'apprentissage dit automatique (détermination d'informations jugées pertinentes à partir d'un ensemble de données d'entraînement), et la planification d'actions ou la prise de décisions. Deux approches du fonctionnement des SIA coexistent : **l'approche symbolique et l'approche connexionniste** (Cardon et al., 2018; McAfee & Brynjolfsson, 2017; Russel & Norving, 2010)⁵.

L'approche symbolique, qui était dominante jusqu'aux années 1980, repose sur l'utilisation de symboles et de règles représentant des connaissances et des raisonnements. Cette approche consiste ainsi à écrire des règles, par exemple de type si-alors, pour développer un système capable de réaliser une tâche cognitive (par exemple, comprendre une phrase sur la base d'une analyse de ces constituants). Les connaissances et les raisonnements sont ainsi explicitement programmés dans un algorithme informatique. Les recherches menées en IA symbolique ont abouti à des applications concrètes plus ou moins abouties (Trappl, 1985) comme :

- les systèmes « experts » ou à *base de connaissances*, qui étaient censés reproduire le raisonnement d'experts humains à l'aide de modèles informatiques basés sur des connaissances et des règles (Dubus, 1995). Par exemple, Mycin, l'un des premiers systèmes « experts », avait été développé pour réaliser de manière automatisée du diagnostic médical « comme un médecin », en partant d'une modélisation des raisonnements et des connaissances que les médecins mobilisent pour poser un diagnostic. Il était envisagé que ces systèmes puissent ainsi remplacer ou assister les humains dans le cadre d'activités professionnelles⁶;
- les *systèmes de traitement du langage* naturel pour la traduction automatique ou encore la compréhension et la génération de textes ;

⁵ L'Intelligence Artificielle Distribuée (IAD) est une troisième approche qui a vu le jour dans les années 1970. Inspirée de l'intelligence collective des insectes sociaux ou encore des sociétés humaines (Institut de Recherche en Informatique de Toulouse, 2001), elle étudie des systèmes multi-agents en visant à concevoir plusieurs agents artificiels avec des capacités d'organisation dans le but d'accomplir une tâche de manière collective.

⁶ Un système expert correspond à « un logiciel à base de connaissances dérivées de l'expertise humaine et capable d'une part de proposer une solution à des problèmes d'un domaine précis et d'autre part d'expliquer les modalités de raisonnement l'ayant conduit à retenir cette solution » (De Terssac et al., 1988) p. 463.

- enfin, *la robotique* qui est considérée comme une des applications importantes de l'IA (Trappl, 1985).

L'approche connexionniste, a été initiée dès les années 1940 par les travaux de McCulloch et Pitts sur les neurones formels (sorte de fonction mathématique simulant le fonctionnement d'un neurone), et s'inspire précisément des réseaux de neurones. Dans ce cadre, la représentation des connaissances et le fonctionnement général du SIA (SIA) ne repose pas sur des règles écrites par le concepteur – et donc programmées informatiquement - mais sur des nœuds, des coefficients, des poids qui s'ajustent selon le type de tâche à réaliser et des données d'entrée. Cet ajustement, qui est principalement de l'optimisation mathématique, correspond au résultat de ce qui est appelé « l'apprentissage » de la machine. C'est cette approche qui domine aujourd'hui, grâce notamment à l'une de ses branches la plus populaire, l'apprentissage profond (*Deep Learning*). L'apprentissage profond est utilisé dans un grand nombre d'applications de l'IA (vision par ordinateur, traitement automatique du langage naturel, reconnaissance de la parole, etc.). Les techniques de l'apprentissage machine ont pu être réinvesties grâce à l'augmentation de la puissance de calcul des machines et les données massives (*Big Data*), essentielles pour le fonctionnement des algorithmes basés sur ce type d'apprentissage. Concrètement, pour réaliser une tâche, la machine « apprend » à partir de beaucoup d'exemples préparés – des données structurées - c'est-à-dire annotés par un humain (par exemple, des clichés radiologiques pour apprendre au système à reconnaître des anomalies). C'est ce qu'on appelle l'apprentissage supervisé dans lequel le système apprend à construire des catégories (par exemple d'objets) et qui est le mode d'apprentissage le plus couramment utilisé⁷ (LeCun et al., 2015). Une autre technique, l'apprentissage non supervisé, consiste à apprendre de manière autonome, c'est-à-dire sans que les données soient étiquetées. Il existe une troisième technique d'apprentissage dite par renforcement dans laquelle le système apprend en interagissant avec l'environnement via un processus algorithmique de « punition-récompense ». Le résultat de l'apprentissage de la machine est généralement la construction d'un "modèle" qui renvoie à ce que le système a « appris » et à la manière dont il a ajusté les paramètres (poids) pour réaliser une tâche (par exemple, reconnaître un objet ou prédire un mot dans une phrase) pour un degré de performance donné.

Les premières périodes de développement de l'IA (entre les années 60 et 80-90) ont été suivies de périodes de ralentissement des recherches (dits « hivers » de l'IA avec une forte baisse des financements) dues essentiellement à l'incapacité des recherches et de systèmes développés à tenir leurs promesses. Mais aujourd'hui, grâce à l'apprentissage profond, l'apparition récente de nouvelles approches comme les *transformeurs* (architecture particulière d'apprentissage profond) et les algorithmes de diffusion (méthodes de génération d'images), et le développement de l'IA générative, de nouvelles promesses ou plutôt un renouvellement des promesses émergent. D'aucuns voient dans cette évolution une révolution technologique qui aura de profondes répercussions économiques et sociales.

⁷ Il y a également deux autres types d'apprentissage moins utilisés que nous ne développerons pas ici : l'apprentissage par renforcement (basé sur le principe d'essai-erreur et de « récompense ») et l'apprentissage non supervisé (qui cherche à identifier des catégories de structures sous-jacentes aux données non annotés).

De manière encore plus marquée que dans la période des systèmes experts, le monde économique (des entreprises, de l'industrie, des services) y voit une opportunité inédite d'automatisation et de transformation des systèmes de production. Sur ce point, il importe de comprendre qu'une nouvelle dynamique circulaire s'installe : l'accroissement des performances de l'IA grâce à ces avancées ouvre de nouvelles perspectives d'automatisation, qui sont vues par les entreprises comme moyen d'optimisation de la performance financière (cf. section 5.4), et par conséquent, en misant sur l'IA, ces entreprises vont a priori renforcer le développement de l'IA. L'engouement autour de l'IA générative semble aller dans ce même sens.

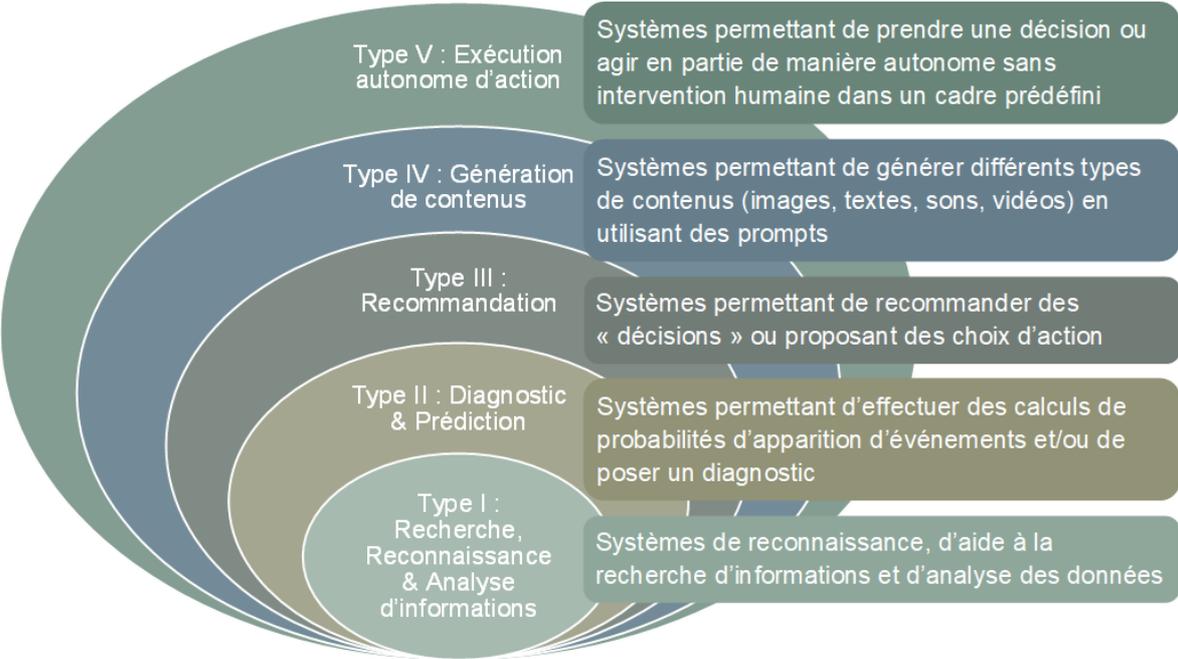
2.2. Les grandes fonctionnalités de l'IA

De nombreux secteurs professionnels sont aujourd'hui concernés par la diffusion des SIA, par exemple, l'industrie, la défense et la sécurité, la justice, l'éducation, la santé, la relation client, la finance, le transport, l'environnement, l'agriculture, ou encore les ressources humaines.... La liste ne cesse de s'allonger. Bien que ces systèmes aient été déployés dans certains environnements de travail, il semble que la plupart des SIA spécifiques – hormis l'IA générative grand public - soient encore dans la plupart des cas en phase de recherche, de prototype ou d'expérimentation, du moins en France.

De façon schématique, on peut distinguer cinq types de systèmes selon les fonctionnalités portées par l'IA (Figure 1. Les types de fonctionnalités des SIA de « nouvelle » génération (d'après Gamkrelidze, 2022)

).

Figure 1. Les types de fonctionnalités des SIA de « nouvelle » génération (d'après Gamkrelidze, 2022)



Le premier type de SIA (type I) se limite aux systèmes de reconnaissance automatique d'éléments (par exemple, dans du texte, des sons, ou des images), d'aide à la recherche d'informations ou encore d'analyse automatique des données (par exemple, les systèmes qui analysent de grands corpus de données pour extraire des informations). Apparu plus récemment, un deuxième type de SIA (type II) regroupe des systèmes plus sophistiqués permettant d'effectuer des calculs de probabilités d'apparition d'événements et/ou de poser un diagnostic (par exemple, médical). Les systèmes du troisième type (type III) vont au-delà du diagnostic : en se fondant sur des analyses de probabilités, ils peuvent proposer des choix d'actions possibles à l'utilisateur ou recommander des « décisions ». A cette classification, on peut ajouter les SIA génératives (type IV) comme Chat GPT ou Dall-E qui ont la particularité de pouvoir générer du texte, des images et/ou d'autres types de contenu en réponse à des requêtes humaines sur la base de données existantes. Ces systèmes, en particulier les grands modèles de langage comme GPT, peuvent être utilisés de différentes manières : chercher des informations, générer des programmes informatiques, générer/reprendre des textes, etc. Enfin, le cinquième type (type V) rassemble des systèmes qui permettent de prendre une décision ou d'agir en partie de manière autonome sans intervention humaine mais tout de même toujours dans un cadre prédéfini ou supervisé (par exemple, les voitures autonomes). Cette grande diversité d'usage et l'accès public gratuit à certains de ces systèmes constituent deux aspects importants qui favorisent l'utilisation de ces systèmes en contexte de travail.

2.3. Limites de l'IA et revers des promesses qu'elle suscite

Si les SIA actuels exhibent des performances remarquables, il existe des limites et points de vigilance à avoir concernant :

- leur fonctionnement (opacité, incertitude sur l'évolution dans le temps des systèmes, questions sur les données utilisées pour l'apprentissage)
- la face cachée des promesses de l'IA et les risques pour le travail et l'emploi (mirage de la collaboration humain-technologie, automatisation, management algorithmique)

2.3.1. Opacité, incertitude, qualité des données et biais

Le problème de l'**opacité** concerne les modèles d'IA dit "boîtes noires" (Rudin, 2019). Ceci renvoie à la difficulté à comprendre la manière dont un système ou modèle d'IA basé sur l'apprentissage produit un résultat spécifique. Ce problème se pose en particulier dans le cas des modèles basés sur l'apprentissage profond du fait de leur grande complexité, notamment en termes de nombre de paramètres manipulés. Plus le nombre de paramètres est large, plus grande est la complexité du modèle. Les spécialistes du domaine ne parviennent pas toujours à expliquer précisément ce qui se passe quand ce type de système « apprend » ou produit un résultat (Zouinar, 2020). Ce problème a généré tout un domaine de recherches dédié à l'**explicabilité** (XAI – *eXplainable AI*) parfois appelé interprétabilité ou transparence dont l'objectif est précisément de trouver les moyens permettant

d'expliquer un résultat donné ou le fonctionnement global d'un SIA. L'explicabilité est même considérée comme un principe éthique fondamentale de l'IA et elle apparaît dans la récente réglementation européenne de l'IA. Il convient toutefois de noter que ce problème de l'opacité/explicabilité n'est pas nouveau. Il s'est posé d'une part avec les « systèmes experts » mais de manière beaucoup moins forte en raison de la nature de ces systèmes qui reposaient sur des règles explicites – programmées - qui pouvaient être retracées, et d'autre part avec l'introduction de systèmes automatisés dans les environnements de travail (par exemple, dans l'aéronautique).

Un deuxième problème lié au premier est **l'incertitude** relative au fonctionnement des modèles d'IA basés sur l'apprentissage. Comme ils fonctionnent de manière statistique, il est difficile de s'assurer à l'avance que le modèle produira toujours les mêmes résultats. Il est donc également difficile de prévoir avec certitude comment le modèle évoluera dans le temps, en fonction de ce qu'il « apprend », c'est à dire des modèles qu'il génère à partir des données qu'il mobilise.

Vient ensuite **l'ambiguïté liée au concept d'apprentissage et les limites des capacités de l'IA**. Le terme « apprentissage » peut prêter à l'IA des promesses inatteignables à l'heure actuelle. En réalité, un algorithme dit "d'apprentissage" n'apprend pas au sens où l'humain peut apprendre. Il s'agit de calculs d'optimisation visant à réduire la marge d'erreur entre le résultat attendu et le résultat donné par le système. Qui plus est, derrière de nombreux SIA se cachent en réalité des humains qui effectuent le travail d'apprentissage (par exemple, l'étiquetage ou la structuration des données) ou réalisent le travail que l'IA est supposée faire (cf. travaux de Casilli). A ce jour, les SIA ne sont ni conscients, ni autonomes au sens où on l'entend pour un être vivant (capacité à définir ses propres buts, capacité à s'auto-organiser, à interpréter un contexte complexe, etc.). Les systèmes actuels relèvent majoritairement de l'IA dite « faible », c'est-à-dire spécialisée dans des tâches précises, à l'opposition de l'IA « forte » - terme introduit par le philosophe (Searle, 1984) - qui fait référence aux systèmes intelligents autonomes et généralistes qui seraient dotés d'un sens commun et d'une conscience (CNIL, 2017). Or, à l'heure actuelle, l'IA « forte » n'existe pas (Askenazy & Bach, 2019). Cela étant, la voie vers des systèmes généralistes, c'est-à-dire capable de réaliser une grande variété de tâches, semble avoir été ouverte par les grands modèles de langage comme Chat-GPT.

Enfin, il existe des **problèmes liés aux données utilisées pour l'apprentissage et à leurs biais, renvoyant à des questions éthiques et juridiques**. Il convient tout d'abord de comprendre que les données mobilisées par les SIA ne sont en fait pas « données » mais des construits sociaux et techniques qui peuvent être entachés de biais ou comporter des stéréotypes pouvant générer des discriminations sociales et conduire à une normalisation de la pensée (Crawford et al., 2023). Les données proviennent de bases construites par des humains et/ou des machines et élaborées par les data-scientists. Les questions concernant les données se posent ainsi à différents niveaux (Gamkrelidze, 2022) :

- l'accès libre et ouvert aux données (*open-data*). Par exemple, l'accès aux décisions de justice pour développer des logiciels d'IA pour le domaine juridique posent la question de la centralisation de ces données sensibles et des critères d'anonymisation et/ou de pseudonymisation. Dans le

domaine de la radiologie, l'accès aux données médicales (par exemple, des images radiologiques et des compte-rendus radiologiques correspondant permettant d'avoir des éléments du diagnostic, qui servent à entraîner des algorithmes pour aboutir à des modèles d'apprentissage) s'effectue dans le cadre de négociations et de partenariats entre les concepteurs de SIA et certaines structures médicales.

- la préparation des données dont les caractéristiques, comme la quantité, l'exhaustivité et la qualité, sont des éléments majeurs dans la conception des SIA. La préparation des données suppose de les nettoyer et de les enrichir (par exemple, de métadonnées) et de les faire valider par des équipes spécialistes des données et/ou du domaine concerné. Par exemple, dans le domaine de la radiologie, des radiologues peuvent être recrutés pour un travail d'annotation qui consiste à indiquer des éléments pertinents sur des images radiologiques, par exemple par des contours sur des images, par une spécification de la pathologie, etc.
- l'exploitation des données recueillies, nettoyées et enrichies, pour un entraînement des algorithmes pour d'une part obtenir des modèles d'apprentissage stabilisés et performants et d'autre part, de les évaluer et les améliorer. Le système n'apprend pas et ne s'améliore pas de manière automatique, sans demande ni intervention humaine.
- la reproduction de biais pouvant générer des discriminations, renvoyant à des questions d'éthique. Bien que certains soutiennent que l'IA pourrait aider à éliminer les préjugés humains grâce à une supposé neutralité des algorithmes, les SIA peuvent également refléter des biais qui peuvent se trouver dans les données. De plus, leur manque de transparence rend difficile la détection et la correction de ces biais. Par conséquent, des décisions importantes telles que l'embauche, la promotion ou le licenciement peuvent être influencées par des critères opaques et potentiellement discriminatoires. Amazon avait ainsi retiré un algorithme de traitement de CV qui avait tendance à favoriser les hommes.
- des questions éthiques et juridiques (Benbouzid & Cardon, 2018 ; CNIL, 2017 ; Villani, 2018) : la question de la déresponsabilisation ou de la dilution de responsabilité en cas d'erreur⁸ ; le cadre légal des données massives et personnelles, essentielles aux algorithmes d'IA ; la question de formes d'affects, de dépendance, de confiance et d'attachement, développées chez l'humain vis-à-vis des machines autonomes ou anthropomorphes.

2.3.2. *La face cachée des promesses de l'IA : vers plus d'automatisation et de subordination algorithmique ?*

Un premier problème quant aux promesses de l'IA renvoie à **l'augmentation potentielle de l'automatisation et la polarisation entre les tâches ayant de la « valeur » et celles « sans valeur ajoutée »**. Ceci peut ainsi accroître la hiérarchisation des rapports de travail qui donne plus de valeur

⁸Cette problématique de responsabilité se pose notamment avec des systèmes d'aide au diagnostic/d'aide à la décision. Par exemple, en cas d'erreur médicale, qui peut être tenu responsable : le praticien, le concepteur, le développeur ... ?

à certains métiers supposés plus qualifiés que d'autres qui continueront pourtant d'être réalisés par des humains là où il n'y aura pas d'IA. A cela s'ajoute la complexification des tâches et potentiellement une déqualification de certains travailleurs (Ferguson, 2019 ; Giblas et al., 2018). Il convient toutefois de noter que ce problème de la déqualification ou perte d'expertise n'est pas nouveau, il se pose depuis l'apparition de l'automatisation et même de la mécanisation. Selon Davenport & Kirby (2015), les « nouveaux » SIA marquent le début d'une ère d'automatisation de tâches « cognitives », ce qui implique que les métiers considérés comme plus ou très qualifiés seraient aussi de plus en plus concernés par l'automatisation. Dans la plupart des discours optimistes sur l'IA, on retrouve l'idée selon laquelle les tâches dites « routinières » pourraient être confiées aux SIA, tandis que les humains se concentreraient sur les tâches imprévisibles ou nécessitant des compétences sociales (p.ex. dimensions relationnelles au travail) (Bradshaw et al., 2018). D'autres scénarios envisagent que des tâches automatisables soient entièrement déléguées aux machines, sauf dans des situations nouvelles et inhabituelles où l'humain pourrait éventuellement remettre en question les « décisions prises » par la machine (McAfee & Brynjolfsson, 2017), sans réflexion sur le rôle de ces tâches routinières dans la structuration de l'expertise humaine.

Dans les cas où l'automatisation se limite à certaines tâches, l'IA et la manière dont le travail est réorganisé pourraient enfin accroître le contrôle des travailleurs par les machines, imposant leur rythme et prescrivant certaines tâches, en mettant à mal l'autonomie de prise de décision des travailleurs, comme cela a été montré par ailleurs pour d'autres technologies (p.ex. Bainbridge, 1983 ; Daniellou, 1986 ; Rabardel, 1995 ; Barcellini, 2022).

Il convient toutefois de souligner que certains voient des aspects positifs dans l'automatisation. Dans certaines entreprises, on estime ainsi que l'automatisation par l'IA pourrait réintroduire « l'aspect humain » dans le travail (Ferguson, 2019), par exemple en incitant les travailleurs à développer des capacités complémentaires à l'IA, celles que les machines ne possèdent pas encore, telles que les compétences sociales, relationnelles, créatives et de précision (Villani, 2018). Mais il reste à établir que ces aspects « positifs » sont bien atteignables, voire souhaitables, et avec quelles conséquences sur les travailleurs.

Un deuxième problème tient à un risque de subordination accrue en lien avec le management algorithmique. Dans le champ des promesses et des discours associés à l'IA, elle est parfois présentée comme un moyen pour repenser les organisations du travail vers plus de transparence, moins de hiérarchie et une coopération accrue (Giblas et al., 2018), en offrant aux travailleurs un accès aux données pour guider leurs décisions, favorisant ainsi un partage vertical et horizontal des informations. Cependant, elle est également – et surtout – mobilisée dans le cadre du management algorithmique, c'est à dire l'utilisation d'algorithmes et de données pour attribuer, optimiser, superviser et évaluer les tâches humaines dans un environnement de travail (Lee et al., 2015). Des procédures informatiques programmées, parfois fondées sur de l'IA, sont ainsi utilisées pour recruter, gérer les employés mais aussi organiser et coordonner le travail humain au sein d'une organisation (Baiooco et al., 2022; Ginioux-Kats et al., 2021). Ceci va cependant aussi vers une surveillance des travailleurs, dont les performances peuvent être évaluées par la machine, ainsi que leur comportement ou leur santé

physique et mentale (De Stefano & Taes, 2021; Kellogg et al., 2020). Le management algorithmique fait aussi appel aux usages des nouvelles technologies comme l'IA pour la prise des décisions concernant les travailleurs (Mollen & Hondrich, 2023). Cette évolution qui va vers un contrôle accru des travailleurs s'inscrirait - dans ce que Zuboff (2019) nomme le « capitalisme de surveillance » - qui réduirait l'autonomie des salariés et deviendrait ainsi source de troubles psychosociaux (Moore, 2019).

Enfin, un dernier problème et non des moindres renvoie au **mirage d'une supposée collaboration humains-technologie**. La collaboration humain-IA est devenue un important courant de pensée sur l'utilisation de l'IA au travail. Cette orientation est intéressante car elle permet de construire une vision non substitutive de l'IA, la vision substitutive étant celle qui consiste à remplacer l'humain par la machine. L'IA est souvent perçue comme un outil qui peut augmenter les capacités humaines et inversement : on parle d'augmentation mutuelle (Daugherty & Wilson, 2018; Jarrahi, 2018). L'humain augmenterait la machine en la « formant » avec des données massives structurées, en fournissant des explications sur ses actions et en surveillant son bon fonctionnement. En retour, la machine pourrait augmenter l'humain en amplifiant ses capacités cognitives, en étendant ses capacités d'interaction et en s'incorporant physiquement à lui. Dans cette perspective, les humains et les machines ne sont plus des adversaires, mais des partenaires symbiotiques (Jarrahi, 2018) qui améliorent mutuellement leur performance. Ces modèles d'augmentation s'inscrivent, en fait, dans une évolution des modèles d'automatisation, qui se concentraient sur la coopération humain-machine depuis les années 1980 (Hoc, 2000; Millot, 1999; Millot & Boy, 2012; Woods et al., 1990). Cependant, les termes de collaboration ou coopération paraissent excessifs, du moins pour le moment, car les capacités des systèmes sont loin de disposer de capacités cognitives et sociales nécessaires à la collaboration. Il importe donc d'être prudent sur les nombreuses promesses de collaboration ou de « symbiose » Humain-IA. De plus, les travaux anciens ont bien mis en évidence les limites d'une collaboration effective entre humain et machine : manque de capacités de communication, de compréhension des intentions et des buts poursuivis par les humains de la part des machines, difficulté à interpréter le contexte d'action pour une machine et à s'ajuster. Une des critiques majeures de ces approches est leur tendance à ignorer l'asymétrie entre les humains et les machines (Salembier & Pavard, 2004; Suchman, 1987). Par ailleurs, ces approches se concentrent souvent sur l'interaction entre un seul agent humain et un agent artificiel, négligeant les dimensions collectives et organisationnelles de l'activité humaine. Ces limitations persistent aujourd'hui, malgré les avancées en termes de performance des SIA (Van Den Bosch & Bronkhorst, 2018; Zouinar, 2020).

Nous avons dressé ici un premier point de vue nuancé sur l'IA, les promesses qu'elle suscite, les interrogations qu'elle soulève et ses limites. Nous allons poursuivre la construction de ce point de vue nuancé et critique par l'exploration des modèles économiques qui tentent d'expliquer les transformations de l'emploi liées à l'IA.

A Retenir...

- L'IA ne désigne pas une technologie ou un type particulier de système mais un ensemble hétérogène d'algorithmes, de systèmes informatiques, de machines (robots) et plus largement de technologies sous différentes formes (logiciels, robotiques etc.) qui visent à imiter des facultés cognitives humaines comme la perception visuelle ou auditive, la production et la compréhension du langage naturel, ou encore le raisonnement. Le terme « Intelligence » est utilisé de manière abusive alors même que les définitions sur ce qu'est l'intelligence humaine fait débats.
- Il existe deux approches en IA : l'approche symbolique et l'approche par apprentissage comme le connexionnisme, qui est devenue dominante aujourd'hui. Dans cette approche, ce qui est appelé « apprentissage » est la production de modèles construits à partir d'un grand nombre de données structurées (i.e. annotées par un humain) et basées sur des optimisations mathématiques.
- Les SIA actuels renvoient à 5 grandes fonctionnalités : recherches, reconnaissances et analyse d'informations ; prédiction et diagnostic, recommandation, exécution autonome d'action, génération de contenu nouveau ;
- Ces systèmes présentent plusieurs limites : opacité et manque d'explicabilité de leur résultat, incertitude sur l'évolution dans le temps des systèmes, nature et biais des données utilisées pour l'apprentissage.
- Au-delà des promesses d'augmentation de l'humain ou de l'automatisation de tâche routinière, il existe un risque accru d'automatisation et de subordination du travail et la réelle collaboration entre humains et IA reste un mirage.

3. Transformations de l'emploi liées à l'IA

Dans les approches économiques, la compréhension des effets du changement technologique sur l'emploi et les revenus a pris appui sur le concept de **biais technologique** qui continue à sous-tendre les travaux les plus récents consacrés à l'IA. Nous allons retracer brièvement la mobilisation de ce concept dans l'analyse des effets du progrès technique, ce qui nous aidera à développer un point de vue critique sur les travaux économiques actuels sur l'IA et leurs prédictions sur l'emploi. Ces approches ont évolué au fil des révolutions industrielles. Alors que l'on décrit dans l'entre-deux guerres le progrès technique comme biaisé en faveur du capital, dans les années 90, on va considérer que les ordinateurs sont biaisés selon la qualification (*skill-biased*) ou selon l'âge (*age-biased*). Puis, dans les années 2000, l'hypothèse dominante est que les TIC sont biaisées selon le caractère routinier des tâches (*routine-biased*). C'est ce modèle basé sur les tâches qui continue à prévaloir aujourd'hui pour penser les effets des IA et de la robotique sur l'emploi et les revenus. Il a cependant été étendu pour intégrer la façon dont ces technologies repoussent la limite de tâches automatisables. Même si elles se sont renouvelées, ces approches demeurent marquées par un déterminisme technologique qui ne laisse pas de place aux choix en matière d'organisation du travail. Nous examinons un modèle alternatif où l'accent est porté sur la production de connaissances nouvelles plutôt que sur l'automatisation des tâches. Tout en étant complémentaire aux IA dans ses effets sur l'innovation, l'investissement dans la capacité d'apprentissage des organisations y apparaît comme protecteur pour l'emploi.

3.1. Le progrès technique biaisé au fil des révolutions technologiques

La définition du *progrès technique biaisé* vient de l'analyse macroéconomique de la production plongée dans le cadre néoclassique d'analyse de la répartition de la valeur entre capital et travail qui s'est développée dans l'entre-deux guerre. Le modèle dit du *progrès technique biaisé* pense l'introduction d'une technologie comme un choc exogène sur la fonction de production agrégée⁹ qui la déplace vers le haut tout en étant susceptible de la déformer en modifiant la productivité relative des facteurs de production et, à l'équilibre, la distribution des revenus qui leur sont associés. Le caractère déterministe sous-jacent à ce concept s'est très largement maintenu jusqu'à ce jour dans la plupart des analyses portant sur les questions d'emploi. Ainsi, la technologie nouvelle apporte avec elle la direction de ses effets sur l'emploi et les revenus – elle les détermine, d'où le concept de techno-déterminisme.

Cependant, au fil des révolutions technologiques, la représentation des facteurs de production et le biais technologique considéré dans la littérature économique ont considérablement évolué (Bouabdallah et al., 1999). En outre, on peut noter que si les enjeux d'emploi reviennent régulièrement dans le débat public au moment où une accélération du développement technologique est ressentie, les technologies

⁹ En microéconomie, La fonction de production exprime la relation entre le volume de la production d'une entreprise et les quantités de facteurs de production utilisés. Il s'agit notamment du volume des services fournis par les travailleurs (facteur travail) et par les équipements (facteur capital) qui sont rémunérés par l'entreprise. La fonction de production agrégée est la transposition de cette relation au niveau d'une économie toute entière. Elle est utilisée pour étudier la croissance économique.

dont les effets sont analysés dans les travaux académiques ne sont pas celles qui émergent, mais plutôt celles pour lesquelles des données empiriques de qualité ont pu être collectées et analysées. Les représentations des enjeux d'emploi dans le débat public se construisent donc au croisement des grandes narrations futurologiques et des concepts élaborés par la recherche académique qui ont été testés sur des jeux de données portant sur les périodes antérieures.

3.1.1. *L'abandon de l'hypothèse d'un progrès technique biaisé en faveur du capital*

La première discussion explicite du biais technologique a été menée par Hicks (1932). Deux facteurs sont alors considérés, le travail et le capital. Plusieurs définitions du biais technologique sont ensuite proposées dans la littérature qui conduisent à des relations différentes entre la performance des facteurs de production et leur rémunération. Les travaux qui privilégient l'analyse de la performance vont évoquer un progrès technique qui « augmente » un facteur de production, ce qui signifie qu'il accroît sa productivité marginale¹⁰. Lorsque les analyses examinent plutôt les effets sur la répartition des revenus entre les facteurs de production, elles décrivent un progrès technique qui « économise » ou « consomme » un facteur par rapport à un autre. Pour Hicks, un progrès technique est biaisé s'il affecte la répartition. Les économistes de l'entre-deux-guerres pensaient plutôt le progrès technique comme biaisé en faveur du capital, et donc économisant le travail avec pour conséquence une réduction de sa part dans le partage de la valeur ajoutée.

Cette hypothèse est ensuite écartée en faveur de la neutralité du progrès technique au cours des trente glorieuses. Cette période correspond à l'*âge d'or* de la 4^e révolution technologique (Perez, 2010) appuyée sur le pétrole, l'automobile et la production de masse. La tendance à la hausse de l'intensité capitaliste¹¹, qui aurait étayé l'hypothèse d'un biais en faveur du capital, est difficile à établir empiriquement alors que le partage de la valeur ajoutée se déplace progressivement en faveur du travail. L'après-guerre est également marquée par le développement des systèmes éducatifs et les économistes s'étonnent du maintien de la croissance du salaire des travailleurs qualifiés relativement aux non-qualifiés en dépit de la croissance de l'offre de cette catégorie de main-d'œuvre. En 1969, Griliches formalise et teste l'hypothèse d'une complémentarité entre le capital et la main d'œuvre qualifiée. Cette complémentarité est différente d'un biais car elle passe par un effet-prix. Ainsi deux facteurs sont complémentaires lorsque la baisse du prix de l'un entraîne l'augmentation de l'usage de l'autre. En présence d'une telle complémentarité, un capital devenu moins onéreux du fait du progrès technique pourrait contribuer à soutenir la demande pour la main d'œuvre éduquée. Dans la pratique, les économistes ont rencontré des difficultés à distinguer empiriquement l'effet du biais technologique des relations de complémentarité ou substitution entre les facteurs résultant d'un mouvement sur les prix.

¹⁰ La productivité marginale d'un facteur est l'augmentation de la production liée à l'utilisation d'une unité supplémentaire d'un facteur, la quantité des autres facteurs de production restant constante

¹¹ Le rapport du volume de capital au volume de travail.

3.1.2. *L'ordinateur personnel biaisé selon la qualification et selon l'âge*

Avec l'entrée dans la 5^e révolution technologique, au début des années 90, marquée par le développement des Technologies de l'Information et de la Communication (TIC), la question du biais technologique se manifeste de nouveau. Le développement technologique connaît alors une première phase d'accélération qui rend l'ordinateur personnel accessible au plus grand nombre. Le point d'entrée des analyses économiques s'est cependant déplacé puisqu'il s'agit d'expliquer une division, au sein du groupe des travailleurs, entre ceux qui bénéficient de la technologie en termes d'emploi et de salaire et ceux qui au contraire sont perdants : les travailleurs non qualifiés et/ou les travailleurs âgés. On va parler d'un progrès technique biaisé selon la qualification (*skill-biased*) ou selon l'âge (*age-biased*). Si des bases de données plus riches, souvent au niveau sectoriel ou microéconomique sont mobilisées, c'est dans un cadre théorique plus flou. La mesure du stock de capital et de son coût d'usage nécessaire à l'évaluation d'un biais en faveur du capital est en général abandonné au profit de la prise en compte de l'hétérogénéité du facteur travail en fonction de la qualification ou de l'âge des travailleurs. Le progrès technique y est approché empiriquement soit de façon résiduelle, soit au moyen de mesures directes (Bouabdallah et al., 1999).

3.1.3. *Les TIC biaisées selon le caractère routinier des tâches et la polarisation des emplois*

La seconde accélération de cette 5^e révolution technologique est celle associée à Internet et au World Wide Web. Elle s'accompagne d'un renouvellement de la question du biais technologique dans les travaux économiques de la première décennie du millénaire. Il prend appui sur les résultats d'une étude de cas américaine qui analyse l'introduction d'une technologie d'imagerie numérique des chèques dans deux bureaux du back office de la banque Cabot, celui « d'en bas » et celui « d'en haut » (Autor et al., 2002). Les employés concernés par la nouvelle technologie effectuent des tâches cognitives qui sont pour une part de nature routinière, puisqu'elles peuvent être entièrement décrites selon une logique procédurale. Ce sont ces tâches qui ont été automatisées. Il s'agit de la saisie du montant inscrit sur le chèque lorsque l'écriture manuscrite est lisible dans le bureau « d'en bas » qui les traite et du transfert du chèque d'un poste à l'autre du bureau « d'en haut » qui gère les exceptions, par exemple les chèques émis sur des comptes qui ont été clôturés ou les chèques frauduleux.

A l'issue de cette étude de cas, les auteurs vont théoriser le progrès technique biaisé selon le caractère routinier des tâches (*routine-biased*) dans une étude empirique appuyée cette fois sur des données quantitatives (Autor et al., 2003). Ils y partitionnent l'espace des tâches en quatre grandes catégories selon qu'elles soient cognitives ou manuelles, routinières ou non routinières. Leur hypothèse est que les ordinateurs automatisent les tâches cognitives ou manuelles « routinières » en se substituant aux travailleurs et qu'ils assistent les travailleurs dans les tâches cognitives « non routinières » de résolution de problèmes et de communication complexes. Ils mobilisent ensuite une base de données américaine, le DOT (« Dictionary of Occupational Titles ») qui décrit et suit dans le temps les tâches exercées par les professions selon une nomenclature fine. Cette information, appariée à des données qui décrivent l'usage des ordinateurs dans le cadre du travail, leur permet de tester et de valider leur hypothèse pour la période 1960-1998. Goos et Manning (2007) vont ensuite observer sur données anglaises un

phénomène de polarisation des emplois : les emplois des professions situées au milieu de l'échelle des salaires ont vu leur part diminuer entre 1979 et 1999 au profit des emplois situés dans les professions les mieux et les moins bien rémunérées. Le biais selon le caractère routinier des tâches fournit une meilleure explication à ce phénomène que le biais selon la qualification car les professions situées au milieu de l'échelle des salaires exercent plus souvent des tâches routinières cognitives que les professions les moins bien payés dont les tâches sont principalement manuelles non routinières.

Ces travaux marquent donc le passage d'une analyse s'intéressant aux inégalités entre groupes professionnels en fonction de leur niveau d'éducation ou de qualification à une analyse où la vulnérabilité des emplois au progrès technique réside dans les caractéristiques des tâches qui y sont exercées. Les tâches (et donc la division du travail), que la théorie économique néoclassique avait reléguées au second plan au profit des facteurs de production, font ainsi leur retour au sein de la fonction de production. Le modèle théorique d'Acemoglu et Restrepo (2018), qui décrit une fonction de production à base de tâches devient une référence pour penser les enjeux emploi de l'automatisation : les facteurs de production produisent N tâches et la combinaison de ces tâches génère la production. Les tâches y sont hiérarchisées en fonction de la facilité avec laquelle elles peuvent être automatisées. Au-delà d'un certain rang I , le travail aura un avantage absolu et comparatif dans la réalisation des tâches. En outre, les méthodes empiriques développées par les travaux consacrés à la polarisation prennent la profession fine comme unité d'observation plutôt que l'emploi. En effet, comme le poids des différentes tâches mesurées à partir du DOT caractérise une nomenclature fine de professions, les analyses empiriques sont conduites sur des agrégats caractérisant des secteurs ou des professions.

3.1.4. L'IA et la robotique repoussent la limite des tâches automatisable

La troisième accélération est celle de l'Intelligence Artificielle (IA) et de la robotique. L'étude empirique qui a relancé le débat sur les enjeux d'emploi associés à cette technologie est celle de Frey et Osborne, rendue publique au début de la seconde décennie du millénaire et publiée en 2017. Elle a marqué les esprits car elle annonçait que 47 % des emplois américains référencés en 2010 avaient un risque élevé d'être automatisés grâce à l'IA et d'avoir disparu à l'horizon 2020-2030. Cette recherche étend l'analyse d'Autor et al. (2003) en posant l'hypothèse que l'IA repousse la limite des tâches automatisables par sa capacité à réaliser certaines tâches non routinières. Les développements de l'apprentissage machine et de la robotique mobile appuyés sur le traitement des données massives permettent en effet la reconnaissance automatique de formes et de structures. Plusieurs verrous en matière d'ingénierie bloqueraient néanmoins la généralisation de l'automatisation à l'ensemble de tâches. Ils concernent l'intelligence créative, l'intelligence sociale et les perceptions complexes associées à certaines manipulations (cf. supra et infra). Les auteurs appliquent ensuite un algorithme d'apprentissage à O*NET, la base de données en ligne qui a succédé au DOT à partir de 2010, pour prédire la probabilité d'automatisation de chacune des 702 professions fines recensées et décrites dans la base. Leur prédiction pessimiste s'appuie sur le choix d'un seuil de probabilité de 0,7, au-delà duquel les professions et emplois associés seraient voués à disparaître.

Alors que Frey et Osborne (2017) font l'hypothèse que l'IA remplace l'humain dans la réalisation de certaines tâches, Tolan et al. (2021) proposent une approche plus nuancée où l'IA développe des *aptitudes cognitives* qui peuvent être mobilisées dans la réalisation de certaines tâches. Ces aptitudes concernent les relations avec les humains, les connaissances ou les objets physiques. Le lien entre IA et tâches n'est donc pas direct, mais médié par les aptitudes cognitives de l'IA dont le développement dépend de l'effort de recherche. Ces auteurs enrichissent également le cadre de mesure de Frey et Osborne (2017) en combinant aux données de type O*NET¹² sur les tâches, des données produites par les chercheurs sur les performances de l'IA (*AI benchmarks*). C'est la mesure des aptitudes cognitives qui permet de faire le lien entre les deux : dans les données de type O*NET, on identifie quelles aptitudes cognitives sont supposées être mobilisées pour la réalisation des tâches tandis que les *AI benchmarks* évaluent les aptitudes des différentes IA testés par les chercheurs et donnent des indications de l'intensité de l'effort de recherche qui y est consacré à une date donnée. Sur la période 2008-2018, parmi les 14 aptitudes cognitives considérées, l'effort de recherche en IA aurait été plus important aussi bien en niveau qu'en évolution dans le domaine du traitement visuel. Mais cette aptitude cognitive n'occupe une place centrale que dans un tout petit nombre de professions, notamment celles associées à l'imagerie médicale. L'ensemble des travailleurs n'est donc que peu exposé aux conséquences du développement des recherches en IA sur le traitement visuel. Cette approche présente l'avantage d'évaluer le potentiel technologique de l'IA en tenant compte de l'orientation de la recherche. En s'en inspirant et s'appuyant sur le modèle d'Acemoglu et Restrepo (2018), Gries et Naudé (2022) proposent de modéliser la production comme le résultat d'une combinaison de tâches produites à partir d'un service humain qui mobilise les aptitudes et le temps de travail des humains conjointement aux aptitudes des IA et aux services fournis par les entreprises qui les développent.

3.2. Le déterminisme technologique au cœur des approches économiques des conséquences de l'IA sur l'emploi

Ces différentes approches s'appuient sur la modélisation de la fonction de production qui représente l'activité économique des entreprises. Parce que la demande de travail de l'entreprise dérive de la fonction de production, ses caractéristiques, qui sont données par la modélisation, vont fortement influencer la nature des résultats obtenus sur les enjeux emploi du progrès technique. Nous avons vu que par rapport aux travaux les plus anciens, les travaux les plus récents ont fait évoluer certains aspects de cette fonction. Dans un premier temps, le facteur capital disparaît et le facteur travail se fragmente en fonction du diplôme, de la qualification ou de l'âge, qui sont des caractéristiques associées aux travailleurs. Avec l'approche par les tâches, c'est la structure même de la fonction qui évolue puisqu'elle devient le résultat de l'emboîtement de deux fonctions : une première décrit la réalisation des tâches à partir des facteurs de production ; une seconde décrit comment la production résulte ensuite de la combinaison de ces tâches. Cette approche marque le retour de la division du travail dans

¹² Outre les données américaines d'O*NET, deux enquêtes auprès des travailleurs sont mobilisées : l'enquête européenne sur les conditions de travail (Eurofound) et les données du Programme pour l'évaluation internationale des compétences des adultes (PIAAC, OCDE).

la fonction de production tout en conservant la représentation canonique à partir des facteurs de production pour la réalisation de chaque tâche. On peut noter également qu'avec le modèle d'Acemoglu et Restrepo (2018), le capital réintègre la fonction de production, puisque les tâches sont réalisées soit à partir de travail uniquement lorsqu'elles ne sont pas automatisables, soit par une combinaison des deux, voire à partir de capital uniquement lorsqu'elles le sont. Enfin, les travaux de Tolan et al. (2021) suggèrent un modèle où les tâches sont réalisées en mobilisant des facteurs de production dont on distingue les aptitudes cognitives.

Avec la notion de biais technologique, le progrès technique arrive comme un choc exogène qui peut déformer la fonction de production et ainsi en changer les propriétés. La productivité relative des facteurs peut être modifiée suite à ce choc et on dira que le facteur de production - qui devient plus efficace techniquement - est « augmenté ». Ce choc peut également se transmettre à la répartition du revenu entre les facteurs en fonction des nouveaux équilibres qui se mettent en place sur les marchés, mais ce n'est pas systématique. On dira alors que le progrès technique est biaisé en faveur du facteur que le progrès technique consomme et qui voit sa part dans le revenu augmenter. Avec le modèle de la fonction de production à base de tâches, le choc associé au progrès technique est plus diffus. Il peut impacter le nombre de tâches N en le réduisant ou en générant de nouvelles tâches intensives en travail. Il peut également déplacer le seuil qui sépare les tâches automatisables des tâches qui ne le sont pas. Dans ce cas, le choc affecte le contenu en tâches de la production, ce qui va impacter l'allocation des tâches aux facteurs de production. Il peut enfin affecter la productivité relative des facteurs dans la réalisation des tâches, on retrouve là la forme classique du choc de productivité. Ce sont ces effets cumulés qui modifient la répartition des tâches entre le capital et le travail, ce qui à son tour est susceptible de modifier le prix relatif des facteurs de production avec des conséquences pour l'emploi et la part des salaires. Acemoglu et Restrepo (2018) envisagent également un progrès technique endogène, c'est-à-dire résultant de décisions prises par les acteurs économiques. Les éléments déterminants dans cette extension de leur approche sont la structure des droits de propriété intellectuelle, qui incite les entreprises à investir dans le progrès technique, et le coût du capital relatif à celui du travail, qui est susceptible d'orienter la Recherche et Développement (R&D) des entreprises vers l'automatisation et/ou vers la création de nouvelles tâches intenses en travail.

Dans l'approche néo-classique standard, qui fonde la notion de biais technologique selon la qualification (*skill biased*), l'organisation du travail ou celle de la production ne sont pas interrogées. La technologie impose directement une certaine combinaison de travail et de capital ou de travail qualifié et non qualifié. Dans l'approche néoclassique renouvelée, qui fonde la notion de biais technologique selon le caractère routinier des tâches (*routine biased*), la division du travail réintègre la représentation de la production au travers du contenu en tâches de la production, mais il n'y a pas à proprement parler de choix d'organisation du travail. Celle-ci dérive mécaniquement de la technologie. Et, dans le cas où des choix économiques orientent le progrès technique, ce sont les services de R&D qui façonnent l'articulation entre la technologie et l'organisation. Pour une entreprise donnée, les impacts de la technologie sur l'emploi, à niveau de production donnée, seront aussi purement mécaniques et le plus souvent défavorables.

Si les théories économiques classique et néo-classique finissent par conclure à l'indétermination de l'effet du progrès technique sur l'emploi, c'est du fait d'un ensemble de mécanismes de compensation qui vont intervenir en aval de la production, sur les marchés des produits, du travail et du capital (Calvino et Virgilito, 2018 ; Vivarelli, 2014) et qui conduisent à la création d'emplois nouveaux. Cependant, ces mécanismes ne sont pas automatiques et ils peuvent être entravés par un grand nombre de facteurs économiques, sociaux et institutionnels. Pour comprendre les effets du progrès technique sur l'emploi, il ne suffit donc pas d'examiner ce qui se passe au niveau des entreprises innovantes, il faut également suivre ce qui se passe au niveau sectoriel où ces entreprises sont en concurrence avec d'autres, au niveau des zones d'emploi où se manifestent les excès et les pénuries de main d'œuvre et s'interroger sur les enjeux de financement des investissements nouveaux. Si la question du lien entre changement technologique et emploi ne peut être tranchée théoriquement, les preuves empiriques apportées par les travaux de recherche doivent être suivies par les pouvoirs publics pour renforcer leur capacité d'anticipation. Hötte et al. (2023) font la synthèse de ces résultats publiés sur les quatre dernières décennies (entre 1988 et 2021) en distinguant cinq grandes catégories de technologies mesurées : les TIC, les robots, l'innovation, la productivité multifactorielle et une catégorie « autres ». Ils concluent que les mécanismes de compensation auraient bien joué leur rôle jusque-là dans la 5^e révolution technologique centrée sur les TIC. Mais, s'il n'y a pas d'effet net négatif en termes de quantité d'emplois, on observe néanmoins des inégalités entre les groupes de travailleurs : les travailleurs faiblement qualifiés, occupant des emplois industriels et réalisant des tâches routinières ont été plus exposés à la perte d'emploi ainsi qu'à la difficulté de trouver un nouvel emploi.

La représentation du lien entre progrès technique et emploi, intrinsèque à ces approches, conduit à des recommandations de politique publique qui sont principalement de nature corrective et qui interviennent en aval du développement et de la mise en œuvre des technologies émergentes. Lorsque l'on constate que les mécanismes de compensation demeurent limités, ou encore lorsque l'angoisse générale monte face aux récits des populations les plus vulnérables au progrès technique, les pouvoirs publics : (1) déploient alors - ou rénovent- leurs *filets de sécurité* ou (2) réfléchissent à la façon *d'équiper* les individus pour faire face aux risques accrus sur le marché du travail, cet équipement passant notamment par de la formation professionnelle.

3.3. Repositionner l'approche économique à la source de la technologie et de ses usages : le rôle protecteur pour l'emploi de la capacité d'apprentissage des organisations

L'étude de cas de Autor et al. (2002) dont découle le modèle de biais technologique selon le caractère routinier des tâches (Autor et al., 2003) inspire encore les recommandations les plus récentes des économistes américains concernant le déploiement des IA (Acemoglu et al., 2023), au travers de l'idée d'un progrès technique orienté vers l'automatisation. Cependant, Autor et al. (2003) n'ont pas intégré dans leur modèle tout ce qu'ils avaient observé sur le terrain à la banque Cabot, et notamment l'importante des marges de manœuvre organisationnelle dans la recomposition des postes de travail à

la suite de l'introduction de l'imagerie numérique des chèques. Cette avancée technologique, qui permet la photographie rapide du recto et du verso du chèque ainsi que la reconnaissance optique des caractères, romps les unités de lieu et de temps nécessaires au traitement du chèque physique et ouvre de nouvelles possibilités d'organisation du travail. Dans le département « d'en bas », qui traite les chèques, le choix organisationnel est de transformer un poste de travail aux tâches diversifiées en plusieurs postes de travail aux tâches spécialisées et moins bien payées. Dans le département « d'en haut » qui s'occupe du traitement des exceptions, une décision opposée est prise : plusieurs postes de travail spécialisés sont recombinaés dans un seul poste aux tâches plus étendues et plus en phase avec les besoins de la clientèle. Les auteurs nous expliquent que dans le département « d'en bas », les managers ont choisi une solution sur étagère : ils ont suivi les conseils du fournisseur d'équipement qui proposait de réorganiser le traitement des chèques selon un modèle de travail standard. C'est une tout autre démarche qui a été poursuivie dans le département « d'en haut ». Le responsable a cherché une solution permettant des gains de productivité, un meilleur service client et des emplois revalorisés. Il a aussi considéré qu'associer les travailleurs aux choix d'organisation permettait de mobiliser les savoirs acquis par l'expérience et d'obtenir une meilleure adhésion au nouveau système. A l'issue du processus participatif, l'accord s'est porté sur le fait que le travail ne devait plus être organisé par type d'exception mais par compte client, de sorte que la même personne s'occuperait de toutes les exceptions : demandes d'opposition, découverts, etc. liées à un compte donné.

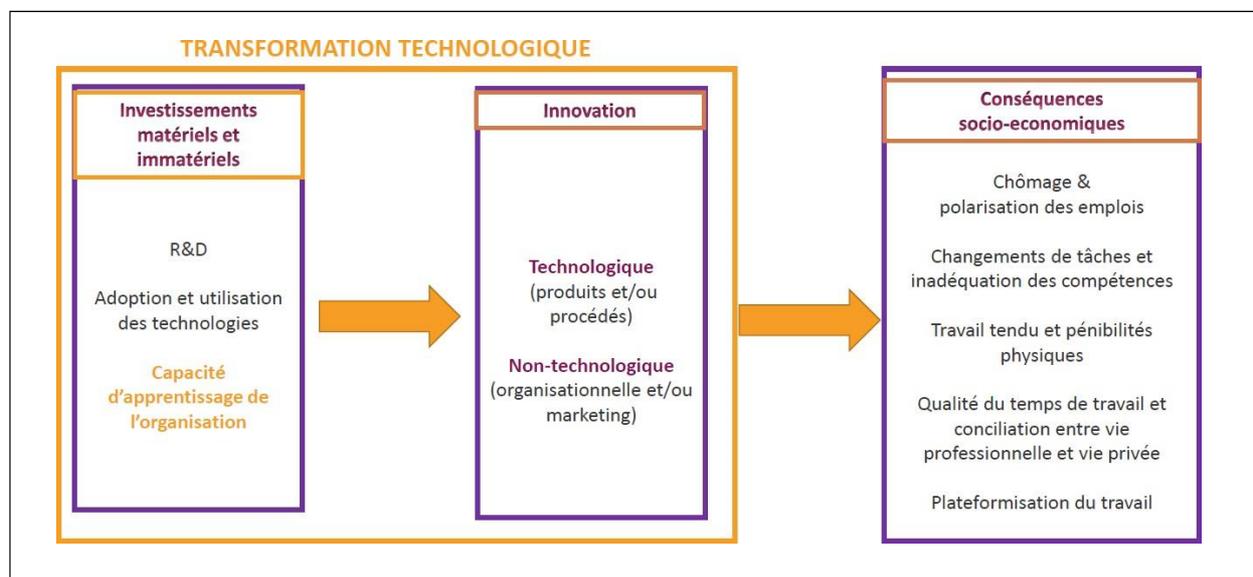
Des études de cas plus récentes illustrent cette marge de manœuvre organisationnelle centrale dans la détermination des usages des technologies et de leurs effets sur l'emploi. Si l'on prend une perspective européenne, on peut citer les monographies comparatives sur l'usage des technologies de l'industrie 4.0 au sein de Cesab-Toyota, Ducati et Lamborghini réalisées par Cirillo et al. (2021), l'analyse historique (1968-2016) de l'apport de la négociation collective à l'innovation chez Lamborghini (Russo et al., 2019) ou encore les études de cas comparative sur la négociation des usages du management algorithmique dans les centres d'appels d'entreprises de télécommunications comparables en Allemagne et en Norvège (Doellgast et al., 2023).

Ce constat appelle à abandonner l'hypothèse selon laquelle la technologie détermine les choix d'organisation afin d'avoir une vision plus complète de la dimension socio-organisationnelle du processus de transformation. Il s'agit de porter le regard plus en amont, tout d'abord, vers les projets de conception des IA (cf. infra) mais aussi vers ceux de leur déploiement en milieu de travail où leurs usages sont déterminés en lien avec des choix organisationnels et des stratégies économiques. Les questions d'emploi vont, en effet, en partie se jouer dans ces choix.

L'approche développée dans le cadre du projet européen [Beyond 4.0](#) s'inscrit dans cette démarche. Elle invite à renouveler la façon dont la transformation technologique est définie et modélisée. Il est en effet courant de considérer qu'il y a une transformation technologique dans une entreprise dès lors qu'elle introduit une technologie émergente dans son système de production. Or l'entreprise peut tout à fait sous-utiliser cette technologie et continuer à traiter ses affaires courantes et à produire « comme d'habitude ». Dans ce cas, la transformation technologique amorcée reste inachevée. La transformation advient lorsque la technologie émergente est mobilisée pour produire des connaissances nouvelles,

sources d'innovations. Et, la technologie, aussi bien conçue soit-elle, n'est pas le seul facteur qui intervient dans ce processus. De façon classique, les économistes de l'innovation considèrent également les dépenses de R&D. Dans les travaux du projet Beyond 4.0, Greenan et Napolitano (2023a) montrent le rôle d'un troisième facteur, la capacité d'apprentissage de l'organisation, identifié dans la littérature sur le changement technologique et l'innovation mais moins souvent défini et mesuré dans les approches empiriques quantitatives. La transformation technologique dépend ainsi des investissements conjoints des entreprises dans les technologies émergentes, la R&D et la capacité d'apprentissage de l'organisation. Et les types d'innovations qui en découlent jouent un rôle médiateur dans la détermination de leurs effets socio-économiques et notamment des risques susceptibles d'impacter les travailleurs. C'est ce qu'illustre la **Figure 2**.

Figure 2. Cadre théorique des analyses quantitatives du projet Beyond 4.0



Les auteures proposent également un cadre de mesure de la capacité d'apprentissage des organisations articulé autour de huit dimensions dont chacune représente une pratique organisationnelle ou une famille de pratiques. La capacité d'apprentissage de l'organisation se construit tout d'abord au niveau des situations de travail où il s'agit de préserver la dimension cognitive du travail (1), d'ouvrir des opportunités de formation (2) et de favoriser l'autonomie dans les tâches cognitives (3). Les travailleurs doivent ainsi être en mesure de régler eux-mêmes les problèmes, d'apprendre et de se former dans le cadre de leur travail. L'apprentissage prend aussi appui sur l'organisation du travail collectif. La mise en place d'équipes de travail autonome (4) favorise la capitalisation et le transfert de connaissances entre travailleurs. L'existence d'un soutien social dans les collectifs (5), apporté par les collègues, ou par la hiérarchie, est un autre élément qui nourrit les échanges et la confiance. Un style d'encadrement coopératif (6) qui autorise et accompagne l'expérimentation, fournit des retours sur expérience, contribue à la résolution des conflits, et favorise la prise de risque que représente l'exploration d'idées nouvelles. Le niveau organisationnel intervient enfin dans le développement de la capacité d'apprentissage, en permettant la participation directe des travailleurs grâce à des espaces de

régulation conjointe – par exemple des espaces de discussion sur le travail (7) et en soutenant leur motivation (8).

Ce cadre de mesure permet d'explorer empiriquement les relations illustrées par la **Figure 2** avec les données de la statistique publique européenne. On observe ainsi que tout en contribuant positivement à l'innovation, l'investissement dans la capacité d'apprentissage des organisations est associé à des effets socioéconomiques globalement favorables pour les travailleurs alors que l'investissement dans les technologies digitales seul et sans réflexion sur les usages socio-organisationnels de la technologie émergente n'a pas cette propriété (Greenan et Napolitano, 2023b ; Cursi et al., 2024). Ses effets socioéconomiques sont en effet la plupart du temps déterminés par le type d'innovation qui prend appui sur la digitalisation. Ainsi, les secteurs qui utilisent ces technologies pour innover en produits génèrent plus d'emplois de meilleure qualité alors que ceux qui les utilisent pour réaliser des innovations marketing¹³ connaissent un chômage accru et des conditions de travail et d'emploi plus défavorables. Ceci est probablement lié au fait que les innovations de nature purement marketing conduisent à la prédation des marchés existants plutôt qu'à leur expansion ou à la création de marchés nouveaux.

Il n'y a donc pas un scénario unique lié à l'introduction d'une technologie émergente, plusieurs scénarios sont possibles, certains seront vertueux pour les humains et l'environnement (redéploiement des compétences au sein des métiers, préservation de la santé des travailleurs, développement des capacités à préserver l'environnement), d'autres seront susceptibles de renforcer l'épuisement des ressources (fortes restructurations des métiers, pertes de savoirs et d'expertise, atteintes à la santé des travailleurs, moindre maîtrise des coûts environnementaux).

Il semble alors qu'un enjeu important des recherches et des politiques publiques est de penser comment agir en amont du processus de développement de la technologie et de définition de ses usages pour favoriser un scénario vertueux plutôt que de n'intervenir qu'en aval pour corriger des problèmes lorsqu'ils se manifestent.

L'analyse critique proposée ici appelle le besoin de considérer de manière empirique les effets de l'IA sur le travail et l'emploi, notamment (1) les liens entre projets d'introduction d'IA, et transformations conjointes du travail et de l'emploi et leur cadrage par des processus de régulations ; et (2) les transformations du travail concrètes intégrant un vision enrichie de la valeur réelle du travail et non de tâches, faisant une place importante à la performance et la santé des travailleurs et la conduite des projets, comme éléments clé pour favoriser la capacité d'apprentissage des organisations.

¹³ Dans l'enquête innovation européenne, l'innovation marketing est définie comme la mise en œuvre de concepts nouveaux ou de stratégies de vente nouvelles qui diffèrent significativement des méthodes de vente existant dans l'entreprise et qui n'avaient pas été utilisés auparavant. Elle inclut les changements significatifs dans le design, l'emballage ou la présentation d'un bien ou d'une prestation de services, son positionnement, sa promotion ou son prix.

A Retenir...

- Les représentations des enjeux d'emploi dans le débat public se construisent au croisement des grandes narrations futurologiques et des concepts élaborés par la recherche académique qui ont été testés sur des jeux de données portant sur les périodes antérieures.
- L'approche économique néoclassique traitant des effets sur l'emploi des technologies émergentes est techno-déterministe. Elle voit l'introduction des technologies comme « un choc exogène » qui apporte avec elle la direction de ses effets sur l'emploi et les revenus.
- L'introduction de technologie émergente au travail – le progrès technologique – y est vue comme source de biais qui transforme le partage de la valeur ajoutée entre les facteurs de production. Selon les périodes et les technologies, ces biais favorisent le capital, le travail qualifié, la main d'œuvre moins âgée, les tâches non routinières ou les tâches non automatisables.
- Pour une approche alternative, il s'agit de considérer les choix possibles en matière d'organisation du travail.
- Plutôt que d'adopter une approche de la production centrée sur les tâches nous examinons une approche de la production centrée sur la production de connaissances nouvelles. Les IA, en association avec la R&D et la capacité d'apprentissage des organisations contribuent à cette production des connaissances nouvelles sources d'innovations pour l'entreprise. L'investissement dans la capacité d'apprentissage est un choix d'organisation aux effets protecteurs sur l'emploi dans un contexte d'adoption de technologies nouvelles.
- Pour comprendre les effets du progrès technique sur l'emploi, il faut également suivre ce qui se passe au niveau sectoriel où les entreprises sont en concurrence avec d'autres, au niveau des zones d'emploi où se manifestent les excès et les pénuries de main d'œuvre et s'interroger sur les enjeux de financement des investissements nouveaux.

4. Les effets sur le travail et sur l'emploi : quelles articulations ?

Les approches économiques dominantes tendent à analyser les effets sur l'emploi de l'introduction d'une technologie sans prendre en compte le contexte de cette introduction. Or, une « nouvelle » technologie, l'IA en l'occurrence, prend pied dans des organisations qui ont une histoire, des stratégies, des concurrents, un cadre institutionnel, des possibilités plus ou moins grandes pour les travailleurs et leurs représentants de prendre part aux discussions, le cas échéant, sur l'intégration et les usages de la technologie.

Nous avons cherché ici à mobiliser des publications qui permettent d'enrichir la compréhension du contexte des organisations qui utilisent ou introduisent des SIA et des conséquences sur le travail. D'abord en proposant 4 types de travaux qui permettent d'articuler effets sur l'emploi et effets sur le travail. Puis en rendant compte des travaux sur les modalités de régulation de l'IA, et les liens entre ces modalités de régulation et les effets sur le travail.

4.1. Analyser à la fois les effets sur le travail et les effets sur l'emploi

Trop peu de travaux analysent conjointement ces deux séries d'effets et leurs interrelations, en particulier au niveau des lieux de travail. Nous retenons ici quatre approches différentes des effets à la fois sur le travail et sur l'emploi : monographie, étude de cas, analyse prospective, analyse sectorielle/systemique.

4.1.1. Un exemple d'analyse monographique

A partir d'une enquête réalisée au sein d'un groupe industriel employant plus de 100 000 salariés, Ferguson (2019) examine les conséquences et les enjeux de plusieurs actions ou expérimentations menées par ce groupe au cours des deux années précédentes, sur la base d'observations participantes à des ateliers internes. Ces ateliers d'échanges, de réflexion, de sensibilisation à l'IA ont été conçus par un sociologue spécialisé dans l'éthique des sciences et des techniques et organisés par le Pôle Métiers support (16 domaines et 12 000 personnes). Ferguson détaille ainsi ce que font les programmes d'intelligence artificielle et leurs conséquences sur l'emploi et sur le travail. Le

Tableau propose une synthèse de ses analyses outil par outil en précisant leurs fonctionnalités de l'IA, les travailleurs concernés, les tâches concernées et les enjeux pour l'entreprise de l'usage d'un SIA, et enfin les conséquences sur l'emploi et le travail.

On constate que plusieurs des expérimentations concernent des outils d'assistance virtuelle censés répondre à des sollicitations - pensées comme simples - de clients ou d'employés. Une seule concerne le remplacement d'expert.

Tableau 1. Synthèse à partir de Ferguson (2019). *En italique commentaires des auteurs du rapport.*
Noms des programmes anonymisés par Ferguson.

Nom du SIA	Fonction	Travailleurs concernés, tâches prises en charge Enjeux	Conséquences sur l'emploi	Conséquences sur le travail
TCOA	assistant virtuel ou « assistant personnel intelligent »	Plateau de sous-traitants (27 personnes) : gestion des relations avec les fournisseurs → demandes les plus simples	en interne : suppression ½ ETP ; création ½ ETP « entraîneur IA » prestataire : suppression 3 ETP	Non décrites
Bot Comptabilité	assistance virtuelle	Réponse aux questions les plus récurrentes des employés de l'entreprise solution au non-remplacement d'un salarié	Suppressions déjà décidées : 7 ETP création 1 ETP « entraîneur IA »	Soulage les agents restants de l'équivalent de 2 ou 3 h par jour
WeGO	assistance virtuelle	Réduction des effectifs de l'agence de voyage interne solution au non-remplacement de deux salariés partis	création ½ ETP « entraîneur IA »	Salarié seul sur son poste
BE 2.0	appui à l'expertise	Départs en retraite d'ingénieurs avec savoir sur des équipements anciens toujours en fonctionnement (pas d'investissement en formation pour les nouveaux)	Fin des interventions des nouveaux retraités	Avant fiches rédigées par les futurs retraités (centaines). L'assistant fait le tri parmi les informations, donne uniquement celles dont les jeunes ingénieurs ont besoin <i>Pas de transmission des connaissances sur les équipements anciens</i>
RightBot	assistance virtuelle	Réponse à des sollicitations simples Traiter environ 10% des questions posées	Pas de réduction des effectifs du service juridique	Décharge les salariés du service de 5 à 10% de leur temps de travail
ComCom	assistance virtuelle	Conseillers commerciaux dans leurs conversations avec les clients Saisie en direct de la conversation → alerte, propose à partir d'un base de connaissance sur les produits et sur les clients		<i>Surcharge cognitive ?</i>
ANSWER	assistance virtuelle	service de comptabilité → routage de mails, réponse automatique à des mails simples, scénarios de réponse pour des demandes plus complexes	7 ETP : pas de réduction des effectifs a priori	Réduction importante de la charge de travail de réponse aux mails : de 6-8h par jour à 2h par jour Activité recentrée sur ce qui demande du savoir-faire
Pepper	robot social	accueil, information et moments ludiques reconnaît les émotions humaines		Déshumanisation, appauvrissement des relations sociales
Julie Desk	assistant virtuel	Planification et annulation de rendez-vous via la messagerie		Déshumanisation, appauvrissement des relations sociales

Ces expérimentations révèlent, en creux, les motifs à la genèse de l'introduction des systèmes d'IA. Les quatre premiers exemples illustrent parfaitement une « logique substitutive », qui va à l'encontre des promesses de collaboration IA-humains discutées plus haut. L'IA prend place dans l'entreprise avec chaque départ (mobilité, mutation, retraite). Dans trois de ces exemples, l'évolution des emplois n'est pas directement liée à l'introduction de programmes d'IA, les décisions étant largement déjà prises. L'IA peut toutefois contribuer à « faire accepter » ces réductions d'effectifs dans les services en question. En l'absence de licenciements au sein de l'entreprise, les conséquences directes sur l'emploi (chez le prestataire pour le TCOA ou sur l'emploi de retraités) ou indirectes (embauche de jeunes ingénieurs) sont peu visibles. Les deux derniers exemples illustrent pour Ferguson une stratégie d'« augmentation » des capacités des travailleurs, en adéquation avec les promesses de l'IA. Cette stratégie n'implique pas *a priori* une réduction des effectifs.

Les conséquences sur le travail sont largement implicites. Ainsi, l'absence de transmission des savoirs professionnels aux jeunes ingénieurs sur les équipements anciens relève d'une logique de « domination-rationalisation » : ces savoirs sont délégués à l'appui à l'expertise. Les risques de dépendance à l'assistance virtuelle et de moindre capacité à répondre à une situation non répertoriée ne sont pas envisagés. Dans le cas de ComCom, les alertes et propositions faites aux salariés en cours de conversation accroissent la « charge cognitive ». Les deux derniers exemples suscitent des commentaires au cours des ateliers sur un « appauvrissement des relations sociales », ce qui pourrait renvoyer à un appauvrissement de la dimension collective du travail dont on sait qu'elle est un facteur d'efficacité et de santé (Caroly et Barcellini, 2013).

4.1.2. Un exemple d'études de cas

L'approche par études de cas permet d'analyser une grande variété de facteurs, de situer le contexte, et de considérer les effets au plus proche de la réalité du terrain.

Une enquête qualitative réalisée en 2021-2022 a permis d'analyser 96 cas d'application des IA dans deux secteurs (industrie manufacturière et finance) et dans huit pays de l'OCDE (Allemagne, Autriche, Canada, Etats-Unis, France, Irlande, Japon, Royaume-Uni). Après une interrogation large sur les différents cas d'usage dans une entreprise, un seul cas d'usage a été retenu : celui qui a le plus d'impact sur le marché du travail. Les questions sur ce cas d'usage ont ensuite été posées à six types d'interlocuteurs : salariés, représentants des salariés, représentants de la direction, responsables des ressources humaines, responsables de la mise en place d'IA et développeurs. La méthode par entretien a aussi permis d'adopter une vision large des IA introduites en entreprise, en partant des exemples indiqués par les interlocuteurs.

Les résultats (Milanez, 2023) mettent en évidence des effets nuancés sur le travail et sur l'emploi. Les cas d'usage d'IA étudiés concernent une large partie des travailleurs, aussi bien des travailleurs très qualifiés que peu qualifiés, aussi bien des travailleurs très « exposés » que « peu exposés » à ces technologies.

Dans trois quarts des cas, l'introduction d'une IA ne réduit pas le volume de travail des travailleurs concernés. Dans la moitié de ces situations, les travailleurs sont déplacés vers d'autres emplois, ou vers de nouveaux emplois. Dans un quart des cas, il y a réduction du volume d'emploi mais l'ajustement passe par d'autres voies les licenciements (réaffectations internes, départs naturels ou volontaires). Dans un cas sur cinq, le rythme des embauches est ralenti ou le sera.

Des changements fréquents interviennent dans la composition des tâches réalisées avec l'IA. Ces changements prennent le plus fréquemment la forme d'une automatisation partielle (l'IA accomplit des tâches simples, le travailleur prend en charge les tâches complexes) ou totale des tâches (déplacement des tâches de travail vers l'IA). Très fréquemment aussi mais un peu moins, il s'agit de changement partiel des tâches (le travailleur reste responsable de l'ensemble de la tâche, l'IA augmentant sa capacité à la réaliser). C'est le cas de figure le plus fréquent dans l'industrie manufacturière. Enfin la création de nouvelles tâches concerne essentiellement des tâches qui nécessitent des compétences en IA.

Dans six cas sur dix, les compétences requises restent les mêmes soit parce que les IA utilisées ont peu d'impact sur les tâches réalisées par les travailleurs, soit parce que les réorganisations les ont réaffectés sur des tâches déjà existantes. Les compétences nécessaires pour travailler avec l'IA apparaissent marginales. Dans quatre cas sur dix en revanche, de nouvelles compétences ou des compétences mieux assurées sont nécessaires. Dans quelques cas les compétences des travailleurs ne sont plus utiles.

Enfin, les travailleurs soulignent des avantages à la mise en place d'IA : réduction de l'ennui avec la diminution des tâches répétitives et peu intéressantes ; réduction de la fatigue et de la pression temporelle lorsque l'IA réduit la surcharge ; réduction de la pénibilité physique lorsque l'IA prend les tâches correspondantes en charge. L'usage des IA conduit aussi dans certains cas à créer des tâches ennuyeuses, à accroître l'intensité et la pression du travail. Nous y reviendrons plus précisément dans la section 5.

4.1.3. Un exemple d'analyse prospective

L'analyse prospective consiste à élaborer des scénarios alternatifs plausibles en combinant différents types de facteurs déterminants. L'analyse prospective réalisée par l'Inrs (2022) est l'une des plus récentes. Elle vise à anticiper les conséquences en termes de santé et sécurité au travail de différents scénarios de développement et d'usage des IA.

Quatre scénarios sont construits à l'issue de l'exercice de prospective¹⁴. Le **Tableau 2** résume les éléments essentiels de ces scénarios. Compte tenu de l'objectif de cet exercice pour l'Inrs, très peu

¹⁴ Voir aussi Inrs (2022), L'intelligence artificielle au service de la santé et de la sécurité au travail, enjeux et perspectives à l'horizon 2035, 18 novembre 2022, <https://www.inrs.fr/footer/actes-evenements/prospective-IA.html>

d'éléments sont donnés sur l'évolution de l'emploi (sauf dans le scénario 3 « Développement démocratique »).

Tableau 2. Synthèse des scénarios prospectifs, Inrs (2022)

	1 Solutions et vision imposées par les géants du numérique	2 Un cadre garanti par les Etats pour l'intégration de l'IA	3 Développement démocratique	4 Déclin de l'IA
Dynamique	Poursuite exubérance technologique Concurrence	Montée des préoccupations environnementales Multiplication des problèmes posés par les usages de l'IA	Croissance économique mondiale, générant de l'emploi et permettant d'investir dans la formation	Développement des usages de l'IA puis déception par rapport aux promesses → déclin/rejet de cette technologie et de ses usages professionnels
Principaux acteurs	Géants mondiaux du numérique	Etats	Les travailleurs (et leurs représentants ?) et les citoyens	Géants mondiaux du numérique
Réglementation /régulation	Morcelée Innovation contrôlée par les géants du numérique	Encadrement réglementaire Décision : développer des systèmes IA sobres, sur des secteurs non critiques et sous supervision humaine	Développement des processus de contrôle démocratique de l'IA	Absence de réglementation harmonisée, Inégalités statutaires entre secteurs + divergences en matière d'éthique entre les grandes zones économiques → augmentation des contentieux en matière du droit des salariés
Comportements des utilisateurs /usages	Normes devenues indispensables	Meilleur contrôle des fabricants Développement de pôles d'expertises Expérimentations visant à démontrer l'innocuité de l'IA	Principe éthique d'explicabilité instauré comme clef d'appropriation Essor de systèmes d'IA hybrides combinant l'apprentissage automatique à des systèmes de raisonnement logique	Diffusion des usages dans de nombreux domaines Déception face aux applications de terrain, failles des systèmes qui provoquent des incidents, accidents ou crises
Travail	L'automatisation progresse et la surveillance s'installe comme outil privilégié de la sécurité au travail.	Usage encadré de l'IA au travail, sous condition de la non dangerosité des dispositifs	Confiance collective dans l'IA mise au service de la santé et de la sécurité au travail	Outils IA intrusifs Traçage systématique et sans leur consentement de l'activité des employés Rejet des systèmes d'IA dans le monde du travail

L'intérêt de ces scénarios est de d'articuler les comportements stratégiques des principaux acteurs (avec une domination ou non par des grandes entreprises qui fixeraient les normes techniques), les comportements des travailleurs dont les situations de travail se retrouvent « IAtisées », le rôle d'une régulation et le développement ou non du secteur. L'usage des SIA au travail est ainsi étroitement lié

aux autres éléments de ces scénarios. La régulation garantie par les États (scénario 2) qui conduit à sélectionner les SIA selon leur coût environnemental, le degré de risque qu'ils impliquent et sous condition de supervision humaine, contraste avec une régulation à l'initiative des producteurs (scénarios 1 et 4) qui tendrait à promouvoir des normes différenciées et peu contraignantes. Le scénario 4 (déclin de l'IA) souligne que des outils très intrusifs, ou des outils dont les performances seraient décevantes pourraient produire un rejet des IA.

L'une des fiches qui précise les éléments intégrés dans l'exercice de prospective mentionne la négociation collective entre employeurs et salariés. Cette fiche « Acceptabilité de l'IA dans le monde du travail pointe des écarts importants entre les représentations des dirigeants et encadrants, et celles des salariés, et semble aborder les questions de négociation collective sous un angle d'adaptation des employés/salariés à l'IA. Comme nous le verrons plus loin, les relations entre employeurs et travailleurs constituent également un mode de régulation.

4.1.4. *Des exemples récents d'effets sectoriels voire systémiques sur le travail et sur l'emploi*

Compte tenu de la durée du processus d'évaluation et des délais de publication, les articles scientifiques dans les domaines retenus portent encore peu sur des exemples récents d'effets sur le travail et l'emploi liés à d'autres évolutions des modèles d'activité économique dans certains secteurs. Un pas de côté est donc nécessaire pour trouver ces exemples, soit dans des revues spécialisées, soit dans des articles de presse/média. Deux exemples sont apparus emblématiques ces derniers mois car les professions concernées se sont mobilisées pour attirer l'attention sur les effets de l'usage des technologies IA.

Le premier exemple a été largement discuté à propos de la grève des scénaristes aux États-Unis¹⁵ : il s'agit de la production de scénarios par des outils d'IA générative et de leurs conséquences sur l'emploi, la rémunération et le travail des scénaristes. Ceux-ci s'inquiètent de l'usage de l'IA générative pour la production de scénarios, qui les priverait de droits d'auteurs (si des scénarios générés par l'IA produisent des droits d'auteurs), capterait leurs productions (si les IA sont entraînées avec leurs productions pour concevoir de nouveaux films ou séries, sans leur accord, sans les rétribuer, ni même les créditer) et les priverait d'emploi (si les IA deviennent la source privilégiée voire unique de production de scénarios) et de revenus (si les premières versions des scénarios sont générées par des IA alors que ce sont les mieux rémunérées, les scénaristes intervenant seulement dans la réécriture, beaucoup moins rémunérée¹⁶). L'échec des négociations entre la Writers Guild of America (WGA), le syndicat qui regroupe près de 11 500 scénaristes, et l'Alliance of Motion Picture and Television Producers (AMPTP), organisation qui représente les grands studios hollywoodiens, les chaînes de télévision et les

¹⁵ Voir A. Defer (2023), « À Hollywood, la fin de la grève des scénaristes par une victoire en demi-teinte sur l'IA », *L'Usine digitale*, 27 septembre, <https://www.usine-digitale.fr/article/les-scenaristes-hollywoodiens-demandent-a-etre-protéges-contre-l-ia.N2130151> ; M. Poitiers (2023), « Les scénaristes hollywoodiens demandent des garanties vis-à-vis de l'IA générative », *L'Usine digitale*, 6 mai, <https://www.usine-digitale.fr/article/les-scenaristes-hollywoodiens-demandent-a-etre-protéges-contre-l-ia.N2130151> ; C. Lesnes (2023), « A Hollywood, la satisfaction des scénaristes après 148 jours de grève », *Le Monde*, 2 octobre, https://www.lemonde.fr/culture/article/2023/10/02/a-hollywood-la-satisfaction-des-scenaristes-apres-148-jours-de-greve_6191968_3246.html.

¹⁶ Comme le souligne M. Poitiers (2023), ces méthodes sont déjà largement employées par les agences de traduction depuis quelques années.

plateformes de streaming, a conduit à une grève de cinq mois, la 2^{ème} plus longue grève de l'histoire de la WGA (du 2 mai au 27 septembre 2023). Cette négociation s'inscrit dans un contexte de forte concurrence par les plateformes dites de streaming. Sous contrainte de rentabilité, les studios souhaitent revenir sur le système de contrats actuels pour passer à un système de paiement à l'heure de travail qui précariserait la profession¹⁷. Les scénaristes demandaient des progressions de leurs rémunérations. Le streaming a démultiplié la durée de vie des œuvres mais les salaires stagnent depuis dix ans. Autre point de revendication : le maintien du processus collectif de création dans les « writers' rooms », que les producteurs veulent réduire.

Le nouvel accord conclu pour trois ans inclut des dispositions qui entérinent mais limitent l'usage de l'IA¹⁸ : 1) « L'IA ne peut pas écrire ou réécrire du matériel littéraire, et le matériel généré par l'IA ne sera pas considéré comme du matériel source en vertu du MBA, ce qui signifie que le matériel généré par l'IA ne peut pas être utilisé pour miner le crédit d'un écrivain ou ses droits séparés » ; 2) « Un scénariste peut choisir d'utiliser l'IA dans le cadre de ses services de rédaction, si l'entreprise y consent et à condition que le scénariste respecte les politiques applicables de la société, mais l'entreprise ne peut pas exiger du scénariste qu'il utilise un logiciel d'IA (par exemple, ChatGPT) dans le cadre de ses services de rédaction » ; 3) « L'entreprise doit informer le scénariste si les documents qui lui sont remis ont été générés par l'IA ou intègrent des éléments générés par l'IA » ; 4) « La WGA se réserve le droit d'affirmer que l'exploitation du matériel des scénaristes pour former l'IA est interdite par le MBA ou toute autre loi ». Par ailleurs, le syndicat a obtenu satisfaction sur d'autres revendications : l'augmentation du salaire de base ; la rémunération des droits « résiduels » (les droits d'auteur sur la rediffusion des créations) ; l'obligation de respecter un nombre minimal de scénaristes dans les « writers' rooms ».

Cet accord est d'autant plus important qu'il est le premier signé dans le secteur cinématographique alors qu'allait s'enclencher une série de négociations pour les trois ans à venir pour d'autres professions du secteur. La grève des scénaristes a été suivie dès la mi-juillet 2023 par une grève du syndicat des acteurs états-uniens, puis par un appel à la grève du syndicat des doubleurs. Les cascadeurs sont également concernés¹⁹. L'IA générative offre en effet des possibilités de doublage et de sous-titrage des films sans recours à des acteurs et doubleurs, des possibilités de créer des « jumeaux » ou « répliques numériques » ou de nouveaux personnages sans traçage des apports des acteurs, et sans rémunération supplémentaire. Des clauses abusives dans les contrats d'acteurs autorisent les studios à créer des versions synthétiques de leurs voix qu'ils peuvent réutiliser sans rémunération. La captation des mouvements et des caractéristiques biométriques est une pratique courante dans le domaine des jeux vidéo. Là aussi, les contrats d'embauche laissent le plus souvent la possibilité de réutiliser ultérieurement ces captations sans rémunération.

¹⁷ Idem M. Poitiers (2023).

¹⁸ Un résumé est accessible via <https://www.wgacontract2023.org/the-campaign/summary-of-the-2023-wga-mba>. Ce qui suit a été traduit par DeepL puis corrigé.

¹⁹ France Info (2023), « Grève à Hollywood : avec la montée en puissance de l'intelligence artificielle, les cascadeurs voient leur métier menacé », FranceTVInfo, 12 août, https://www.francetvinfo.fr/culture/cinema/greve-a-hollywood-avec-la-montee-en-puissance-de-l-intelligence-artificielle-les-cascadeurs-voient-leur-metier-menace_6003053.html

La grève des scénaristes laisse penser que la pression à l'utilisation de l'IA générative provient des producteurs. Des exemples de formations proposées en France par des scénaristes ou par des établissements d'enseignement sur l'usage de logiciels d'IA générative mettent en avant l'intérêt pour un scénariste d'utiliser ces outils en tant qu'assistants pour raccourcir le temps de production d'un scénario ou de sa description, donc en particulier pour améliorer la rémunération nette de ces productions.

Le second exemple est celui de la production de contenu par des outils d'IA générative et de leurs conséquences sur l'emploi, la rémunération et le travail des journalistes et plus largement des professions impliquées dans l'édition d'articles pour la presse et les médias. Deux événements récents ont contribué à rendre visibles ces effets : l'action en justice du New York Times aux Etats-Unis ; la grève engagée par les salariés d'Onclusive en France.

Le New York Times a engagé le 27 décembre 2023 une action en justice contre Microsoft et OpenAI pour violation des droits d'auteur²⁰. Le journal met en cause l'utilisation de millions d'articles qu'il a publiés pour l'entraînement d'Open AI, sans son consentement et sans compensation. Les contenus créés par Open AI et les autres outils d'IA générative de textes et photos, les chatbots créés et entraînés captent des flux d'investissements énormes en rentabilisant des contenus qu'ils utilisent gratuitement, sans avoir soutenu le coût des investissements et du fonctionnement d'une équipe de rédaction. Ils constituent une concurrence déloyale vis-à-vis des médias réellement producteurs de contenus informatifs. En détournant les flux de consultation vers des sites ou des chatbots livrant une information non produite et vérifiée par des journalistes, ces pratiques réduisent la rémunération des médias et ont conduit à la disparition de nombreux journaux et médias dont le modèle d'activité était plus fragile que celui du New York Times. Ce comportement est ce que l'on désigne en économie comme un comportement de « passager clandestin ». Des actions en justice sur les mêmes fondements ont été intentées par des auteurs (parmi lesquels John Grisham) et par l'agence de photographie et banque d'images Getty Images. Le New York Times s'inquiète aussi de la moindre qualité des contenus fournis par ces outils. Il ne s'agit donc pas seulement de concurrence déloyale mais du risque de dégradation de la qualité de l'information et de la disparition de la plupart des acteurs professionnels du secteur. S'y ajoute le risque qu'une mauvaise information puisse être attribuée au New York Times, dégradant sa réputation auprès des lecteurs. D'autres producteurs d'informations (Associated Press, Axel Springer) ont déjà conclu des accords sur le respect des droits d'auteur. Le développement de tels accords répondrait à la question de la rémunération des droits d'auteurs mais pas à celle des conséquences sur la qualité des informations accessibles. Le New York Times envisage par ailleurs d'intégrer cette technologie dans son processus de production.

Onclusive, entreprise de veille média et de communication, a annoncé en septembre 2023 son intention de supprimer plus de la moitié de ses emplois (218 postes sur les 383 actuels) et de les remplacer par

²⁰ M. M. Grynbaum & R. Mac (2023), « The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work, The New York Times, 27 December, <https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html>.

l'usage d'IA générative. La mise en place d'un logiciel de suivi en temps réel de la télévision, de la radio, des actualités en ligne, des publications papier et des plateformes de réseaux sociaux a déjà permis aux effectifs de passer de 1 000 salariés début 2022 à moins de 400 salariés au moment de l'annonce. Outre les effets sur la structuration du secteur et sur la baisse de qualité des informations fournies²¹, les représentants des salariés pointent le basculement vers le micro-travail masqué de travailleurs à Madagascar.

L'usage de l'IA générative est déjà avéré dans ces deux cas mais elle semblerait n'avoir pas encore atteint une fréquence ou des modalités d'usage telles que les conséquences en matière d'emploi et de travail seraient massives. Les réactions des collectifs de travailleurs posent en tout cas la question de la réglementation de ces usages, des outils et des possibilités qu'elles ouvrent ou pas, de prendre en compte à la fois les conséquences sur l'emploi, sur le travail, voire sur les caractéristiques des produits et services.

4.2. Régulation du développement et des usages de l'IA

La notion de régulation est très largement utilisée pour aborder à la fois la mise en place de règles, la nature de ces règles (niveau de définition, degré de contrainte), leur champ d'application. Dans le champ notamment de l'IA, cette notion fait largement référence à la notion de « *regulation* » (en anglais) employée notamment en droit états-unien, qui recouvre une palette d'outils allant de la charte interne à une entreprise à la loi, en passant par la charte de bonne conduite, les recommandations, la jurisprudence.

La réflexion sur une régulation de l'IA a beaucoup évolué au cours des six dernières années, au fur et à mesure du développement des technologies et des usages des IA. Les réflexions initiales s'inscrivent surtout dans une analyse de l'innovation, en termes de processus et d'effets. Assez rapidement des réflexions sur la régulation des conséquences de l'IA ont été engagées, notamment dans le cadre d'élaboration des stratégies nationales. Les conséquences sur l'emploi et le travail n'en représentent qu'une très faible partie.

L'adoption en décembre 2023 de la directive européenne sur l'intelligence artificielle dite « AI Act » tend à centrer l'attention sur une discussion des modalités juridiques de régulation, et en particulier ce qu'elles régulent et leur caractère plus ou moins contraignant. La plupart des travaux qui envisagent une régulation de l'IA mobilisent une analyse du droit, combinée ou non à des analyses d'autres disciplines.

4.2.1. Réglementer ou pas, avec quels instruments ?

Différents arguments sont avancés en faveur et en défaveur d'une régulation sous forme de réglementation par exemple dans l'analyse de l'impact de l'IA sur le marché du travail de l'OCDE (2023).

²¹ T. Le Bonniec (2024), « Dégénérescence numérique du journalisme », Blog Médiapart, 10 février, <https://blogs.mediapart.fr/thomas-le-bonnec/blog/100224/degenerescence-numerique-du-journalisme>.

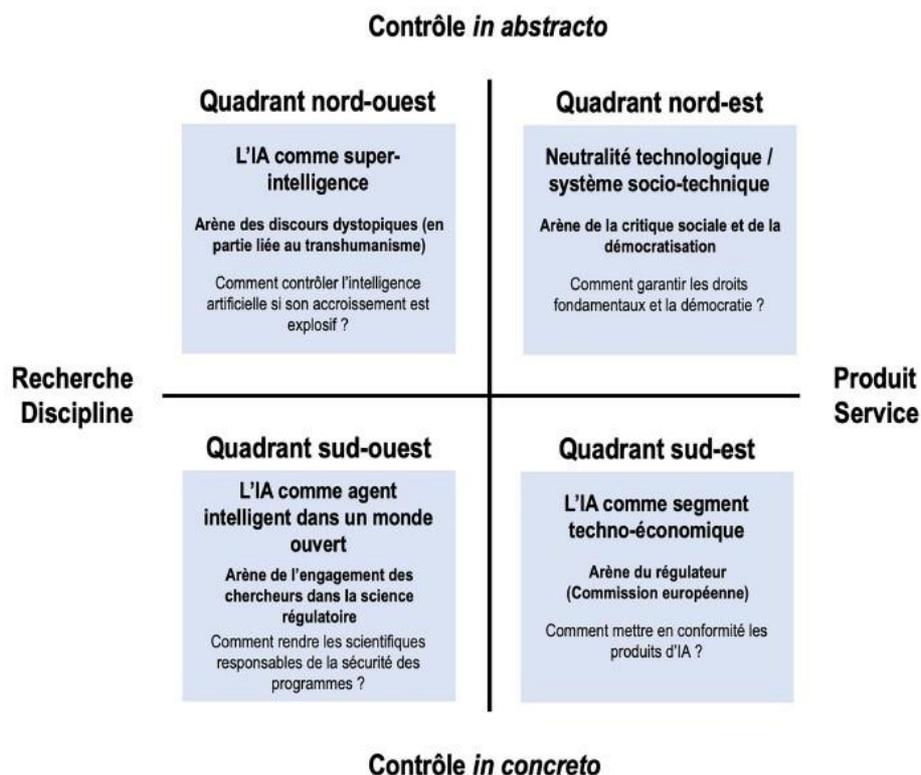
Ces arguments relèvent pour la plupart d'une analyse de l'innovation comme source de croissance. Il s'agit alors de laisser l'innovation se diffuser, de ne pas mettre en place d'obstacles qui pourraient limiter les effets positifs de l'innovation, de ne pas pénaliser la compétitivité des entreprises utilisatrices. Les arguments en faveur d'une réglementation relèvent plutôt d'une approche évolutionniste, d'une approche intégrant le rôle des institutions. La réglementation peut protéger des innovations dans les premiers temps de leur développement, par exemple en fixant des normes techniques, ou favoriser leur diffusion en réduisant leurs éventuels effets défavorables.

Les modalités de régulation ont fait l'objet de nombreuses études : en France par le GT3 Cellule Ethique du numérique en santé (ministère de la Santé), la Haute autorité de la santé (HAS), au niveau communautaire par le Groupe d'experts de haut niveau de la Commission européenne, au niveau international par l'OMS, l'UNESCO, l'OCDE, les Ministères de la santé au Royaume-Uni et au Canada, etc.

Sur quoi la régulation porte-t-elle ? Cette question préalable s'impose en raison du flou du terme IA et de ce qu'il désigne. Benbouzid, Meneceur et Smuha (2022) distinguent quatre champs de régulation en croisant deux dimensions. La première dimension tient à la distinction entre l'IA ensemble de savoirs produits, discipline scientifique, et l'IA ensemble de produits et services dérivés des savoirs, systèmes techniques visant à accomplir des actions qui demanderaient une certaine intelligence si elles étaient accomplies par des humains. La seconde dimension met en regard un contrôle de portée générale, qualifié d'*in abstracto*, qui porte sur les « risques existentiels », l'éthique ou le droit des personnes, et un contrôle dit *in concreto* qui porte sur « l'analyse du réel au cas par cas des situations dans lesquelles opèrent les systèmes » (Benbouzid *et al.*, 2022 ; p. 37). Le découpage en quadrants (voir Figure 3) permet en particulier de situer ce qui se joue d'une part dans le champ du contrôle *in abstracto* (partie nord) et d'autre part en ce qui concerne les produits et services (côté est).

Les deux quadrants nord nous permettent d'envisager l'empan du champ concerné par des régulations générales. Ce contrôle peut prendre la forme de règles en matière d'éthique, de droits fondamentaux (nord-est) à charge pour les acteurs économiques de s'y conformer, selon le caractère plus ou moins contraignant de ce contrôle. Une illustration en est la Recommandation de l'UNESCO sur l'éthique de l'intelligence artificielle qui pose 14 valeurs et des principes. Les pays membres de l'Unesco se sont engagés en novembre 2021 à respecter cette recommandation. Ces valeurs et principes ne sont toutefois pas hiérarchisés et les conflits entre eux ne sont pas explicités. L'expression « IA de confiance » traduit un choix explicité par la Recommandation : créer les conditions de confiance des citoyens, travailleurs et consommateurs pour favoriser un développement et une diffusion des outils d'IA. On retrouve dans cette logique les dispositions visant à sécuriser l'usage et le stockage des données, la cybersécurité, à assurer qu'un consentement est requis, à empêcher l'utilisation de données portant atteinte aux droits fondamentaux et à la vie privée. La notion « IA centrée sur l'humain » largement présente dans les réflexions des universitaires, dans les régulations produites au niveau international, traduit également le souci de garder un contrôle humain sur des systèmes de contrôle et de décision intégrant de l'IA.

Figure 3 : L'espace social de la régulation selon 4 arènes normatives différentes (repris de Ben Benbouzid et al., 2022)



Le contrôle des produits et services peut aussi prendre la forme de règles concrètes sur les produits et services, limitant par exemple certains usages ou les subordonnant à une évaluation (sud-est). La proposition de directive de la Commission Européenne rendue publique le 21 avril 2021 et votée par le Parlement européen en décembre 2023 s'inscrit dans ce quadrant. On peut aussi y situer l'engagement signé le 5 février 2024 par huit très grandes entreprises du secteur de la technologie (GSMA, INNIT, Lenovo Group, LG AI Research, Mastercard, Microsoft, Salesforce et Telefonica) de mettre en œuvre les principes et valeurs de la Recommandation de l'Unesco à chaque étape, de la conception, du développement, de l'achat, de la vente et de l'utilisation de l'IA²².

Les instruments juridiques pour réguler l'IA font l'objet d'une littérature très abondante, en particulier la directive communautaire sur l'intelligence artificielle. Cette littérature permet de comprendre comment la régulation de l'intelligence artificielle est conçue et de questionner la prise en compte des effets sur l'emploi et le travail. A la différence de la Recommandation de l'Unesco, dont la prise en compte repose sur une auto-régulation des acteurs, publics comme privés, la directive communautaire vise à réguler plus directement un certain nombre de comportements. L'objectif de créer une « IA de confiance » est à la fois éthique et économique : le respect des droits des individus (consommateurs et citoyens) est

²² Voir le document signé est accessible à l'adresse suivante : https://www.unesco.org/sites/default/files/médias/fichiers/2024/02/gfeai_private_sector_05022024.pdf.

essentiel à la diffusion des technologies et des usages des IA. Le choix européen est de réguler diversement les usages de l'IA selon le niveau de risques qu'ils présentent du point de vue des droits fondamentaux, de la santé et de la sécurité. Cette approche graduée selon les risques vise à concilier respect des droits fondamentaux et sécurité de l'IA d'un côté, et limitation des coûts de mise en conformité de l'autre (Bertrand, 2023). La régulation forte concerne uniquement les systèmes d'IA qualifiés de « à haut risque ». Dans la directive sur l'IA, les usages de l'IA relatifs à l'emploi, la gestion des travailleurs et l'accès à l'emploi indépendant relèvent de ce niveau de risque. Ils incluent les processus de recrutement, la gestion des ressources humaines et l'organisation du travail, l'appréciation des performances et des comportements, les décisions de rupture des relations contractuelles avec les travailleurs. Toutefois l'évaluation uniquement interne à laquelle ces usages sont soumis semble assez faible comparée aux risques pour les travailleurs (Molina, Butollo, Makó *et al.*, 2023).

Quels que soient les instruments juridiques mis en place, ils ne peuvent être considérés isolément. Leur effectivité tient aussi à d'autres instruments juridiques, qu'ils complètent ou renforcent (Martin, 2023 ; Poulet, 2023). Ainsi la directive communautaire sur l'IA complète la directive communautaire sur la numérisation de l'économie et le Règlement sur la protection des données (RGPD). Les conséquences en termes de discrimination de l'usage d'algorithmes pour la gestion des travailleurs (par exemple dans le cas de la livraison de produits alimentaires) sont aussi couvertes par les directives communautaires anti-discrimination transposées dans les droits nationaux. L'effectivité de ces différents instruments dépend de la date de leur entrée en vigueur, avec un délai plus ou moins long par rapport à leur adoption et une entrée en vigueur progressive dans certains cas. Ainsi la directive sur les services numériques est entrée en vigueur le 16 novembre 2022 pour les plus grands moteurs de recherche et plateformes, et s'applique à tous les producteurs et diffuseurs de services numériques depuis le 16 février 2024.

Ces sources juridiques combinées posent la question de leur cohérence et de leurs lacunes. Poulet (2023) en pointe de nombreuses à propos de la santé. Par exemple, il alerte sur le fait que les données anonymisées, dont l'usage est autorisé, peuvent désormais être repersonnalisées du fait de la puissance des IA actuelles. Ponce del Castillo (2020) souligne la nécessité de mettre en place un « régime », un ensemble de règles plus structuré et plus cohérent pour renforcer la loi, les droits humains et la participation des travailleurs et de leurs représentants. Dans la directive sur l'IA, aucune régulation spécifique, ni par la législation des États membres ni par la négociation collective n'est prévue, alors qu'elle l'est dans l'article 88 du Règlement sur la protection des données (RGPD). Dans sa version finale de décembre 2023, le projet de directive sur les travailleurs des plateformes comportait plusieurs dispositions importantes : une présomption de travail salarié, une obligation de transparence à l'égard des travailleurs et de leurs représentants quant aux algorithmes utilisés, l'interdiction de décisions importantes par des systèmes sans contrôle humain, l'obligation d'évaluer les conséquences de décisions prises par ou avec des systèmes automatisés sur les conditions de travail, la santé, la sécurité et les droits fondamentaux, la protection des données des travailleurs (salariés ou non-

salariés). Son rejet par le Conseil européen²³ signifie l'abandon du projet après plus de deux années de procédure législative. Le développement de l'emploi via des plateformes d'emploi (ubérisation, micro-travail) dépend donc de décisions de juges nationaux en cas d'action en justice de travailleurs demandant la requalification de leur statut d'emploi en celui de salarié, ou d'accords collectifs.

4.2.2. *Dialogue social et accords collectifs*

Le développement des outils d'IA et de leurs usages, avérés ou potentiels, a des conséquences à différents niveaux, notamment sur les organisations, sur les relations d'emploi et sur le fonctionnement du marché du travail. La négociation collective, la participation des salariés et de leurs représentants constituent des modes de régulation en principe adaptés à la prise en compte des effets réels et potentiels des IA sur les travailleurs, des effets différenciés selon les secteurs d'activité et selon les entreprises, et à des ajustements plus rapides aux évolutions du contexte.

Dans quelle mesure ces modes de régulation contribuent-ils à modeler les IA ? Quelle place prennent-ils dans la régulation d'ensemble des IA ? Quels effets ont-ils ? Ces trois questions structurent largement les travaux.

Des travaux quantitatifs ont été conduits par l'OCDE sur la façon dont la négociation collective, la participation des salariés et de leurs représentants contribue à configurer les IA et leurs usages. Krämer et Cazes (2022) ont produit une revue de littérature très complète sur les liens entre dialogue social et IA. Elles pointent la rareté des travaux portant à la fois sur des outils d'IA (et pas sur des robots) et sur des indicateurs de négociation collective, et l'absence de travaux concernant l'effet de la représentation des travailleurs sur la qualité de l'emploi et les conditions de travail. A partir des données de la 3^{ème} enquête européenne des entreprises sur les risques nouveaux et émergents (ESENER-3²⁴) et d'un modèle de régression probit, elles concluent à l'existence d'une interaction entre la représentation des salariés et certaines conditions de travail, mais le lien de causalité reste indéterminé.

De très nombreux travaux qualitatifs ont été menés sur les modalités et le rôle de la régulation des IA par la négociation collective dans un cadre national ou en comparaison internationale, sur des cas d'entreprise ou de secteurs (Cazes, 2023 ; De Stefano et Doellgast, 2023). Comme le relèvent De Stefano et Doellgast (2023), ces analyses mobilisent plusieurs disciplines et sous-disciplines : relations professionnelles, systèmes d'emploi, droit, sciences politiques, gestion, *social sciences*...

²³ Les deux dernières versions du texte, résultats d'un compromis avec le Parlement européen, ont été rejetées par le Conseil européen le 22 décembre 2023 et le 16 février 2024.

²⁴ En 2019, l'Agence européenne pour la sécurité et la santé au travail (EU-OSHA) a interrogé 45 420 établissements de différents secteurs dans 33 pays sur les risques pour la santé et la sécurité sur le lieu de travail. L'enquête ESENER-3 comprend des informations détaillées sur les différentes formes de représentation des travailleurs sur le lieu de travail, sur les différents types de technologies utilisées (avec une attention particulière à la numérisation) et sur certains aspects des conditions de travail non monétaires.

L'analyse des accords collectifs à différents niveaux (national, sectoriel, organisation) est très présente par exemple dans les travaux récents publiés par l'OCDE (Cazes, 2023 ; Krämer et Cazes, 2022), par l'Institut syndical européen/ETUI (Lamannis, 2023 ; De Stefano et Doellgast, 2023).

Molina, Butollo, Makó *et al.* (2023) sont un exemple d'analyse comparative couplant relations professionnelles et droit : ils ont analysé les instruments de régulation des usages de l'IA et leur évolution dans quatre pays européens. Dans deux d'entre eux, une adaptation incrémentale semble s'être mise en place. Au Danemark, la loi ou les accords collectifs de coopération (là où ils s'appliquent) obligent les directions à discuter avec les salariés et à les informer sur l'introduction de nouvelles technologies et leurs implications, y compris sur l'organisation du travail. Des accords ont été signés récemment dans des secteurs très exposés mais aussi dans des activités où se diffuse l'usage des IA pour réguler les usages. En Allemagne, la loi donne aux conseils d'entreprise des droits concernant l'information, le conseil et la codétermination qui s'appliquent aux usages de l'IA, y compris à la gestion des ressources humaines par des outils d'IA. En 2021, la loi sur la modernisation des conseils d'entreprise a renforcé les droits des conseils d'entreprise en ce qui concerne l'introduction et l'usage de l'IA. Les conseils peuvent faire appel à un expert pour évaluer l'IA ou le management par algorithme. En Espagne, ce sont surtout des lois qui régissent l'usage de l'IA : la transposition du RGPD et une loi de 2021 sur les droits des travailleurs des plateformes, issue d'un accord des partenaires sociaux. En Hongrie enfin, il y a peu de dispositions légales à part celles qui résulte de l'application du RGPD, des modalités de régulation plutôt souples, très peu de dialogue social sur le sujet et une absence d'accords sectoriels.

A l'exception de l'Accord-cadre des partenaires sociaux européens sur la numérisation signé le 22 juin 2020 par BusinessEurope, la CES, le CEEP et SMEunited, les accords conclus par des partenaires sociaux le sont surtout dans des secteurs et dans des entreprises très exposés (Banque et Assurance au niveau européen, livraison de produits alimentaires dans de nombreux pays) et concernent surtout le management algorithmique. La négociation collective sur le développement et les usages de l'IA présentent de très grandes différences selon les pays, différences qui tiennent beaucoup au modèle de relations professionnelles qui prévaut (Molina, Butollo, Makó *et al.*, 2023). Un point à souligner est la possibilité très inégale selon les systèmes de relations professionnelles de discuter avec les directions voire de négocier l'organisation du travail. Cette possibilité existe dans des pays où les syndicats sont très implantés, avec des instances de co-détermination et dans les entreprises couvertes par des accords collectifs le prévoyant. C'est pourquoi l'Accord-cadre des partenaires sociaux européens sur la numérisation ne mentionne pas l'organisation du travail²⁵.

²⁵ Voir l'intervention de Maxime LEGRAND, président de la Confédération Européenne des Cadres lors de la séquence 3 « Quelle place pour le dialogue social dans l'utilisation de l'IA au travail ? » du colloque de la 37^{ème} session de l'INTEFP, 30 juin 2023. https://www.intefp.tv/video/649-S_quence_3_Quelle_place_pour_le_dialogue_social_dans_l_utilisation_de_l_IA_au_travail_1_re_partie_des_ateliers_de_la_37_me_session_nationale

A Retenir...

- Les travaux qui permettent d'articuler les effets sur l'emploi et les effets sur le travail sont relativement peu nombreux.
- Deux voies d'articulation ont été repérées. La première passe par l'analyse d'éléments de contexte plus larges. Par exemple pour comprendre en quoi l'introduction d'un SIA peut modifier le travail des personnes qui l'utilisent, la répartition des tâches, réduire les emplois ou accompagner leur réduction, créer des emplois. La seconde passe par les modes de régulation, leurs champs d'application et leur prise en compte ou non des questions d'emploi et de travail.
- Les approches monographiques ou par études de cas, au niveau d'entreprises ou de secteurs, sont les plus à même d'envisager les effets fins sur le travail, sur l'organisation du travail et de permettre de les relier aux effets sur les emplois. Une partie des effets sur l'emploi est peu visible (délocalisation, sous-traitance, départs naturels, réduction des effectifs déjà engagée).
- Les travaux sur la régulation de l'IA soulignent la nécessité de penser une régulation d'ensemble pour pallier les incohérences et les lacunes des instruments juridiques actuels.
- La possibilité de réguler les conséquences des SIA sur le travail et l'emploi dépend beaucoup du modèle de relations professionnelles. De grandes différences existent entre les pays.

5. Usages de l'IA et transformations du travail : de la projection aux usages en situation réelle de travail

L'objectif de cette section est d'apporter un éclairage sur les usages actuels ou potentiels de l'IA dans des contextes professionnels, tout en mettant en lumière les transformations effectives et/ou potentielles du travail et les conséquences de ces transformations pour les travailleurs et leurs activités. Il s'agit ici d'aborder non seulement ces conséquences en termes de modification des tâches, de performance ou de risques, mais aussi et surtout en termes de transformation des activités concrètes réalisées *in situ*. Ce sont elles qui in fine déterminent la valeur accordée (par exemple, en termes de performance) à « l'IA-tisation » (c'est-à-dire la mise en place d'un SIA dans un environnement de travail) et permettent de saisir les conséquences effectives de la transformation technologique sur le travail et les travailleurs (performance, santé, qualité, fiabilité, autonomie...).

Pour répondre à cet objectif, nous nous appuyons sur les travaux de recherche identifiés selon la méthodologie décrite en Annexes 1. Trois approches empiriques - parfois combinées - peuvent être distinguées pour produire les connaissances scientifiques dans ces travaux :

- **Les études qui (1) explorent les points de vue des acteurs** (potentiellement) concernés (employés quel que soit leur niveau hiérarchique) par l'IA, sur ce qu'elle peut apporter, sur ses risques, et/ou **(2) produisent des scénarios d'usage potentiels** – des projections - impliquant des transformations du travail. Ces études s'appuient le plus souvent sur des scénarios d'usages réels ou potentiels – qui ne correspondent pas forcément à ceux qui seront in fine mis en œuvre en situation - et sur des points de vue qui ne sont pas toujours étayés par des expériences d'usage concrètes.
- **Les études qui analysent les usages et les conséquences de l'introduction de l'IA dans le cadre d'expérimentations.** La majorité des études reposant sur cette approche est centrée sur l'analyse des « effets » de l'utilisation sur la performance, en comparant l'humain à l'IA ou dans le cadre de différents scénarios de « couplage » humain-IA. Une bonne partie de ces études est réalisée dans le secteur médical (notamment en radiologie) et souvent sur la base de tâche uniquement individuelle. La généralisation des résultats de ces études expérimentales de travail à des situations de travail réelles - plus complexes et impliquant une diversité de professionnels - est donc à discuter.
- **La troisième approche renvoie aux études des SIA en situation réelle, c'est-à-dire dans des conditions opérationnelles.** Bien que ces études soient moins nombreuses, elles apportent des éléments pour une compréhension plus approfondie des transformations et des conséquences de l'utilisation de ces systèmes dans des contextes réels.

Les sections suivantes exposent les éléments principaux abordés et mis en avant dans ces études, nous n'en mobiliserons que certaines pour étayer notre propos et proposerons des « zooms » sur certaines d'entre elles particulièrement illustratives. Sur cette base, nous discutons : (1) de l'utilité de la valeurs et des risques perçus concernant la performance du travail réalisé avec une IA, , nous discutons

des performances des SIA, les apports et les risques liés aux leurs usages ainsi que les enjeux de transformation du travail avec les facteurs à considérer ; (2) des connaissances produites expérimentalement sur la performance dans des tâches réalisées avec une IA ; (3) des connaissances produites par des études s'intéressant à des situations de travail réel et enfin (4) les intentions à la base des projets d'IA et leurs conduites.

Les études explorant le point de vue des acteurs – parfois en lien avec des expériences concrètes d'usage de l'IA – permettent de mettre au jour la valeur accordée a priori aux tâches « IAtisées », à l'utilité perçue par les professionnels, mais aussi aux risques concernant la performance et la santé au travail (qualité, fiabilité...), que l'IA pourrait entraîner.

5.1. Utilité, valeur et risques perçus pour la performance du travail réalisé avec une IA

5.1.1. Un point de vue nuancé sur les bénéfices-risques de l'IA et une valeur a priori accordée aux tâches automatisables discutables

Deux études portant sur l'automatisation de la consultation et de la modification de dossiers patients par des médecins généralistes mettent en avant les points de vue nuancés des acteurs (médecins) sur les apports et risques liés à l'introduction potentielle de SIA ainsi que sur la valeur accordée a priori aux tâches automatisables.

(Navarro et al., 2022) se sont ainsi intéressés aux points de vue des médecins généralistes sur un SIA s'appuyant sur des technologies du traitement automatique du langage naturel pour l'automatisation du traitement de la documentation clinique (dossier médical du patient). En menant des entretiens semi-directifs avec dix médecins généralistes, les auteurs cherchaient à explorer les points de vue de ces professionnels en s'appuyant sur divers scénarios pour imaginer comment l'IA pourrait concrètement être utilisée dans le travail.

Du côté des fonctionnalités techniques proposées, le système permet l'extraction d'informations, la transcription via la reconnaissance vocale, la synthèse de contenu de données sur des patients mobilisables à différentes phases de leur travail (avant, pendant et après une consultation). Le système envisagé soutiendrait la possibilité de garder le contrôle des dossiers patients et serait « explicable », c'est-à-dire qu'il pourrait générer des explications permettant de comprendre ses sorties.

Du côté des apports et de l'utilité perçus, les médecins ont une vision positive des systèmes de traitement automatisé des textes. L'automatisation des tâches de documentation clinique (extraction d'informations, analyse de données cliniques, etc.) est supposée leur permettre de consacrer davantage de temps à la prise en charge de leurs patients, ce qui confère une utilité *a priori* au SIA.

Du côté des « risques » et de la performance au travail, les médecins mettent en avant des obstacles à surmonter pour assurer la mise en œuvre et le développement d'un usage pertinent de l'IA, en particulier, des écueils liés à l'automatisation (par exemple, lorsque l'automatisation ne permet pas de

détecter et corriger des erreurs), les questions juridiques et réglementaires liées à la confidentialité et la sécurité des données médicales, ainsi que l'interopérabilité avec les systèmes déjà existants. Ces risques et écueils potentiels questionnent les possibilités de réaliser un travail de qualité donc la valeur réelle et l'intérêt du SIA, son intérêt. Si dans cette étude, le point de vue des médecins généralistes interrogés paraît plutôt positif, l'étude de (Willis & Jarrahi, 2019) la nuance en complétant cette première identification des risques associés à l'automatisation totale du traitement de la documentation clinique. Le travail de Willis et Jarrahi (2019) qui porte lui aussi sur le travail des médecins généralistes, se concentre sur le traitement de la documentation clinique (constitution du dossier médical du patient) afin d'explorer les possibilités et les problèmes éventuels liés à l'automatisation de certaines pratiques cliniques. Après avoir analysé le travail des médecins (à l'aide d'entretiens et d'observations), les auteurs construisent un scénario d'usage potentiel d'un SIA.

Les fonctionnalités en jeu concernent l'utilisation de la reconnaissance vocale pour transformer automatiquement en texte la discussion entre le patient et le médecin lors de la consultation et l'intégrer automatiquement dans le dossier informatisé du patient ; après la consultation clinique, le médecin pourrait modifier le résumé produit automatiquement. Ce scénario envisage donc une automatisation du processus de documentation des échanges entre le patient et son médecin afin de permettre une interaction directe avec le patient sans interruption. La valeur ajoutée du système résiderait dans le fait de favoriser le plus possible de réels échanges entre le patient et le médecin, d'automatiser un travail de documentation considéré comme fastidieux, et la tâche d'écriture qui lui est associée.

Cette étude illustre l'écart existant entre valeur projetée a priori et la valeur réelle de la tâche « *IAisée* » pour la qualité du travail. En effet, la projection sur la valeur de la tâche d'écriture ne tient pas compte du rôle de son rôle : celle-ci soutient la réflexion, la mobilisation des expériences passées du médecin et son analyse critique des choix potentiels de diagnostic et de traitement et in fine la prise de décision thérapeutique. Priver brutalement les médecins de cette tâche, sans réflexion sur le développement d'autres modalités de « *raisonnement dans l'action* » du médecin, reviendrait potentiellement à dégrader la prise en charge du patient (qualité et fiabilité du travail). De plus, la transcription complète des échanges n'est pas utile pour le médecin qui sélectionne des informations pour poser son diagnostic comme cela a été montré par le passé lors de l'introduction des premiers échographes (Ochanine, 1978).

5.1.2. Du besoin de fiabilité et de contrôle de la qualité des « décisions » des SIA

Des études réalisées dans les domaines des Ressources Humaines (RH), du journalisme et de la justice renforcent les résultats précédents tout en les étendant et en soulignant les enjeux d'explicabilité et de contrôle des SIA notamment en lien avec les dimensions de fiabilité des résultats/informations produits par l'IA et d'éthique, qui sont deux éléments essentiels de la performance du au ? travail.

Dans le domaine des RH, les SIA proposent désormais des *fonctions* permettant le traitement de grande quantité d'informations pour sélectionner les candidats appropriés pour des postes et d'évaluer leurs profils. La *valeur a priori* de l'IA réside dans l'automatisation de tâches considérées comme répétitives

de recherche et de sélection des candidats, pour « *gagner du temps* ». Cependant, Chen, (2022) identifie utilement des critères de réussite liés à la performance dans différentes tâches d'un processus de recrutement réalisés à l'aide de l'IA. Par exemple, pour la tâche concernant la diffusion d'une offre d'emploi, l'exactitude dans la description de poste est un critère considéré comme important. Il s'agit ici d'aider les recruteurs à répondre aux exigences dans la rédaction des offres d'emploi. Lors de la tâche de sélection, la valeur de l'IA reviendrait pour le recruteur à pouvoir s'assurer de la fiabilité de l'IA en comparant ses décisions à leurs propres décisions. Si l'IA peut aider ici à trier rapidement lorsqu'il y a beaucoup de candidats, il s'agit néanmoins de garantir la qualité et la pertinence des résultats produits par l'IA : ne pas passer à côté des « talents » potentiels et les sélectionner. Enfin, l'étude évoque un aspect important de performance du travail, lors de l'étape d'évaluation des candidats : l'équité et l'impartialité de l'évaluation conduite par l'IA. Partant de ces postulats, Chen (2022) explore les points de vue de différents utilisateurs potentiels en interrogeant 15 responsables des ressources humaines (recruteurs, managers).

Du côté des apports, il ressort que pour les interviewés, l'IA permettrait un recrutement plus rapide, une embauche de qualité, une diminution de la charge de travail et une réduction de la discrimination. On retrouve donc là aussi une perception positive de l'IA, mais avec un certain nombre de conditions qui révèlent en creux des risques perçus.

Du côté des risques concernant la performance du travail et leurs « préventions » :

- les interviewés estiment qu'il est plus facile de faire confiance au SIA s'il peut être testé et évalué sur la base d'expériences d'embauche antérieures réussies au sein de l'entreprise, en lien avec le besoin de fiabilité des résultats produits ;
- ils insistent sur la nécessité de comprendre pourquoi un SIA « décide » si un candidat doit être rejeté ou sélectionné renvoyant aux enjeux d'explicabilité déjà évoqués ;
- ils indiquent que le classement des notes des candidats fourni par le SIA ne devrait pas être une décision finale afin de donner aux professionnels le sentiment de « garder le contrôle » sur le processus d'embauche « guidé » par l'IA ;
- de plus, les interviewés souhaiteraient disposer de la possibilité de modifier les décisions inappropriées de l'IA ;
- enfin, ils considèrent que pour la construction de la confiance envers l'IA, il est nécessaire de tester sa fiabilité en comparant ces « décisions » de l'IA à celles d'un recruteur humain.

Une autre étude menée dans le secteur du journalisme cette fois (Amaya Sánchez, 2022) renforce cette identification de risques en lien avec les dimensions éthiques et juridiques liées aux résultats de l'IA. Là aussi, les personnes interrogées s'accordent sur *une valeur a priori* de l'IA concernant une amélioration des capacités des journalistes en termes de « gain du temps » et « d'augmentation de l'efficacité » des processus de production d'informations. Toutefois, les interviewés soulignent l'émergence de *questions et risques éthiques et juridiques* (risques de désinformation, pluralité et diversité des informations, propriété des données, règles de protection des données, etc.) qui, pour l'auteur, soulignent la nécessité d'un contrôle et d'une supervision continus des processus qui sont gérés par l'IA. Enfin, les personnes

interrogées insistent sur le rôle du journaliste humain en tant qu'agent irremplaçable et en tant que professionnel qui doit superviser les résultats de l'IA.

Enfin, une dernière étude réalisée dans le domaine juridique (Licoppe & Dumoulin, 2019) met en évidence les questions de responsabilités, d'indépendance et d'autonomie des professionnels face aux « décisions » de l'IA. L'étude concerne la mise à disposition d'un SIA relevant de la « justice prédictive » (une notion critiquée)²⁶ à des magistrats. La valeur *a priori* du système résiderait dans une « aide à la décision judiciaire » dont la performance serait à renforcer. Dans cette étude, il s'agissait de comprendre ce que peut apporter le système pour définir la grille d'indemnisation après licenciement et le référentiel d'indemnisation des dommages corporels.

Du côté de l'utilité perçue, les magistrats interrogés considèrent que les SIA basés sur les données de justice ouvertes présentent un intérêt potentiel, sous réserve d'interpréter davantage les grilles et barèmes produits dans la continuité des grilles et barèmes existants, plutôt que de les considérer comme une rupture – des nouveautés sans les replacer dans leur contexte. Les magistrats reconnaissent que ce type de système permet une nouvelle visualisation de données et introduit de nouvelles possibilités de recherche, en complément des dispositifs numériques ou non dont ils disposent. *Du côté des risques*, les professionnels soulignent les limites du système en termes de pertinence des résultats et sa limitation lorsqu'il s'agit d'introduire des critères de recherche originaux. Enfin, le déploiement de ce système questionne l'indépendance et l'autonomie des juges dans le traitement singulier des cas dans le cas où les SIA entraîneraient une forme de standardisation des décisions. Cette valeur essentielle de la performance du travail en termes de qualité et d'éthique pourrait être mise à mal par les SIA.

5.1.3. L'IA au service de la fiabilité des actions

Sujan et al. (2020) explorent un scénario d'usage d'une pompe à perfusion autonome dans les soins intensifs à partir d'une analyse du « système sociotechnique ». Ils mettent en évidence quatre dimensions essentielles à prendre en compte pour s'assurer de la « sûreté » d'une action lorsqu'un SIA est intégrée dans une situation de travail « à risque », tel que le milieu médical :

- Le besoin de fonctionnalités de « transfert » renvoyant à la capacité du système à reconnaître ses limites et à confier le contrôle de l'action aux professionnels de santé ;
- Le besoin de soutenir des exigences de performance variable renvoyant à la capacité du système à assister les professionnels de santé dans leur gestion des priorités et des exigences des situations spécifiques ;

²⁶ La notion de « justice prédictive » a été critiquée à de nombreuses reprises (CEPEJ, 2019; Meneceur, 2020) tout d'abord pour un abus de langage. Ce terme est considéré comme ambigu. En effet, « la prédiction », soit le fait d'évoquer « par avance des événements futurs » (CEPEJ, 2019) p. 32, serait à différencier de « la prévision » qui elle « résulte de l'observation (*visere*, voir) d'un ensemble de données afin d'envisager une situation future » (p. 32). Les systèmes de justice dite prédictive s'inscriraient davantage dans une visée prévisionnelle. Ils fournissent des probabilités concernant l'issue d'un litige à partir de la modélisation statistique des décisions de justice déjà rendues, avec l'objectif final d'aider à la prise de décision. Or, ces modélisations ne permettraient pas de rendre compte ou de s'approcher à la complexité d'un raisonnement juridique. Ces systèmes seraient accompagnés aussi des risques d'interprétations erronées des décisions de justice ou encore des biais de discrimination.

- La prévention du biais d'automatisation renvoyant au risque de sur-confiance et dépendance excessive envers le système ;
- La supervision du système : les professionnels de santé sont amenés à utiliser mais aussi surveiller le système c'est-à-dire comprendre son fonctionnement ou des dysfonctionnements.

Les préoccupations relatives à la sûreté des actions avec une IA impliquent donc un certain nombre de risques concrets pour la santé des patients.

5.1.4. Conséquences pour la santé et le sens du travail

D'autres études apportent des éléments intéressants sur la manière dont l'introduction de SIA pourrait affecter positivement ou négativement la santé des travailleurs et le rapport que ceux-ci construisent et entretiennent avec leur travail, c'est-à-dire le sens du travail.

L'étude de Ferguson (2019) s'appuie sur un groupe de travail « Humain-IA » mis en place au sein d'une entreprise, chargé de réfléchir aux principes d'une IA au service de l'humain. La valeur a priori des SIA réside là aussi dans la reprise des discours autour de la prise en charge de tâches considérées comme routinières. Cette étude complète cependant un certain nombre de risques pour la performance du travail en lien notamment avec la santé et le sens du travail : les interviewés soulignent en effet un appauvrissement intellectuel ou encore une déresponsabilisation et une perte d'autonomie face aux décisions prises par l'IA. Les travailleurs craignent de perdre leurs compétences et leur autonomie. Face à ces risques, l'entreprise étudiée met l'accent sur les recommandations concernant l'explicabilité, la transparence et les dimensions éthiques à intégrer dès la conception.

Zoom sur les points de vue des acteurs sur l'IA dans le domaine juridique (Gamkrelidze, 2022)

Cette étude permet de donner à voir dans un cas spécifique et détaillé un certain nombre d'éléments mentionnés dans cette section. Sa présentation est proposée avec une visée d'intégration de ceux-ci dans un exemple concret.

Méthodologie de l'étude

14 personnes ont été interviewées exerçant métiers différents (avocats, magistrats, juristes, responsables et fondateurs d'entreprises développant des SIA juridiques).

Résultats

L'analyse des entretiens a permis d'identifier plusieurs fonctionnalités IA considérées comme potentiellement utiles par les interviewés pour leur métier (avocat, magistrat ou juriste). Ces fonctionnalités se structurent autour d'une part de la recherche, l'analyse et la synthèse d'informations, et, d'autre part, le calcul d'estimations (par exemple, probabilités de gain ou perte d'un procès) dans une perspective d'aide à la prise de décisions (voir Tableau 3). Plusieurs bénéfices possibles sont pointés par les interviewés comme le gain de temps ou des améliorations en termes de performance

Tableau 3. Fonctionnalités de l'IA décrites comme utiles par des acteurs du domaine juridique

Fonctionnalités de l'IA		Utilité
RECHERCHE - SYNTHÈSE - ANALYSE	Recherche, veille et analyse des décisions de justice : guidée, fine avec une haute précision	Identifier des jurisprudences pertinentes avec potentiellement un degré supérieur pour devancer l'adversaire Disposer d'une meilleure vision de la jurisprudence et renforcer la sécurité juridique c'est-à-dire réduire les erreurs potentielles dans la prise de décision des professionnels du droit : - Identifier des éléments pertinents et cohérents dans les bases de données juridiques qui ne cessent d'augmenter - Évaluer la cohérence d'une décision, identifier les points à examiner selon les questions juridiques afin d'aider au raisonnement juridique du juge
	Synthèse de recherche d'informations	Identifier et mettre en avant de manière synthétique des informations et des éléments importants et pertinents vis-à-vis du sujet traité
	Aide à la production des documents	Élaborer des documents intelligibles avec une forme et un fond amélioré
	Traitement et analyse automatique des documents	Effectuer une première analyse de nombreux documents constituant des dossiers juridiques afin d'identifier des éléments pertinents selon les questions et les points à traiter Extraire des informations clés et des éléments de réponse dans des situations où les professionnels du droit doivent organiser et passer en revue un grand nombre de documents (par exemple, dans le cas des audits juridiques)
ESTIMATIONS	Calcul de probabilité (de succès de l'issue d'un litige), d'estimation (des barèmes) : ensemble de systèmes qualifiés de « prédictifs »	Identifier des moyennes et des tendances de décisions (par exemple, des tendances prévisionnelles pour l'administration de la justice, la gestion des tribunaux) Disposer d'une vision globale sur ces tendances pour aider à la prise de décision (servir de « GPS ») aux : - magistrats : avoir une vision sur des affaires similaires, aider dans l'acte de juger (utile seulement dans des cas complexes ou ceux dans lesquels il n'y a pas encore de jurisprudence) - professionnels du droit : notamment pour les « généralistes » et les moins expérimentés, aider un avocat dans l'acte de conseiller (utile pour un avocat expert si seulement le système donne une analyse fine et poussée) - assureurs (estimation des indemnisations et des risques) - directions juridiques (estimation des risques) Objectiver certaines décisions de justice (par exemple, les décisions du conseil de prud'hommes qui seraient prises par des magistrats non professionnels) Utile sur la dimension technique du droit (plus de légitimité dans certains domaines comme le civil, commercial et administratif)

Mais plusieurs craintes ont aussi été exprimées qui recourent celles identifiées dans d'autres secteurs. Ainsi, par exemple, certains interviewés se sont montrés préoccupés par l'automatisation des tâches qui sont le plus « facilement » automatisables (par exemple, celles qui reposent sur l'application de règles et/ou sont répétitives). Ces tâches composent une bonne partie de certains métiers juridiques ; ces derniers pourraient ainsi être menacés de disparition. Cependant, cette vision est contestée par d'autres, qui estiment que les métiers du droit reposent sur de nombreuses compétences et dimensions humaines complexes, qui vont au-delà de la simple application de règles, ce qui questionne une nouvelle fois la définition de la valeur accordée à une tâche. Au-delà de cette divergence de point de vue, l'ensemble des interviewés estime que la transformation et l'évolution de ces métiers est inéluctable du fait du développement de l'IA. Un autre point est que cette évolution entraînerait l'émergence d'une situation concurrentielle entre les professionnels qui utilisent ou n'utilisent pas de SIA, ce qui est également évoqué dans d'autres secteurs, comme la radiologie.

Plusieurs bénéfices potentiels mais aussi les risques évoqués sont résumés dans le tableau ci-dessous :

Tableau 4. Bénéfices potentiels et risques perçus par les interviewés dans le domaine juridique

Bénéfices potentiels	Risques perçus
- Amélioration de la productivité	- Risque pour l'apprentissage, le développement de l'expérience et des compétences
- Gain de temps sur des tâches dites « simples » et « répétitives »	- Risque de perte de compétences
- Amélioration de la qualité du travail : Meilleure construction de la stratégie juridique (notamment pour les novices), Meilleur conseil avec des éléments « objectivés » et plus d'accompagnement du client, Meilleures analyses juridiques, Gains financiers, Accessibilité des données juridiques pour une meilleure compréhension des décisions judiciaires	- Risque de « profilage » des magistrats, c'est-à-dire la mise en visibilité des pratiques individuelles - Risque pour l'indépendance et l'impartialité des magistrats - Risque de sur-confiance et de prescription - Risque de standardisation des décisions de justice - Risque d'erreurs engageant la responsabilité et la remise en question des compétences

L'étude a également permis de recueillir des retours sur les systèmes de justice dite « prédictive » qui ont été expérimentés par certains avocats et magistrats interviewés. Ces systèmes se basent sur l'analyse des données de décisions judiciaires dans le but de fournir une évaluation de l'incertitude juridique en utilisant des probabilités. Ils visent à « prédire » l'issue d'une affaire judiciaire ou à estimer les montants potentiels des indemnisations.

Certains avocats qui ont testé ce type de systèmes espéraient économiser du temps en les utilisant, tandis que d'autres espéraient être surpris en découvrant des précédents juridiques spécifiques qu'ils n'auraient pas trouvés sans l'aide de ces systèmes. Les magistrats, quant à eux, cherchaient à évaluer les performances de ces systèmes dans l'analyse de la jurisprudence et à améliorer la qualité des décisions juridiques.

Au-delà de ces attentes, les avocats et magistrats interrogés expriment certaines critiques, principalement en ce qui concerne l'écart entre les avantages théoriques vantés par les entreprises vendant ces systèmes et les avantages réels pour les professionnels du droit. Malgré les avantages théoriques promis, tels qu'un gain de temps, une productivité accrue et la découverte de nouvelles informations utiles, les professionnels ayant testé ces systèmes constatent peu ou pas de véritables avantages. Ils se retrouvent ainsi confrontés à une réalité « décevante » par rapport aux promesses formulées. Dans la plupart des cas, ces systèmes semblent simplement fournir des informations et des éléments déjà connus par ces professionnels du droit : « *ça n'apportait absolument aucune plus-value. Ça ne sortait rien, ça ne m'apportait rien que je ne savais déjà [...] ça fait 12 ans que je suis avocat, j'avais probablement plus de données intéressantes* » (Avocat 5). Ces systèmes sont ainsi considérés comme des « gadgets » par certains avocats. Cette critique émane d'avocats spécialisés ayant une grande expérience, ce qui peut suggérer un lien entre l'inutilité de ces systèmes et le niveau d'expertise des professionnels. La déception face aux promesses des fonctionnalités prédictives des SIA reflète la confrontation entre les promesses et la réalité de leurs apports.

Une autre limite majeure de ces systèmes serait leur incapacité à incorporer les connaissances fines et l'expertise des professionnels du droit. De plus, les facteurs contextuels et humains, tels que les dimensions émotionnelles ou les considérations affectives, ne semblent pas être intégrés par les systèmes de justice dite prédictive : « *On ne peut pas se fier toujours qu'aux chiffres, je veux dire, il faut une analyse et encore une fois c'est de l'humain. Il y a tellement de variables. Les souffrances endurées, c'est un bon exemple parce que quand on dit un 4/7 [l'évaluation du dommage pour l'estimation du montant de l'indemnisation], ça peut être une amputation, ça peut être une fracture de jambe, ça peut être un viol, tout ça c'est différent. Parfois, il y a des experts qui mettent 6/7 parce que la victime a eu conscience qu'elle allait mourir, c'est-à-dire qu'on est dans un accident de la circulation classique, elle a un grave accident et elle a conscience qu'elle va mourir et elle finit par mourir, donc ils disent : « Elle a souffert moralement, on lui met 6/7 », mais à combien vous indemnisez la conscience que vous allez mourir ? Vous ne pouvez pas et l'intelligence artificielle a beau faire ce qu'elle veut, ça reste de l'humain, profondément de l'humain.* » (Avocat 5). Ainsi, ces systèmes ne pourraient pas appréhender pleinement les éléments nécessaires à une interprétation précise et exhaustive des décisions de justice.

A Retenir...

- Dans l'ensemble, les études examinées montrent que les travailleurs ont globalement des points de vue plutôt « réalistes » et nuancés sur l'IA :
 - Ils projettent des apports potentiels de l'IA, parfois en lien avec les discours de promotions de l'IA par exemple l'automatisation de tâches répétitives, le gain de temps, une amélioration de la prise de décision. La valeur a priori des tâches automatisées par l'IA resterait cependant à questionner et à confronter à une valeur en action, basées sur des expériences concrètes.
 - Ils identifient de nombreux risques et limites concernant la performance du travail (qualité, fiabilité) mais aussi la santé : perte d'emploi, opacité des systèmes, perte d'expertise et d'autonomie ; perte du sens du travail ; biais d'automatisation - sur confiance, pertinence, éthique des résultats de l'IA.
 - Ils s'interrogent sur l'utilité « réelle » des systèmes qui sont proposés dans les secteurs professionnels qui les concernent. Sur ce dernier constat, il convient de noter que ce questionnement de l'utilité réelle n'est pas seulement une projection a priori mais repose sur des expériences vécues de certains des acteurs interrogés comme dans l'étude de Gamkrelidze (2022).
- Concernant la « prévention » des risques de dégradation de la performance du travail réalisé avec un SIA, plusieurs éléments sont soulignés :
 - Le besoin de transparence sur les apports et les limites de l'IA (ce qu'elle permet de faire et de ne pas faire) ;
 - Le besoin de compréhension des résultats fournis par l'IA (explicabilité) ;
 - La prévention d'un risque de normalisation/standardisation des raisonnements mis en œuvre par les professionnels ;
 - La nécessité d'un pouvoir de contrôle sur l'IA en lien avec des questions de fiabilité du travail, de responsabilité et d'indépendances des actions entreprises à l'aide des résultats produits par une IA ;
 - La nécessité de réflexions sur les compétences à agir avec une IA qui concerne à la fois des démarches de développement de nouvelles compétences et de remobilisation d'expériences et de savoir-faire,
 - Des réflexions la sur-confiance/la méfiance envers l'IA.
- Ces résultats renforcent l'importance, pour aller plus loin, de la compréhension fine de l'activité dans ses dimensions physiques, cognitives et sociales – afin d'identifier et d'évaluer les effets transformatifs des SIA, et de soutenir le développement de nouvelles façons de faire et/ou la mobilisation d'expériences passées, ce qui est souvent un impensé de l'introduction des systèmes techniques dans les environnements de travail (cf. section 5.3).

5.2. L'IA à l'épreuve des expérimentations : une performance accrue dans les tâches réalisées à l'aide de l'IA, mais une performance à pondérer

L'amélioration de la « performance » entendue au sens d'efficacité (par exemple dans la détection et le traitement de maladies, la prédiction de pannes, l'optimisation des processus de travail, le gain de temps) sont au centre des discours qui promeuvent l'IA. Beaucoup d'études expérimentales ont ainsi été réalisées pour montrer les gains de performance qui pourraient être obtenus grâce à l'IA (y compris les IA génératives) soit (1) en comparant l'efficacité de l'humain seul à celle de l'IA seule, soit (2) en comparant l'efficacité d'un humain assisté d'une IA et d'un humain sans IA.

Ces études aboutissent souvent à la conclusion que l'IA est « meilleure » que le travailleur humain, en tout cas dans ces conditions expérimentales. Des pistes de réflexion sont aussi identifiables sur la manière de « combiner » l'humain et l'IA dans une logique de complémentarité et non pas de substituabilité.

Nous reprenons ici les études expérimentales sur l'efficacité de l'IA ou d'un couple Humain-IA dans l'aide au diagnostic médical lié à l'imagerie médicale vue comme un modèle – un cas d'école – des performances associées à l'IA en situation expérimentale²⁷. Puis nous présentons une synthèse des travaux s'intéressant à la performance des tâches réalisées avec de l'IA générative dans différents secteurs professionnels, avant de conclure sur les limites de ces études.

5.2.1. Des diagnostics d'imagerie accrus dépendants du niveau d'expertise et des caractéristiques techniques de l'IA

Les études que nous choisissons de reprendre ici proposent toutes des SIA visant l'aide à la prise de décision diagnostique dans le cas: de cancers du sein (Calisto et al., 2022) ou de la peau (Tschandl et al., 2020), d'examen optique de lésions colorectales (Reverberi et al., 2022), d'échographies mammaires (Park et al., 2019) et enfin de détection de fracture (Oppenheimer et al., 2023). Elles s'intéressent le plus souvent à l'efficacité d'un *couple humain seul + IA* vs. Un humain, et s'appuient sur des expérimentations où l'on demande à plusieurs médecins seuls de poser un diagnostic sur la base d'images et à ces mêmes médecins de poser un diagnostic avec l'aide d'une IA « en second lecteur », c'est-à-dire après un premier diagnostic de l'humain. La *valeur a priori* de ce type de systèmes d'aide au diagnostic serait d'améliorer la précision dans la détection de lésions/pathologies, voir la rapidité, ce qui permettrait de réduire la charge de travail des radiologues tout en fiabilisant le diagnostic.

Ces études concluent toutes à une amélioration de la performance du couple « humain seul + IA » comparée à celle de l'IA seule en termes de précision des diagnostics et/ou de temps mis pour établir un diagnostic. Par exemple, l'étude de Calisto et al. (2022) sur la classification et la segmentation de

²⁷ Hors du secteur médical mais toujours dans l'aide au diagnostic, l'étude de (Geex, 2018) réalisée auprès d'avocats pour repérer des problèmes dans les textes des accords de confidentialité. Elle met en avant la supériorité du SIA par rapport à un avocat pour l'identification des problèmes (94% de taux de précision pour le SIA contre 85% pour les avocats).

cancer du sein montre qu'en situation expérimentale l'utilisation de l'IA réduit le temps (moins 3 minutes par patient pour poser un diagnostic avec le SIA) et un effet significatif de la diminution d'erreurs de diagnostic (diminution du pourcentage des faux positifs et négatifs). L'étude de Oppenheimer et al. (2023), réalisée auprès d'internes en radiologie apporte néanmoins des nuances à ce résultat. Les résultats montrent une augmentation de la capacité des internes assistés par l'IA à détecter les maladies (augmentation de la sensibilité à la détection de fractures), sans qu'ils perdent la spécificité dans leur interprétation (détection des résultats négatifs i.e. de l'absence de fracture – vrais négatifs)²⁸. Si la sensibilité de l'IA seule était légèrement supérieure à celle des internes, sa spécificité (détection des vrais négatifs) restait bien inférieure. Ceci appellerait donc des scénarios qui vont plus vers des couples humain-IA (complémentarité entre l'humain et la machine) que vers l'automatisation complète (substitution de l'humain par la machine).

Cependant, les performances de diagnostic améliorées d'un couple « humain seul + IA » varient suivant le niveau d'expertise des médecins – ces bénéfices seraient supérieurs pour les novices (Reverberi et al., 2022 ; Tschandl et al., 2020)), ou les caractéristiques techniques du système. Ces « bénéfices » sont dépendants du niveau de fiabilité du système : une IA défectueuse peut au contraire induire en erreur l'ensemble des médecins, y compris les experts (Tschandl et al., 2020). Ils peuvent être compromis par la sur-dépendance ou la sous-dépendance envers l'IA, une sur-confiance ou une méfiance envers ces résultats, en lien notamment avec l'opacité du SIA (Reverberi et al., 2022). En effet, ces auteurs mettent en avant qu'en étant capables de juger quand ils pouvaient avoir confiance ou se méfier de l'IA, les médecins pourraient mobiliser leurs connaissances au mieux et ainsi améliorer les résultats de la décision, notamment dans les situations où l'IA risque de ne pas fonctionner correctement. Ceci suppose que la conception du SIA propose des éléments permettant de construire une distance critique par rapport à la fiabilité des résultats de l'IA.

5.2.2. *L'IA générative : une source d'amélioration de la performance au travail ?*

L'apparition de l'IA générative a suscité de nombreuses interrogations concernant ce qu'elle pourrait apporter en contexte de travail, notamment en termes de performance. De nombreuses études empiriques ont ainsi été menées sur ce sujet. Plusieurs de ces études qui ont été réalisées dans différents domaines comme la programmation (Peng et al., 2023), le secteur des services (Dell'Acqua et al., 2023 ; Brynjolfsson et al., 2023) ou les tâches de rédaction (Noy and Zhang, 2023) montrent une amélioration de la performance dans la réalisation de tâches relatives à ces domaines (en termes d'efficacité et de productivité). Par exemple, l'étude de Dell'Acqua et al. (2023) a consisté à faire utiliser GPT4 par des employés (plus de 700) d'un cabinet de consultants (Boston Consulting Group) dans un cadre expérimental sur deux tâches supposées réalistes, c'est-à-dire représentatives du travail habituellement réalisé par ces employés : conceptualisation et développement de produits, résolution

²⁸ La sensibilité d'un algorithme fait référence à sa capacité à détecter correctement les résultats positifs, tandis que la spécificité correspond à sa capacité à identifier correctement les résultats négatifs.

de problèmes. Trois configurations ont été testées : sans IA, avec IA, avec IA incluant une présentation globale du système ("*overview*") permettant au salarié de se familiariser avec à GPT4, en particulier la manière de formuler des questions. Les résultats montrent que les performances (qualité des productions, temps de réalisation des tâches) sont meilleures avec l'IA (gain de plus de 45% en termes de qualité), la qualité des productions étant évaluée par des experts des domaines de tâches sur différents critères (créativité, qualité de l'écrit, pensée analytique, pouvoir de persuasion). Ces performances sont même supérieures pour les consultants qui ont bénéficié de la présentation de GPT. Cette étude comme celle de Brynjolfsson et al. (2023) dans un contexte serviciel (utilisation de GPT par les assistants de clientèle d'une entreprise informatique) montrent que les améliorations constatées semblent bénéficier surtout aux travailleurs moins qualifiés, sans que leur niveau d'expertise ne soit évoqué. Selon Brynjolfsson et al. (2023), une explication possible est que les modèles d'IA diffuseraient les savoirs tacites des experts lorsqu'ils l'utilisent, ce qui profiteraient aux travailleurs moins qualifiés, considérés ici comme moins experts. Cet « effet » de l'usage d'IA générative comme GPT est supposé potentiellement réduire les inégalités entre travailleurs d'un groupe professionnel de différents niveaux d'expertise ou de qualification du point de vue de la performance. Une récente remet cependant cette hypothèse question. Dans le cadre d'une large étude expérimentale de même nature que celles mentionnées, c'est-à-dire comparant la performance de personnes avec (500) ou sans (500) GPT sur des tâches représentatives d'activités professionnelles (par exemple, rédaction de courriels professionnels, tâche de compréhension de textes), Haslberger et al. (2023) identifient également une amélioration globale de la performance « grâce » à l'IA, mais constatent des inégalités. Ils observent des inégalités entre les jeunes et les plus âgés, l'IA bénéficiant beaucoup plus aux premiers. Les auteurs en concluent qu'à mesure que l'IA pénètre les lieux de travail, des actions de formation/accompagnement des salariés plus âgés seront importantes. Mais il convient de noter que les tâches utilisées dans cette étude n'étaient pas celles réalisées au quotidien dans leur travail par les participants, elles avaient été fabriquées en les supposant comme représentatives de tâches professionnelles. Ce qui questionne la validité écologique de cette étude.

Cependant, des effets négatifs voire aucun effet sur la performance sont également mis en avant. L'étude de Dell'Acqua et al. (2023) met en avant une dégradation de la performance (augmentation du nombre d'erreurs) lorsque les tâches sont plus difficilement réalisables de façon directe avec l'IA (GPT) par exemple difficulté de formulation des « invit » (prompts) permettant d'obtenir des réponses du système ou les risques de diminution de la créativité / d'uniformisation lorsque tous les travailleurs utilisent un même système qui produit des réponses similaires. D'après les auteurs cela peut être dû à la tendance à faire trop confiance à l'IA affectant négativement la performance dans les situations où elle n'est pas la plus adaptée ou à une utilisation des sorties de l'IA sans distance critique, ce qui rejoint les résultats mis en avant dans le secteur médical.

Comme pour les SIA d'aide à la décision diagnostique, ceci met en évidence le fait que l'amélioration de la performance liée à l'utilisation d'IA générative n'est pas systématique, elle peut varier selon les tâches, le niveau d'expertise ou les caractéristiques du système.

5.2.3. *Limites des résultats expérimentaux*

Les résultats des études centrées sur la performance fournissent des indications intéressantes sur les effets et le potentiel des SIA pour améliorer la performance des travailleurs (précision ou temps) dans différents contextes. Des pistes de réflexion sont aussi identifiables sur la manière de « combiner » l'humain et l'IA dans une logique de complémentarité et non pas de substituabilité. Dans cette perspective, les études montrent différentes transformations possibles des tâches de travail impliquées par l'inclusion de l'IA.

Plusieurs éléments appellent à prendre ces résultats avec prudence quant à leur transposition en situation de travail réelle.

Une première limite renvoie à la notion de performance telle que mobilisée par ces études, l'évaluation de la performance est réduite ici à un pan technique du travail et une tâche en particulier – p.ex. la prise de décision diagnostique souvent sur la base d'images – et non l'ensemble de la relation de soin dans laquelle le diagnostic n'est qu'une étape. Ces études se concentrent sur la performance dans des conditions bien définies et dans une acception réduite synonyme d'efficacité ou d'efficience excluant d'emblée d'autres éléments de la performance au travail : par exemple les conséquences sur la santé des professionnels et le sens du travail (vu ici comme un critère de performance), en lien avec un risque perte d'autonomie décisionnelle introduite par l'IA.

Une seconde limite tient au fait que la performance de plusieurs humains (par exemple deux experts, ou couple expert-novices ou des groupes de médecin) ne semble pas être évaluée expérimentalement alors même que le travail collectif humain, les possibilités de débats et de concertation sont des facteurs de performance. On pourrait alors en conclure que ces études incorporent un « biais organisationnel » puisque que des scénarios impliquant plusieurs humains agissant collectivement vs IA seuls ou collectifs humains - IA ne semblent pas testés.

Une troisième limite en lien avec la perspective située et systémique que nous adoptons, cette efficacité accrue de l'AI ou du couple Humain+IA en situation expérimentale ne dit rien de l'efficacité réelle de l'IA introduite dans la complexité de situations de travail où les choix organisationnels liés à l'introduction de l'IA (par exemple travail en effectif réduit ou sous contraintes de temps, empêchement du travail collectif) peuvent venir au contraire « faire baisser » cette performance « globale » en termes de santé, de qualité et de fiabilité. Nous reviendrons sur ces éléments dans la section suivante dans le « zoom » sur l'introduction de l'IA en radiologie. Qui plus est la performance n'est qu'un aspect du travail. Par conséquent, elle ne suffit pas à caractériser pleinement les transformations effectives des activités de travail liées à l'IA. En effet, la réalisation d'une activité de travail est influencée par une dynamique complexe d'interactions entre facteurs « individuels », tels que les caractéristiques des travailleurs, et facteurs externes, tels que les caractéristiques de la tâche à accomplir et l'organisation du travail, sans oublier les interactions avec les collègues de travail. Compte tenu de cette complexité dynamique, les études basées uniquement sur des situations expérimentales et des tâches spécifiques sont insuffisantes pour comprendre de façon holistique les effets du déploiement et de l'usage de l'IA sur le travail. Elles ne permettent pas d'aborder comment l'IA s'intègre réellement dans les pratiques

individuelles et collectives ou comment elle modifie la division du travail, les rôles, les compétences, les relations et les responsabilités des travailleurs.

A Retenir...

- Plusieurs études expérimentales dans différents secteurs mettent en avant une augmentation de la performance (en termes de temps, d'efficacité, de précision, de productivité, etc.) dans différents types de tâches (diagnostic médical, résolution de problème, production de textes ou de programmes) réalisés à l'aide de l'IA.
- Cependant, cette performance accrue n'est pas systématique et universelle :
 - Elle bénéficie surtout aux travailleurs les moins experts ;
 - Elle est meilleure lorsqu'elle est précédée de « tutoriels » ;
 - Elle ne concerne que des tâches qui sont « à la portée de l'IA » ;
 - Elle est dépendante du niveau de fiabilité du système et de ses possibilités de soutenir la construction d'une distance critique vis-à-vis de ce qu'il produit.
- En conséquence, et dans certaines situations, il peut y avoir des effets négatifs voir pas d'effet positif sur la performance humaine ou des risques de diminution de la créativité / d'uniformisation de la pensée et des produits de l'action et donc une baisse de performance qualitative.
- Les études expérimentales présentent des limites qui amènent à considérer avec prudence leurs résultats en situations de travail réelles : vision de la performance réduite à une performance technique pour certaines tâches du travail et excluant les effets sur la santé des professionnels, absence de prise en compte du rôle de la dimension collective et systémique du travail sur la performance.

5.3. Usages de l'IA en situation de travail réel : conséquences sur les professionnels et leurs activités

Les limites précédemment établies appellent à rechercher des travaux portant sur des SIA mis en place dans des situations réelles de travail et permettent d'identifier les transformations du travail et organisationnelles qu'entraînent l'introduction et l'usage de ces systèmes, ainsi que les conséquences sur les travailleurs et leurs activités. Nous reprenons ici plusieurs d'entre elles particulièrement illustratives :

- dans le milieu médical toujours en lien avec l'introduction de SIA d'aide au diagnostic – détection d'anomalie au thorax notamment par des internes (Anichini & Geffroy, 2021), diagnostic de cancers

- du sein et du poumon et détermination de l'âge osseux (Lebovitz et al., 2023) – ou système d'alerte - identification et traitement de la septicémie pour des médecins et infirmiers (Henry et al., 2022),
- mais également les systèmes de recommandations pour les Ressources Humaines – p.ex. recommandations des postes à des salariés d'une grande entreprise pour favoriser la mobilité interne (Lévy, 2019) ;
 - et enfin l'introduction de chatbots dans différents secteurs qui jusqu'à l'arrivée des IA génératives représentaient près de 60% des IAs en situations de travail - Chatbot juridique, Chatbot RH pour l'accompagnement des salariés sur les questions des ressources humaines et Chatbot R&D destiné à « capitaliser les connaissances » d'experts (Gras Gentiletti et al., 2022), Chatbot de relation clientèle à la réception d'un hôtel (Flandrin et al., 2022) ; Chatbot marketing (Wang et al., 2023).

Dans ces résultats, on retrouve certains des éléments concernant les apports et les risques des SIA décrits dans les études de la section 5.1 par exemple concernant les écarts entre les valeurs projetées *a priori* des SIA pour assister certaines tâches et leurs remises en question par la plus grande richesse des activités, les risques pour le développement de l'expertise des professionnels, le risque de standardisation des pratiques, de fiabilité de l'action et de sa dépendance à l'explicabilité des systèmes - mais en les étayant de l'expérience concrète des acteurs et en mettant en avant la dépendance organisationnelle des usages : pression temporelle, augmentation de la charge de travail dans des contextes déjà contraints, autonomie dans la décision d'utiliser ou non un SIA, entrave potentielle à la coopération entre professionnels. Par ailleurs, différents modes d'usages sont identifiés- non-usage dans l'activité faute de pertinence du système ou de coût temporel associé compte tenu de l'intensité du travail, division du travail IA-Humains en fonction des « capacités » de l'humain ou de la machine, « augmentation » effective de l'activité, augmentation effective de l'activité avec développement de pratiques réflexives soutenant une distance critique par rapport aux résultats produits par l'IA.

5.3.1. Vers une vision élargie de la valeur et de la performance d'une activité

Anichini & Geffroy (2021) ont étudié l'utilisation réelle d'un SIA pour l'aide à la détection d'anomalies sur des radiographies du thorax, en particulier par des radiologues internes. Les résultats montrent la complexité de l'activité de diagnostic des radiologies qui implique une activité d'interprétation des radiologues, qui va au-delà d'un « simple » repérage des anomalies sur des images tel que soutenu par les SIA. Cette activité repose notamment sur l'utilisation d'autres sources d'informations sur les patients pour poser un diagnostic « de qualité », et c'est ce qui en fait sa fiabilité et sa valeur au-delà de la détection de l'anomalie²⁹. Autrement dit contrairement à l'IA, l'interprétation du radiologue ne définit pas les images comme normales ou pathologiques par la simple présence ou absence de lésions. Cette vision étendue du diagnostic peut donc entraîner des divergences entre le radiologue et l'IA dans l'identification de ce qui est normal ou pathologique, puisque l'IA est aveugle à ces informations. L'évaluation de l'efficacité du système par les médecins – et donc leur prise de distance critique par

²⁹ Ceci rejoint en ce sens la discussion sur la « tâche d'écriture » des médecins généraliste évoquée en section 5.1 et de sa valeur interprétative

rapport à ses résultats - passe notamment par l'évaluation de la pertinence des anomalies détectées par le système pour différencier le normal du pathologique, si tant est que les fonctionnalités techniques du système le permettent.

5.3.2. *Des SIA augmentant ou entravant l'apprentissage et le développement de l'expertise*

La même étude (Anichini & Geffroy, 2021) met en avant le rôle du système sur les processus d'apprentissage des internes. D'une part, le système peut aider les internes à confirmer leur interprétation et à diminuer les demandes d'examens supplémentaires pour assurer leurs diagnostics. Ceci est également retrouvé chez Henry et al. (2022) où les professionnels de santé sont préoccupés par le risque d'une dépendance excessive au système, entraînant la dégradation des compétences cliniques. Des résultats similaires se retrouvent chez Lebovitz et al. (2023) dans leur étude de trois services de radiologie dans un grand hôpital américain où des outils d'IA étaient utilisés pour le diagnostic de cancers du sein et du poumon et pour la détermination de l'âge osseux. L'usage du SIA d'aide au diagnostic pourrait entraver leurs processus d'apprentissage et de développement des compétences qui visent à développer les capacités de différenciation du normal et du pathologique en adoptant un point de vue holistique au-delà de la détection d'une lésion. Certains internes apprenants expriment en effet le fait que le système peut les faire douter sur leurs propres compétences, en particulier lorsqu'il montre des anomalies qu'ils n'ont pas vu, ce qui peut altérer le processus d'apprentissage. Ceci est aggravé par l'élimination des images classées comme "normales" par le système qui ne peuvent être analysées par les internes alors qu'elles sont une source de développement de connaissances. Ainsi, l'utilisation du système, en particulier quand celle-ci intervient au début du processus d'apprentissage de la pratique radiologique, peut altérer négativement l'acquisition des compétences des internes en formation. Ainsi, les internes ne consultent l'avis de l'IA qu'après avoir regardé les images pour éviter que le système perturbe la construction de leurs « routines visuelles d'analyse » des radiographies – construites au fil de l'expérience. On peut donc faire l'hypothèse que des choix organisationnels d'introduction de l'IA qui imposeraient son usage en « premier lecteur » (dans des contextes de pression temporelle, de manque d'effectifs, ou parce que les procédures ou une réglementation l'imposeraient) pourraient venir entraver ce processus d'apprentissage pourtant nécessaire au développement de l'expertise et faire courir un risque pour la performance de l'action.

5.3.3. *Dualité entre opacité/explicabilité des systèmes, fiabilité et pertinence de l'action et dépendance organisationnelle des usages de l'IA*

Plusieurs études révèlent : (1) les conséquences de l'opacité des SIA – de leur manque d'explicité - sur la fiabilité du travail (p.ex. risque de standardisation des soins entraînant un risque pour la prise en charge singulière d'un patient), (2) leur pertinence ou non en situation et (3) la diversité d'usages de ces systèmes en fonction des marges de manœuvre organisationnelle (pression temporelle, autonomie ou non vis à vis du choix de l'utilisation de l'IA) et de l'intensité du travail.

L'étude d'Anichini & Geffroy (2021) souligne que l'opacité du système implique une prise de risque pour les professionnels (accepter des résultats sans tout comprendre) et renvoie donc une question de fiabilité de l'action, lorsque le radiologue se fie aux résultats du système dont le fonctionnement n'est pas toujours complètement compréhensible. On retrouve ici le problème de l'opacité déjà évoqué qui introduit de l'incertitude dans l'action et questionne la responsabilité médicale et la fiabilité de la décision. Ceci se retrouve également, chez Lebovitz et al. (2023) dans leur étude au sein de trois services de radiologie. L'opacité a eu pour effet d'accroître l'incertitude des radiologues, car les résultats des outils d'IA divergeaient souvent de leur jugement initial dans l'interprétation des radiographies sans disposer d'explications de ces résultats.

Des risques de standardisation des soins sont également évoqués. Anichini & Geffroy (2021) soulignent qu'avec l'IA se joue le risque d'une standardisation du travail des radiologues à partir du moment où elle ne laisse plus de place à la singularité et la variabilité des pratiques qui sont essentielles dans le métier de radiologue. L'étude de Henry et al. (2022) examine la mise en place et l'utilisation d'un système pour aider les professionnels de santé (les infirmiers et les médecins) à identifier et à traiter la septicémie. Ils évoquent ce risque de standardisation en lien avec des choix organisationnels issus des organismes de réglementation qui pourraient contraindre les cliniciens à appliquer les recommandations du système.

Lebovitz et al. (2023) montrent eux aussi que l'opacité de l'IA constitue un problème important pour les professionnels, mais qu'ils peuvent développer des stratégies pour y faire face en fonction des contextes. Dans un seul des trois départements étudiés, les professionnels ont systématiquement intégré les résultats de l'IA dans leurs jugements finaux, mettant en œuvre ce que les auteurs appellent une « augmentation engagée » (*engaged augmentation*). Ces professionnels ont développé des « pratiques interrogatives » de l'IA, c'est-à-dire des pratiques mises en œuvre pour établir un lien entre leurs propres connaissances et celles de l'IA. Cependant, ces pratiques sont dépendantes des marges de manœuvre laissées ou non aux professionnels compte tenu des contextes organisationnels (p.ex. pression temporelle, contrainte à se servir de l'IA ou non) : elles ont un coût car elles exigent du temps et la réalisation d'analyses supplémentaires, ce qui est problématique dans les situations de forte charge de travail qui sont habituelles en radiologie. Les radiologues des deux autres départements n'ont pas adopté de telles pratiques et n'ont pas intégré les données de l'IA dans leurs décisions finales. Pour les auteurs, il s'agit d'une forme "d'augmentation non engagée"³⁰ puisque les résultats de l'IA sont ignorés par manque de pertinence pratique pour les professionnels. Ceci s'explique d'une part par la difficulté qu'ont eu les radiologues de ces départements à relier leur expertise aux résultats du système utilisé (par exemple, le mode de calcul de l'âge osseux) et d'autre part à l'absence de valeur ajoutée réelle de l'IA pour ces radiologues, qui ajoute des tâches supplémentaires de vérification et de compréhension des résultats qui nécessitent trop de temps compte tenu de leur charge de travail.

³⁰ Nous reprenons ici les propositions des auteurs, bien qu'il ne s'agisse pas d'un problème d'engagement ou non des professionnels, mais bien de pertinence.

La compréhension limitée du fonctionnement des SIA et la façon dont les professionnels s'en emparent est également évoquée dans les travaux de Henry et al. (2022). La valeur *a priori* du système étudié et qui alerte en cas d'un changement dans l'état du patient serait d'aider à la priorisation des patients dans le service des urgences dans un contexte de surcharge informationnelle et d'intensification. Les auteurs constatent dans un premier temps que l'opacité du système ne constitue pas nécessairement un obstacle à son utilisation parce que les professionnels semblent « s'en accommoder » - autrement dit y voir une utilité et/ou une pertinence - dans leur contexte d'action contraint. Les médecins utilisent le système comme un système d'assistance au diagnostic, tout en étant conscients que ce système ne dispose pas de leur expertise clinique. Pour les infirmiers et les médecins, il s'agit d'une « deuxième paire d'yeux » permettant de faire une surveillance continue. Les auteurs constatent que la confiance dans le système se construit de différentes manières : en observant le fonctionnement du système dans différentes situations d'usage mais aussi en prenant en compte des études externes et des retours d'expérience des collègues. Les échanges avec les concepteurs jouent également un rôle bénéfique pour l'appropriation du système.

Dans un tout autre domaine, celui des ressources humaines, Lévy (2019) met en évidence les difficultés rencontrées par des conseillers RH dans l'usage d'un SIA dont l'objectif était de favoriser la mobilité interne en recommandant des postes à des salariés d'une grande entreprise à travers le croisement des profils des salariés et des offres de poste en interne. L'étude montre le peu d'utilisation du système. Contrairement à l'utilité et l'usage projetés, ce système est davantage devenu un outil de réflexion pour les conseillers en pointant le manque de pertinence des propositions des postes aux salariés et en les amenant à réfléchir sur leurs critères de recommandation. Si au départ les services RH manifestaient un certain intérêt envers ces SIA qui étaient *a priori* envisagés comme inévitables dans leur métier, la confrontation à l'utilisation du système les a conduits à le délaisser car il manquait de pertinence pratique. Une fois de plus, l'opacité du fonctionnement du système semble au moins en partie expliquer leur faible utilisation.

Cette diversité d'usage en fonction des contextes organisationnels et des possibilités pour les professionnels de développer des stratégies pour faire face aux limites de l'IA peuvent être rapprochés des observations de l'étude semi-expérimentale de Dell'Acqua et al. (2023) qui identifiait un usage de type "*centaure*" avec une division du travail entre humains et IA et distribution des tâches selon les capacités de chacun ; et un autre de type "*cyborg*" - un travail en tandem avec l'IA aux frontières de ses capacités, Lebovitz et al. (2023) y ajoutant la nécessité de développement d'activités réflexives dans la configuration « cyborg », et l'absence d'intégration des résultats de l'IA dans l'activité faute de pertinence ou de coûts de contrôle des résultats de l'IA trop élevé (appelé de façon approximative « augmentation non engagée »), retrouvée chez Lebovitz et al. (2023) et Lévy (2019).

Enfin, on constate le développement d'usages informels – parfois cachés- de systèmes d'IA génératives en situations et l'émergence de règles organisationnelles visant à en interdire l'utilisation. D'après une récente étude (étude BlackBerry 2023) ; 75% des 2000 entreprises interrogées (responsables « Information Technology ») mettaient en place ou envisageaient cette interdiction. D'autres cherchent à mettre en place un accès sécurisé, et enfin d'autres laissent faire.

5.3.4. *Le cas des chatbots : entre intensification du travail ? et ses entraves à l'interaction directe entre « clients » et experts*

Une dernière catégorie d'études s'intéresse au cas emblématique des chatbots utilisés le plus souvent pour assister les tâches de relations de service avec des clients ou entre différentes entités d'une entreprise. Comme pour les systèmes précédents, l'introduction des chatbots n'est pas sans effet sur l'intensification du travail, mais constitue également une entrave à la coopération directe entre des experts et leurs clients (qu'ils soient internes ou externes à l'entreprise ou l'institution).

Gras Gentiletti et al. (2022) ont étudié trois types de chatbots déployés dans une grande entreprise (Chatbot juridique destiné aux connaissances juridiques, Chatbot RH pour l'accompagnement des salariés sur les questions des ressources humaines et Chatbot R&D destiné à capitaliser les connaissances des experts). Cette étude met en lumière le manque d'adaptation de ces systèmes aux contextes d'utilisation et aux environnements dans lesquels ils sont mis en place. L'introduction des chatbots limiterait également l'accès direct aux collègues (par exemple, les juristes) alors que l'étude met en évidence la nécessité pour les professionnels de l'entreprise de pouvoir interagir avec des experts humains car les chatbots ne disposent pas de toutes les connaissances et l'expérience permettant de répondre à la diversité et à la complexité des demandes. Cependant, les choix organisationnels (procédures de travail relayées par la direction) incitent fortement les juristes à renvoyer les salariés vers le chatbot lorsque la réponse peut être donnée par ce dernier. Les professionnels « clients » du service équipé du chatbot essaient donc d'utiliser les chatbots pour obtenir un contact direct avec les juristes qu'ils recherchent en formulant explicitement leurs demandes. Ils utilisent également d'autres moyens, tels que les forums d'entreprise, pour contacter les experts. Ainsi, ils cherchent d'autres ressources que les chatbots pour obtenir l'aide dont ils ont besoin dans leur activité. Par ailleurs, les résultats ont aussi montré un travail supplémentaire pour les experts (comme les juristes), induit par les chatbots car ils nécessitent d'être alimentés en données structurées par des humains. Les SIA mis en place accroissent ainsi la charge de travail de certains salariés, cette tâche supplémentaire s'ajoutant à leur travail. L'étude de Flandrin et al. (2022) sur un chatbot mis en place dans le cadre du travail des réceptionnistes d'un hôtel avec l'objectif d'améliorer les conditions de travail à la réception identifie des difficultés similaires. Elle montre que la mise en place du chatbot ne réduit pas la charge de travail des réceptionnistes mais plutôt crée de nouvelles exigences et induit de nouvelles tâches : les réceptionnistes doivent par exemple paramétrer mais aussi veiller au bon fonctionnement du système.

A contrario, l'étude de Wang et al. (2023) constate surtout des avantages dans l'utilisation de chatbots en entreprise, plus précisément dans le secteur du marketing. Ces derniers permettraient d'augmenter l'efficacité des employés, de réduire la charge de travail, de permettre des modalités de travail plus flexibles. Les résultats de l'étude indiquent que les employés se sentent rassurés par l'utilisation d'un chatbot, car cela leur permet d'accéder à une quantité plus importante d'informations, de connaissances et de ressources pour améliorer leur performance au travail. Les participants ont perçu les chatbots comme des outils utiles pour les opérations commerciales, fiables pour atteindre des objectifs

professionnels et dignes de confiance pour la protection des données. Il convient cependant de remarquer que cette étude repose principalement sur des questionnaires, donc des éléments déclaratifs qui ne permettent pas de disposer d'informations quantitatives précises, par exemple sur l'augmentation de l'efficacité rapportée par les enquêtés.

Zoom sur les usages de SIA et ses conséquences dans le secteur de la radiologie (Gamkrelidze, 2022)

Deux SIA assistant le travail en radiologie sont ici considérés.

- **Un système de dictée à Reconnaissance Vocale (RV)** mis en place dans un service d'imagerie médicale d'un hôpital universitaire en France. Ce système comprend un logiciel de reconnaissance vocale, une interface de traitement de texte et un microphone filaire, et est intégré au système d'archivage et de transmission d'images (PACS). Le système de RV est basé sur des algorithmes d'IA pour la reconnaissance vocale et transcrit automatiquement le texte dicté. Il est conçu pour transcrire le compte rendu radiologique dicté par un radiologue, qui est le produit essentiel de leur travail afin de transmettre et de laisser une trace de l'analyse et l'interprétation des examens d'imagerie.
- **Un système de Détection des Fractures (DF)** utilisé dans les services d'imagerie et d'urgence de divers types d'organisations en France, y compris les hôpitaux généraux et les cliniques privées. Ce système utilise l'apprentissage profond pour traiter automatiquement les images de radiologie conventionnelle, communément appelées radiographies ou imagerie aux rayons X, afin de détecter les fractures osseuses. Le système a été conçu pour fournir des résultats d'analyse des images radiographiques afin d'aider à établir un diagnostic, généralement réalisé par des radiologues. Une fois les images acquises, elles sont automatiquement envoyées à un serveur pour être analysées, puis intégrées dans le dossier numérique du patient. Le radiologue peut accéder aux images d'origine ainsi qu'au rapport d'analyse du système, qui comprend des images avec les indications des résultats du système lorsqu'il détecte des fractures. Le système fournit trois types d'informations en fonction de son niveau de certitude et en indiquant les zones de fracture sur les images : absence de fracture, une ou plusieurs fractures probables et présence d'une ou plusieurs fractures. Ces informations sont les seuls éléments explicatifs fournis sur les résultats du système.

Figure 4. Photos d'un radiologue utilisant le système de dictée vocale pour établir son compte rendu (à gauche) et de l'interface du système de détection de fractures (à droite)



Méthodologie de l'étude

Des entretiens et des analyses du travail par observations ont été utilisés auprès de plusieurs professionnels (radiologues, manipulateurs, médecins urgentistes, secrétaires, informaticiens, cadre de santé).

Les SIA comme réponse au manque de ressources dans le secteur hospitalier

L'intégration des SIA en radiologie est influencée par un contexte socio-organisationnel marqué par des ressources limitées et une charge de travail élevée. La mise en place du système RV est liée à la volonté de la direction de remédier aux retards dans la transcription des comptes-rendus radiologiques réalisée par les secrétaires dans un contexte d'une baisse des effectifs. La valeur affectée *a priori* à ces système étant de les décharger de cette tâche – perçue comme sans valeur ajoutée - pour se concentrer sur les tâches administratives. Avant la mise en place du système, les radiologues dictaient les comptes-rendus radiologiques, que les secrétaires transcrivaient, relisaient et imprimaient pour validation et signature par les radiologues. Le processus était asynchrone et pouvait prendre plusieurs jours, mais la coopération entre les radiologues et les secrétaires garantissait la qualité des comptes-rendus et, par conséquent, la qualité des soins et la sécurité des patients. La mise en place de la dictée RV a nécessité de repenser cette organisation du travail. Son intégration dans une nouvelle version du PACS a alors permis de résoudre les problèmes de retards dans la transcription, mais cela n'est pas sans conséquence sur la performance du travail au sens fort (santé, qualité, fiabilité) des radiologues et des secrétaires. Dans l'exemple du système DF, il s'agissait d'une réponse à l'évolution technologique de la radiologie et de la nécessité de participer à cette évolution plutôt que d'être « *à la traîne* ». Les principaux objectifs étaient de faire face à l'augmentation du volume d'examens et au manque de radiologues, ce qui a entraîné une augmentation de la charge de travail, notamment dans les hôpitaux.

Face aux problèmes socio-organisationnels complexes, l'IA est donc souvent présentée comme une solution rapide qui peut faire gagner du temps, accroître l'efficacité et aider les travailleurs à faire face à des conditions qui se détériorent du fait de l'augmentation des contraintes budgétaires dans les hôpitaux. Cependant, la mise en place de SIA dans de telles circonstances peut être perçue comme une mesure compensatoire pour les lacunes socio-organisationnelles, semblable à une « prothèse organisationnelle ». Cette approche fait écho au « solutionnisme technologique » (Morozov, 2013) évoqué plus haut.

Faire face aux limites de l'IA : travail d'adaptation et construction de sens des résultats de l'IA

L'étude montre que les travailleurs utilisant des SIA doivent adapter leur façon de travailler pour tenir compte des limites de ces systèmes et les rendre bénéfiques pour leurs tâches. Cependant, cette adaptation peut être difficile et entraîner des coûts. L'utilisation de la dictée RV peut

entraîner des erreurs de transcription et un appauvrissement de la structure grammaticale des phrases, ce qui peut nuire à la qualité de langue et donc à la qualité du compte-rendu et *in fine* à la qualité des soins. Dépasser cet écueil implique pour les radiologues des actions supplémentaires – et donc une charge de travail supplémentaire - qui visent à comprendre les performances et les limites du système, à ajuster leurs propres activités en conséquence et à appauvrir le niveau de langue utilisé (diversité du lexique utilisé, structure grammaticale des phrases). Cet appauvrissement n'est potentiellement pas sans conséquence sur la santé des radiologues (charge accrue, sentiment d'être limité par la machine, fatigue psychologique) et sur la précision et donc la qualité des compte-rendus. Par ailleurs, la réalisation d'un compte-rendu « fiable » était auparavant le fruit d'un travail collectif entre radiologues et secrétaires médicales qui ne réalisaient pas uniquement une tâche de transcription mais bien une activité d'évaluation et de repérages des incohérences potentielles dans un compte-rendu. L'introduction du système de RV prive donc l'organisation de cette barrière de sécurité et appauvrit le travail des secrétaires - qui était un maillon essentiel de la fiabilité des soins - venant ainsi questionner le sens de leur travail – en particulier le sentiment d'utilité. Cette tâche de transcription considérée *a priori* comme « sans valeur ajoutée, routinière et chronophage » cachait en fait une activité de grande valeur pour l'organisation – en tant que barrière de sécurité- et pour la santé des travailleuses.

Dans le cas du système de détection des fractures, il existe des différences entre l'interprétation humaine et celle de la machine : alors que le radiologue se base sur un contexte clinique pour ses interprétations, le système de détection des fractures fonctionne sans une telle compréhension. Ainsi, les radiologues doivent recontextualiser les résultats en s'appuyant sur leur expertise, leurs connaissances du système et la situation clinique pour leur donner du sens. Bien que le système puisse orienter l'interprétation des images, c'est finalement le radiologue qui évalue la pertinence des résultats en tenant compte du contexte clinique et des informations sur le patient. L'expertise du radiologue et le contexte clinique jouent un rôle crucial dans la compréhension des résultats du système de détection des fractures et dans sa responsabilité au sein des pratiques radiologiques. Par ailleurs, dans certains contextes, ces systèmes étaient de fait utilisés par des radiologues novices ou des internes aux urgences - laissés seuls et sans pairs experts par le manque d'effectifs et l'intensification du travail dans les hôpitaux - pour apprendre à interpréter et contrôler les résultats de l'IA.

Incidences potentielles et effectives liées aux usages de l'IA : apports et risques variables selon les situations et les professionnels

Le tableau 5 synthétise les apports et les risques associés aux deux systèmes.

Tableau 5. Apports, incidences et risques associés aux systèmes de dictée vocale et de détection de fracture en radiologie.

	Apports	Incidences négatives & Risques
Système de dictée à Reconnaissance Vocale (RV)	<ul style="list-style-type: none"> - Gain de temps dans le processus de production d'un compte-rendu - Soutenir la construction du contenu et de la forme d'un compte-rendu intelligible et pertinent par les radiologues 	<ul style="list-style-type: none"> - Disparition d'une dimension constructive de l'activité des secrétaires - Perte de sécurité des radiologues sans la relecture des comptes-rendus par les secrétaires - Augmentation de la charge de travail des radiologues - Dégradation de la qualité orthographique
Système de Détection des Fractures (DF)	<ul style="list-style-type: none"> - Système DF en tant que « filet de sécurité » - Aider les internes radiologues dans l'interprétation des radiographies (souvent sans accompagnement par un senior) - Gain de temps mitigé et difficile à évaluer de façon objective 	<ul style="list-style-type: none"> - Risque de sur-confiance : nécessité de garder le contrôle - Risque d'empêchement de l'apprentissage et de l'acquisition des compétences - Risque de perte de temps (par exemple dans le cas de faux positifs incitant le radiologue à passer du temps à confirmer ou infirmer une fracture alors qu'en réalité il n'y a pas de fracture)

Dualité entre opacité/explicabilité, et fiabilité pertinence de l'action

Dans la ligne des résultats des études présentées plus haut, nos résultats soulignent une nouvelle fois l'importance de comprendre les éléments soutenant les résultats proposés par des SIA dans certaines situations. L'opacité du processus d'apprentissage des IA et des capacités d'amélioration du système de dictée RV limite la contribution des radiologues au développement du système. De même, le manque d'explications concernant les résultats du système DF peut entraver la compréhension des radiologues vis-à-vis des fractures détectées par le système mais qui ne leur sont pas visibles. Les informations fournies par le système, telles que le degré de certitude et la zone entourée, ne permettent pas toujours aux radiologues de comprendre explicitement la façon dont la fracture est détectée. Ainsi, le radiologue ne peut pas donner de sens au résultat.

5.3.5. Vers une vision élargie et située de la performance

L'introduction des SIA dans les situations de travail ne peut donc pas être considérée comme une solution miracle pour augmenter l'efficacité et améliorer la performance sans prendre en compte les effets réels sur les travailleurs, leurs activités mais aussi l'organisation du travail. En effet, la mise en place de l'IA peut impliquer de penser ou repenser l'organisation du travail dans sa globalité et non

uniquement la répartition des tâches entre les SIA et les humains, ce qui est pourtant principalement évoqué dans les réflexions actuelles (Gamkrelidze, 2022). L'utilisation de ces systèmes peut dans une certaine mesure soutenir les travailleurs, mais soulève également certains enjeux et risques pour la performance, dont des enjeux de santé au sens large qui englobe le vécu et le rapport subjectif au travail, les possibilités de travail collectif qui est un facteur de préservation de la santé. Le travail peut nuire à la santé notamment en cas de sur-sollicitations physiques ou psychologiques dans des environnements de travail contraints, mais aussi offrir une opportunité de construction de celle-ci (Delgoulet & Vidal-Gomel, 2013), par le biais du développement des compétences et des possibilités offertes aux travailleurs d'être acteurs du processus de construction de leur travail et donc de leur santé. Pour favoriser un travail performant et en santé, l'enjeu est aussi de trouver et de maintenir un équilibre entre les ressources (techniques, organisationnelles, humaines) et les contraintes mobilisables par les travailleurs, tout en intégrant les exigences de performance. La performance ne peut être réduite ici à la productivité ou la rentabilité. Elle doit être abordée « du point de vue des organisations (production, qualité, sécurité, délai...), mais aussi des personnes qui ont à cœur de fournir un travail de qualité » (Delgoulet & Vidal-Gomel, 2013, p.22), de la valeur qu'ils accordent à leurs activités. L'enjeu majeur de l'introduction de l'IA au travail consiste alors à chercher le meilleur compromis entre les critères de santé et de performance, voire une performance qui englobe les enjeux de santé. Pour cela, il est essentiel de dépasser les discours centrés sur des bénéfices « universels » de l'IA et ainsi, tenir compte du caractère variable et singulier des situations de travail où l'IA pourrait être intégrée, que ce soit en termes d'utilité et de pertinence, de risques ou de dégradations dans les activités de travail (Gamkrelidze, 2022). On voit donc apparaître ici le besoin d'explorer les modalités de conduite de projets d'intégration d'IA au travail, favorisant ou non, le développement d'activités performantes et préservante pour la santé.

A Retenir...

- Les études empiriques sur les transformations réelles du travail liées à l'introduction des SIA nuancent les résultats expérimentaux quant à la performance des IA. Les SIA introduit, comme second lecteur dans certaines situations, peut constituer un « filet de sécurité », mais surtout pour les professionnels novices et comme palliatif dans des situations de travail dégradées.
- Seule une analyse fine et systémique du travail réel ancrée dans l'expérience des professionnels permet d'appréhender la valeur réelle de l'activité -avec ou sans un SIA,- au-delà des projections théoriques sur ce qui fait valeur.
- Ces travaux identifient des effets sur la santé des professionnels (augmentation de la charge de travail, perte de sens au travail, entrave au travail collectif), une transformation potentielle de la construction de l'expérience professionnelle et des risques sur la fiabilité des actions menées avec une IA, notamment en lien avec l'opacité des systèmes.
- On observe un effet des contextes et des choix organisationnels sur les développements.
- Les usages pertinents impliquent la mise en œuvre de « pratiques interrogatives » soutenant la construction d'une distance critique pour comprendre les résultats de l'IA, afin donner du sens entre les interprétations des professionnels et les résultats de l'IA. Ces pratiques plus que nécessaires ne sont souvent pas soutenues dans des organisations du travail « sous-contraintes » et se révèlent donc coûteuses en temps pour des professionnels qui ont déjà une charge de travail importante.

5.4. La conduite des transformations du travail par l'IA : un impensé des discours sur l'IA et des démarches de conduites de projets

Les résultats précédents montrent bien l'importance de l'expérience et des activités des professionnels dans le développement d'usages pertinents et performants de SIA au travail. Ils soulignent également l'importance des choix organisationnels liés à l'introduction de ces systèmes dans les situations de travail comme un élément déterminant de leurs usages effectifs, ces usages étant souvent éloignés des promesses « théoriques ». Cependant, les démarches de conception et d'accompagnement des projets permettant de construire ces usages pertinents de l'IA restent le plus souvent impensées – voire opaques - non seulement dans les discours sur l'IA au travail, mais aussi dans les entreprises ou institutions. Ce manque ne permet pas d'apprécier de façon claire comment les systèmes ont été introduits et déployés et plaide donc pour une attention renouvelée sur la manière dont les décisions se prennent dans les projets d'IA et sur leurs conduites.

Pourtant, plusieurs démarches visant à « encadrer » la conception des SIA existent, même s'ils sont parfois critiquables. Certains se centrent sur la prise en charge des questions éthiques ou sur la « place des humains » dans le processus de conception en s'inspirant de démarches anciennes développées en ergonomie/facteurs humains et dans le champ de l'Interaction Humain-Machine, ou des principes de la conduite du changement. Nous reprenons certaines de ces propositions pour en souligner les limites et prôner des démarches de conduite de projets « centrées travail » et qui donnent une valeur à la fois à l'expérience et l'activité des professionnels et à la pluralité des logiques qui peuvent mobiliser dans les prises de décisions.

5.4.1. *Des principes éthiques et des démarches centrées « humain » pour encadrer la conception et l'intégration de systèmes d'IA*

Des premiers éléments de cadrage des processus de conception viennent d'une volonté de prise en charge des questions éthiques liées à la diffusion des SIA. L'Europe propose des « Lignes directrices en matière d'éthique pour une IA digne de confiance » (AI HLEG, 2019), une IA centrée sur l'humain, respectueuse des lois et règlements, respectant des principes et valeurs éthiques. Dans le même sens, la CNIL³¹ (CNIL, 2017) insiste sur la prise en compte de deux principes fondamentaux dans la conception des SIA - le principe de loyauté et le principe de vigilance³² – pour limiter des phénomènes de « sur-confiance » ou de « déresponsabilisation » dans les prises de décisions, favorisés par le caractère opaque des systèmes d'IA. La CNIL fait ainsi la proposition d'introduire « *une obligation de redevabilité ou d'organisation de la responsabilité* » (p. 51), à travers « *une attribution explicite des responsabilités impliquées par son fonctionnement* » (p. 51). Elle souligne le besoin d'encadrement quant à la quantité et à la qualité des données mobilisées pour éviter des dysfonctionnements « en bout de chaîne ».

Un deuxième groupe de propositions concerne l'encadrement des démarches de conception de systèmes d'IA par des démarches dites centrées Humain (« *Human-Centered AI* » : HCAI). Ces démarches reprennent notamment des principes de la conception centrée utilisateur (Norman and Draper, 1986 ; Bannon, 2011), en y adjoignant des critères éthiques. Plusieurs chercheurs (Reidl, 2019 ; Xu, 2019; Shneiderman, 2021, Schmager, et al., 2023, etc.) ou institutions portent ces démarches, par exemple aux Etats-Unis d'Amérique où des instituts de recherches dédiés à l'HCAI ont été créés par Stanford University, UC Berkeley, et le MIT. . Globalement, ces démarches mettent en avant le principe d'augmentation de l'humain par l'IA plutôt que son remplacement. De plus, l'humain doit pouvoir mettre en œuvre un contrôle critique de l'IA, renvoyant à l'importance du principe d'explicabilité des SIA (Schoonderwoerd et al., 2021). Ces principes sont opérationnalisés de différentes manières. Par exemple pour tenter d'opérationnaliser concrètement les principes éthiques généraux décrits

³¹ La Commission nationale de l'informatique et des libertés.

³² Le principe de loyauté vise à encadrer ce que fait l'algorithme, ce principe est complexe compte tenu des résultats inattendus, voir incompréhensibles, d'une IA. Le principe de vigilance est pour orienter les conceptions des systèmes d'IA, en « obligeant » les concepteurs à prendre en compte ce caractère imprévisible et évolutif des systèmes.

précédemment, (Shneiderman, 2020) propose des mesures pratiques pour une gouvernance de l'IA centrée humain : concevoir des SIA fiables et transparents conçus sur la base de « bonnes pratiques » d'ingénierie logicielle ; chercher à développer une culture de sécurité autour de l'IA soutenue par des stratégies de gestions pertinentes ; chercher à favoriser la confiance a priori envers l'IA à travers des processus de certification et une surveillance indépendante de son introduction au travail par secteurs d'activités. (Xu, 2019; Xu et al., 2021) proposent quant à eux de développer un approche compréhensive dite « *Human-Centered AI* » (HAI ou HCAI)³³ (voir l'Annexe 2 pour une description plus fine de ces deux approches) qui s'appuie sur les principes suivants : une conception en accord avec les principes éthiques, i.e. qui évite la discrimination et soutient des pratiques équitables et justes ; une conception qui reflète l'intelligence humaine ; un processus de conception qui favorise des SIA explicables, compréhensibles, utiles et utilisables. Enfin, Hagemann, Rieth, Suresh et Kirchner (2023) proposent une approche collective de la conception de SIA appelée « *team-centered IA* ». Cette perspective propose d'intégrer dans la conception des SIA des critères essentiels à un travail collectif de qualité :

- une réactivité dans les interactions entre membres de l'équipe) ;
- le besoin de construire une « conscience de la situation de travail partagée » par les différents partenaires du travail collectif pour agir là aussi de manière pertinente en fonction du travail accompli par les autres ;
- la flexibilité de la prise de décision.

La prise en charge de ces critères renvoie à des exigences techniques telles que des fonctions d'IA soutenant l'articulation des objectifs des différents partenaires de l'équipe, la communication et la prise de décision collective. Ceci implique de soutenir des échanges communicationnels entre les membres de l'équipe.

Un troisième ensemble de démarches concerne des propositions d'encadrement des projets d'intégration des SIA, au-delà de la conception de l'interaction entre l'humain et la machine. Ces démarches sont principalement portées par des entreprises ou cabinets de conseil. Certaines affichent une volonté de donner une place importante aux travailleurs dans la conception et l'intégration de l'IA dans les situations de travail. Par exemple, la démarche proposée par (Nerce et al., 2018) pour « *l'adoption de l'IA dans les organisations* » met en avant son objectif de « *garder l'humain au cœur de la transformation* » (Nerce et al., 2018 ; p. 82) à travers un programme « *d'acculturation* » dans le but de « *démystifier* » l'IA et « *d'engager* » les salariés dans le projet de transformation. Ceci passe notamment par la création d'« IA Labs » permettant d'identifier et de créer des cas d'usages d'IA et, les « besoins » en termes de transmission des savoirs concernant ces systèmes et leurs usages. D'autres démarches visent à mieux appréhender les enjeux techniques liés à l'introduction de l'IA, comme la nécessité d'avoir une architecture informatique compatible avec les systèmes d'IA (Fischer & de Broca, 2018). Il s'agit de créer des conditions pour « être prêt pour l'IA » (« *AI-Ready* ») (Académie des technologies,

³³ HAI est aussi un institut de recherche créé récemment par l'université de Stanford, Berkeley et MIT.

2018). Dans le même sens (Fischer & de Broca, 2018) mettent en avant l'importance du développement d'une politique de « gouvernance des données » dans les organisations en lien avec les enjeux éthiques et de responsabilité et d'une politique globale de conduite de projets soulignant l'importance des plusieurs points :

- la communication et de la sensibilisation en interne autour des projets d'IA,
- le besoin d'alignement du projet d'introduction des SIA sur la stratégie globale de l'entreprise,
- un partenariat métiers-DSI (la direction des systèmes d'information) dans l'intégration des systèmes d'IA,
- le repérage des talents en interne,
- la stimulation des équipes via des challenges IA,
- le développement d'un centre de compétences dédié à la « *data intelligence* ».

Ces dernières démarches insistent sur des éléments essentiels de la conduite de projets et du besoin de réflexion sur une infrastructure technique et organisationnelle autour de l'IA. Bien que ces éléments soient effectivement importants, ils mobilisent parfois des éléments de langage des démarches de la conduite du changement qui sont discutables et insuffisants. Bien qu'ils mettent l'accent sur la communication, le « besoin » de pédagogie et de formation pour favoriser la réussite de projets, ils sont souvent pensés de manière descendante. Sous couvert d'éléments de langage séduisants, ces approches semblent se focaliser sur la technologie en tant que solution et à traiter au second plan tous les enjeux liés au travail (organisation, santé, présence de ressources dans la situation pour agir de manière efficace...). Ceci est critiquable pour la position résiduelle qu'elle donne à l'humain et à son travail réel vis-à-vis de l'outil technique (Rabardel, 1995; Woods, 1996).

Examinons alors ce qu'il en est des conduites de projets effectivement mises en place dans le cadre de projets d'implantation d'IA.

5.4.2. Des intentions de transformations du travail liés à l'IA renvoyant à des logiques partielles et un impensé de la conduite de projets

Comme nous l'avons souligné, les conduites de projets d'intégrations d'IA au travail restent opaques. Dans la majorité des études que nous avons mobilisées, les intentions qui sous-tendent la genèse des projets d'introduction d'IA ne sont pas décrites, de même que les démarches de conception des SIA et de conduite de projet. Nous pouvons toutefois identifier quelques grandes intentions sur la base des travaux de Gamkrelidze (2022), de Ferguson (2019) et des cas des secteurs de la presse et du journalisme évoqués plus haut.

Les intentions d'introduction d'un SIA naissent parfois au travers de rencontres et d'interactions entre des spécialistes de l'IA et des professionnels de certains secteurs (médical par exemple). Dans les entretiens réalisés avec une cinquantaine de professionnels (concepteurs ou utilisateurs potentiels d'une IA), Gamkrelidze (2022) révèle que la majorité des SIA évoqués sont issus des rencontres entre des spécialistes de l'IA et des métiers opérationnels qui sont en mesure d'identifier des besoins et des

problématiques de leur domaine d'activité. Les choix d'intégrer et d'utiliser des systèmes sont souvent motivés par exemple par des chefs de service ou des cadres de la direction portant un certain intérêt à l'IA. Les choix sont aussi orientés par de nombreuses propositions de la part des start-ups qui font tester leurs systèmes par des utilisateurs ciblés, de manière gratuite ou à prix réduit. Différents cadres de partenariats sont alors explicités par les interviewés, par exemple les collaborations entre les industriels, les start-ups et les opérationnels comme les radiologues qui vont activement participer au développement d'un système en échange d'une utilisation gratuite de ce même système. Cette forme de « collaboration » peut sembler intéressante pour prendre en compte les spécificités des différents milieux professionnels, mais elle conduit souvent malgré tout à intégrer en situation de travail des systèmes génériques que ne sont pas forcément adaptés, et l'implication des professionnels se révèle parfois coûteuses en temps pour eux.

Au-delà de ces rencontres, on retrouve des intentions d'introduction de l'IA qui reprennent les promesses des discours de l'IA pour « libérer » les travailleurs des tâches pénibles, répétitives et les concentrer sur les tâches à plus forte « valeur ajoutée », dont nous avons déjà discutés les limites. L'IA peut également être vue comme la « solution » technique à des problèmes qui ont des causes plus systémiques. C'est le cas par exemple de l'introduction du système de dictée à reconnaissance vocale dans le service de radiologie, qui est en partie censée répondre au problème de retard de transcriptions des comptes rendus radiologiques, ou de l'introduction d'un système de détection de fracture (voir « Zoom Sur.. » de la section 5.3), qui est pensé dans certaines structures comme un moyen de pallier le déficit de radiologues, notamment aux urgences. Dans certains cas, on pourrait donc penser que les promesses de l'IA sont au service de modèles gestionnaires de réductions des coûts en lien avec des rationalités financières. Le cas de la société Onclusive est emblématique de ce point de vue : les promesses de l'IA sont instrumentalisées pour justifier un plan social radical obéissant à des logiques de rationalisation des coûts. Dans le même sens, l'IA est présentée comme une évolution technologique « inévitable » du domaine journalistique marqué par des logiques économiques centrées sur le besoin de rationalisation des flux de travail (Sirén-Heikel et al., 2023). Dans d'autres situations, le projet d'introduction au travail de l'IA vise à faire face à une forme de « bazar informationnel » liés à l'empilement de systèmes d'informations ou de procédures – comme le cas des Chatbots. Enfin, certains professionnels décrivent une crainte d'être « à la traîne » par rapport à d'autres si l'on ne s'équipe pas d'IA et cherchent à s'informer et/ou expérimenter pour évaluer la pertinence des SIA. D'autres, plutôt les décideurs obéissent à une forme de « course technologique » et souhaitent introduire des SIA sans forcément questionner leur pertinence de la solution IA dans leurs contextes propres. Ceci s'inscrit dans une forme d'isomorphisme institutionnelle, traduisant la convergence de certaines décisions techniques ou organisationnelles dans des organisations d'un même champ (DiMaggio et Powell, 1983).

Quant aux modalités de conduite de projets effectivement mise en œuvre pour introduire une IA dans organisations, elles sont peu voire pas décrites. Quand elles le sont, elles reprennent les principes de la conception « centrée humain » (itération avec et participation des utilisateurs, attention aux recueils des besoins utilisateurs) sans que l'on ne sache réellement ni si, ni comment ils ont été opérationnalisés

concrètement. Une phase importante mise en avant, par exemple dans le cadre d'un projet de conception et d'introduction d'un chatbot juridique dans une grande entreprise (Gamkrelidze, 2022), porte sur le recueil des besoins des métiers opérationnels (futurs utilisateurs) à travers des ateliers, des discussions, des entretiens ou encore de l'expérience et des connaissances personnelles de certains concepteurs qui évoquent les IA-Labs mentionnés plus haut. Le travail de (Flandrin et al., 2022) révèle que la mise en place du chatbot qu'il a étudié a été imposée sans implication de l'équipe opérationnelle dans la démarche de conception. A contrario, dans le cas du système d'IA étudié par (Henry et al., 2022), l'équipe a été intégrée durant les différentes phases. Afin d'expliquer le fonctionnement et l'utilisation du système, plusieurs rencontres ont été réalisées entre les concepteurs – qui déployaient le système dans les services de l'hôpital – et les utilisateurs ciblés qui étaient les professionnels de santé. De plus, des moyens ont été mis en place pour permettre aux utilisateurs de poser des questions concernant le système tout au long de son utilisation. Ceci a très probablement favorisé l'appropriation du système.

Cependant, la réelle implication des travailleurs en tant qu'acteurs à part entière des projets et la prise en considération de leurs activités, de leur expérience et de ce qui a du sens et de la valeur dans leur travail apparaît faible voire inexistante. Globalement, ces constats liés à la prise de décision en entreprise et à la conduite de projets ne sont ni nouveaux, ni propres à l'IA et se retrouvent dans la plupart des projets (par exemple Mintzberg et al., 1976 ; Daniellou, 1992). Aux vues des potentialités des SIA, on peut néanmoins se questionner sur les risques accrus pour les entreprises et les travailleurs de ses projets insuffisamment pensés de manière holistique et systémique, c'est-à-dire qui tiennent compte des différents aspects de l'activité humaine (physique, cognitif, social, émotionnel, collectif), des interdépendances entre les différentes structures qui composent une organisation, etc. On retrouve des écueils classiques des projets de conception (Barcellini, 2020) comme la prégnance des logiques techniques ou financières, ou encore peu de prise en compte de la complexité, de la variabilité du travail et précisément de ces aspects.

5.4.3. Vers des démarches de conduite de projet centrées travail qui donnent une valeur à l'expérience et l'activité des professionnels

Pour dépasser ces écueils, l'ergonomie propose des démarches alternatives permettant la prise en compte des transformations systémiques du travail d'un point de vue collectif, organisationnel et technique dans la conduite de projet techniques et organisationnels (Daniellou, 1992 ; Barcellini, 2015). Cette approche souligne l'importance de donner une possibilité aux professionnels de pouvoir penser et débattre de transformations futures de leur travail et de ses conséquences collectives, organisationnelles et sur leur santé ; et des critères de qualité de leur travail – de ce qui a de la valeur dans leur activité au-delà de choix descendants. Cette approche insiste aussi sur la nécessité de construire socialement les projets et d'enrichir leurs définitions, c'est-à-dire de créer des moments de mise en discussion des différentes logiques à l'œuvre (financière et technique, mais aussi organisationnelle, liée à la qualité du travail, à la santé, à la préservation et aux développements des compétences) et de la construction d'une vision systémique du « problème » que l'introduction de l'IA

chercherait à résoudre. Ceci implique également les acteurs du dialogue social, par exemple dans des démarches couplant dialogue social formel et informel et conduite de projets centrée travail (Galey et al., 2023 ; Lefeuvre et Barcellini, 2023). Ceci ne va pas de soi et invite à une inversion des rapports de pouvoir à l'œuvre dans les entreprises, et donc un engagement politique local – au sein des entreprises – et au-delà.

Nous reprenons ici certains des principes de ces conduites de projets « centrée travail » augmentés, d'autres plus spécifiques à la conception de systèmes d'IA.

Le premier renvoie au soutien de la participation effective (Darses & Reuzeau, 2004; Williams, 1983) des travailleurs et des autres acteurs impliqués dans le projet, et ceci dès les premières phases des projets de conception. Ceci implique de dégager du temps aux professionnels et une réflexion sur une organisation méthodique de la participation permettant une collaboration et des débats réels autour du projet. L'ergonomie propose des méthodes en ce sens, comme :

(1) l'analyse ergonomique du travail qui permet de donner à voir et de mettre en discussion le travail tel qu'il se fait, les stratégies mises en œuvre par les professionnels, leurs ajustements, et la coopération qui caractérisent leur travail ; la façon dont ils font fassent aux contradictions organisationnelles et/ou à une diversité de situations :et enfin ce qui a de la valeur dans leur travail ;

(2) la simulation de l'activité qui permet aux professionnels de se projeter dans une future situation de travail (Maline, 1994 ; Barcellini, 2015 ; Van Belleghem, 2018).

Le second renvoie à la mise en place d'observatoires des effets de la mise en place et l'usage des SIA sur le travail pour soutenir les ajustements organisationnels et techniques nécessaires pour aboutir à des systèmes qui conviennent non seulement aux responsables mais aussi utilisateurs. Cette évaluation à travers l'observatoire devrait être ancrée dans un processus continu dans les organisations, notamment parce que l'IA se caractérise par des possibilités d'évolution rapide (Gamkrelidze, 2022).

Le troisième renvoie aux développements des capacités critiques d'usages des SIA. L'accent est actuellement mis sur le besoin pour les travailleurs de se concentrer sur le développement de compétences sociales ou relationnelles en complément des tâches cognitives réalisées par l'IA. Or, avec l'usage croissant des systèmes d'IA et leurs avancées rapides, le véritable défi résiderait dans le développement et le maintien des capacités critiques et de contrôle envers les résultats et les informations produits par les SIA. Le développement de telles capacités permettrait aux travailleurs de prendre des décisions éclairées, avec des risques réduits, lorsqu'ils utilisent l'IA. Ceci peut être facilité par : les fonctionnalités des systèmes eux-mêmes ; des formations spécifiques ; des temps dédiés aux évaluations des systèmes dans des situations réelles, considérés comme des moments légitimes et des étapes à part entière de ces démarches ; des outils réflexifs et accessibles à l'ensemble des acteurs pour pouvoir co-construire ces évolutions liées à la fois à l'IA, aux activités et aux organisations du travail.

A Retenir...

- Des propositions de démarches de conception d'IA existent, elles reprennent des principes éthiques à respecter, des recommandations pour assurer la compatibilité de l'IA l'infrastructure technologique existante, ou des principes de modèles de conduite du changement (communication, pédagogie dans les projets) ou d'une conception d'IA centrée humain.
- Les intentions derrière les projets d'introduction d'IA au travail sont opaques dans les études et cas mobilisés. Lorsqu'elles sont mises à jour, on retrouve des intentions parfois marquées par des logiques de rationalisation financière, de course technologique, de solutionnisme technologique. Les conduites de projets sont le plus souvent impensées.
- Le développement de démarches de conduite de projets d'IA centrées travail sont nécessaires pour enrichir les logiques prises en compte dans la définition des projets d'IA et permettre aux professionnels de pouvoir penser et pouvoir débattre des transformations de leur travail et de ce qui fait valeur pour eux.
- Le développement de ces démarches de conduites de projet impliquent de repenser les rapports de force dans les prises de décisions et dans la conduite des projets et de porter une vision émancipatrice du travail dans les projets : un travail qui soit source de développement de l'expérience, de santé et porteur d'un sens.

6. Recommandations et perspectives

Le travail présenté dans ce rapport avait notamment pour objectif de porter un point de vue critique et empiriquement informé sur les promesses et les craintes associées aux transformations du travail et de l'emploi liées à l'IA. Deux constats s'imposent. Premièrement, malgré un nombre croissant de publications portant sur l'IA, les travaux font relativement peu de place à ses liens avec les évolutions du travail et de l'emploi explorées par des approches systémiques et situées. Beaucoup d'études restent expérimentales – et elles s'inscrivent encore trop dans une approche déterministe de latechnologie qui néglige le rôle des choix organisationnels, de la conduite de projet et la place des activités et de l'expérience des professionnels. Or, ces éléments sont centraux quant à d'une part la compréhension des transformations du travail et de l'emploi liées à l'IA, et d'autre part l'introduction « vertueuse » de celle-ci dans les organisations. Deuxièmement, beaucoup de publications sont consacrées à certains secteurs professionnels (par exemple la médecine) ou des fonctionnalités ou SIA emblématiques (par exemple, les chatbots, l'automatisation de tâches répétitives, l'aide au diagnostic ou recherche et traitement d'informations).

Dans cette partie conclusive, nous nous appuyons sur ces constats pour proposer des perspectives et recommandations qui favorisent le développement d'usages **soutenables** socialement et économiquement de l'IA dans la sphère professionnelle, lorsque sa mise en place peut avoir du sens pour l'ensemble des acteurs concernés. Par soutenables économiquement et socialement, nous entendons ici des usages qui préservent l'emploi et favorisent la soutenabilité du travail (Gaudart et al., 2019 : Gollac et al., 2008).

6.1. Les 4 piliers du développement d'usages soutenables de l'IA dans la sphère professionnelle

Les approches économiques dominantes qui évaluent les conséquences de l'IA sur l'emploi prennent appui sur un modèle de la production où les choix d'organisation du travail ne sont pas considérés et sur un modèle du travail pensé comme une liste de tâches éloignée du réel du travail. Dans ce cadre, les SIA sont mobilisés pour automatiser les tâches existantes avec des effets destructeurs sur l'emploi. C'est la frontière des tâches que les SIA ne savent pas encore exécuter qui fixe, à une date donnée, le périmètre de ces effets sur l'emploi. Par ailleurs, une partie des effets sur l'emploi est peu visible (délocalisation, sous-traitance, départs naturels, réduction des effectifs déjà engagée).

Plutôt que d'adopter cette approche de la production et du travail centrée sur les tâches, il semble fructueux d'envisager les usages des SIA dans une approche centrée sur la production de connaissances nouvelles, l'activité et l'expérience des professionnels qui sont sources d'innovation. Dans un contexte d'adoption de technologies nouvelles, **l'investissement dans la capacité d'apprentissage de l'organisation, et dans la valeur du travail et de l'expérience des professionnels**, sont des choix d'organisation du travail aux effets protecteurs sur l'emploi et le travail

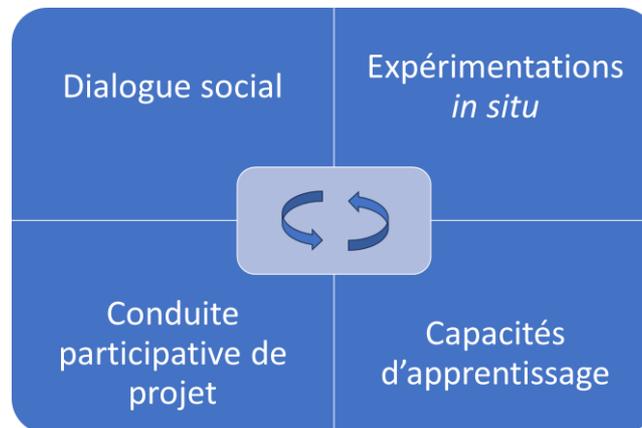
(sa performance, la santé des professionnels, et la possibilité de développement de leur expérience). La façon dont les SIA sont conçus et dont leurs usages sont déterminés à l'occasion de leur déploiement dans les contextes de travail réel est une des clés des évolutions favorables de l'emploi et du travail. Le développement des usages de l'IA est bien lié à une question de répartition du pouvoir dans les organisations et pas uniquement à ses fonctionnalités. Dans ce cadre, les régulations qui se jouent au sein des collectifs de travail et avec les concepteurs sont essentielles. Mais, les travaux sur la régulation de l'IA soulignent la nécessité de penser une régulation d'ensemble pour pallier les incohérences et les lacunes des instruments juridiques actuels. La possibilité de réguler les conséquences des SIA sur le travail et l'emploi dépend beaucoup du modèle de relations professionnelles. De grandes différences existent entre les pays.

Les travaux s'intéressant aux transformations du travail peignent **un tableau nuancé des apports réels de l'IA** – ses transformations sont dépendantes des choix organisationnels qui sont opérés dans les entreprises ou institutions. En effet, des études expérimentales montrent des bénéfices pour la performance, mais ces résultats dépendent de l'expertise des professionnels et restent cantonnés à certaines tâches à la portée de l'IA (par exemple les tâches de diagnostic radiologique, de recherche d'informations ou de production de textes). Qui plus est, contrairement à l'hypothèse que l'IA (en particulier générative) pourrait réduire les inégalités du point de vue de l'expertise, **ces bénéfices sont inégaux selon les profils des salariés** (par exemple, les experts seniors semblent moins tirer profit de l'IA – ChatGPT- en termes de performance pour certaines tâches que les plus jeunes). En situation de travail réel, il existe **des effets négatifs de l'IA dans certains contextes** (par exemple charge de travail accrue, sens du travail affecté, entraves à la coopération et au développement de l'expérience professionnelle) et **l'opacité** des SIA fait peser un risque tant pour les questions d'apprentissage et de développement de l'expérience que de fiabilité des actions réalisées avec l'IA. Ici encore, quand nous parlons « d'effets » de l'IA, mais il s'agit en fait d'un raccourci de langage : car, c'est bien la manière dont un système est choisi, déployé, intégré dans l'organisation, et dont le projet de transformation a été conduit, qui est déterminante et non pas seulement la technique en elle-même. Par ailleurs, notre analyse de la littérature souligne la faible place accordée aux professionnels et à leurs expériences dans la conduite de projets visant l'intégration de systèmes d'IA au travail ; ces conduites de projets sont de fait insuffisamment pensées ou restent cantonnées à des principes séduisants qui, de fait, ne sont pas opérationnalisés. Enfin, un développement rapide des usages des systèmes d'IA génératives semble en cours sans qu'ils ne soient encore bien compris et documentés.

Nos travaux convergent vers une critique du déterminisme et du solutionnisme technique qui prévalent dans les discours sur l'IA dans la sphère professionnelle. A la lumière de notre approche croisée (ergonomie et économie du travail/de l'emploi) et de notre revue de la littérature, nous révélons l'importance de penser de manière conjointe : (1) les choix organisationnels autour de l'introduction de l'IA, (2) la conduite des projets et (3) le dialogue social pour soutenir la capacité d'apprentissage des organisations et la construction sociale des transformations du travail, ainsi que (4) la conduite et la documentation d'expérimentation d'usage d'IA *in situ*. Par *in situ*, il faut entendre ici la confrontation des

SIA aux conditions réelles du travail. Notre analyse critique de la littérature montre que ces quatre éléments constituent au moins en partie les piliers du développement d'usages soutenables, économiquement et socialement, de l'IA dans la sphère professionnelle.

Figure 5 : Conditions pour le développement d'usages soutenables de l'IA dans la sphère professionnelle



Soutenable, car le couplage de ces éléments est vu comme une condition pouvant préserver l'emploi, favoriser la performance des organisations, aider à développer des usages pertinents de l'IA, permettant aux professionnels d'agir efficacement sans dégrader leur santé et en développant leur expérience (Figure 5). Car, au-delà de l'IA, l'enjeu central est le futur du travail et la place qu'il occupe dans l'existence humaine : avec ou sans l'IA, il nous semble que le point central est de promouvoir un travail soutenable, c'est-à-dire « biocompatible », adapté aux propriétés fonctionnelles de l'organisme humain, à son évolution au fil de l'existence et ne dégradant pas la santé ; « ergo-compatible », et donc propice à l'élaboration de stratégies de travail et « socio-compatible », donc favorable à l'épanouissement dans les sphères familiale et sociale, à la maîtrise d'un projet de vie (Gaudart et al., 2019 ; Gollac, Guyot et Volkoff, 2008). Il s'agit aussi de promouvoir un retour à une pensée systémique des transformations du travail explorant la pertinence effective de l'introduction des SIA à travers des conduites de projets participatives et s'appuyant sur un modèle précis du travail réel. Enfin, discuter, expérimenter, mettre en œuvre des pratiques réflexives et accompagner le déploiement de l'IA, apparaissent comme des conditions tout autant nécessaires que les régulations formelles (par exemple des conduites de projet participatives et centrées sur le travail réel) pour aller vers ces transformations soutenables du travail et de l'emploi, pour améliorer la qualité des connaissances produites et faciliter l'exploration de savoirs nouveaux sur le processus de production des entreprises. Au vu des avancées rapides actuelles, il est donc essentiel de promouvoir la conduite et la documentation d'expérimentation de développement d'usage *in situ*, c'est-à-dire dans des milieux de travail réels.

Or, d'un point de vue économique, le développement de la capacité d'apprentissage des organisations est une question d'investissement primordiale pour aller vers des usages soutenables de l'IA, et ce d'autant plus que cet investissement a stagné sur les deux dernières décennies. Dans un contexte de

globalisation des économies, puis face à la grande récession, les entreprises auraient en effet privilégié un investissement dans les technologies digitales orienté vers la rationalisation et l'optimisation de leurs processus de production tout au long des chaînes de valeur (Bodrózic et Adler, 2018), mouvement que nous retrouvons concernant l'introduction de l'IA. Le développement de la capacité d'apprentissage des organisations en aurait été négligé, notamment en France. Les dispositifs existants de dialogue social administré pourraient être mobilisés pour favoriser les investissements dans la capacité d'apprentissage des organisations. Les conduites et la documentation d'expérimentations *in situ*, à travers des modalités spécifiques d'accompagnement de dispositifs de développement et d'appropriation des usages des IA – notamment génératives - en milieu de travail et dans le milieu éducatif, sont essentielles pour suivre le tempo très rapide des évolutions alors même que nous sommes en phase d'apprentissage du travail avec ces technologies.

Inversement, si l'on ne promeut pas cette capacité d'apprentissage des organisations, les scénarios d'épuisement des travailleurs (par exemple de leur santé, de leur expérience, de leurs ressources) et de l'environnement pourraient se développer dans une course à la réduction des coûts et à la traque aux tâches prétendument « sans valeur ajoutée ». Dans ce sens, une concentration trop forte des entreprises sur le marché des produits (De Loeker et al. 2020), et de ses applications dérivées (Korinek et Vipra, 2023), mais aussi une concentration trop importante des employeurs sur le marché du travail (Araki et al. 2023) sont des éléments de contexte défavorables aux scénarios d'usages soutenables de déploiement de l'IA. Une attention particulière doit donc être portée aux risques introduits par des formes d'innovation qui favoriseraient une concentration sur les marchés des produits et du travail – et donc un appauvrissement de la diversité conduisant à une insoutenabilité économique et sociale des transformations du travail et de l'emploi liées l'IA.

6.2. Perspectives

Plusieurs perspectives découlent de notre analyse de la littérature :

- Une attention particulière doit être portée à la façon dont les formes d'innovation sont soutenues par les autorités publiques au travers de dispositifs comme le crédit d'impôt recherche. Ainsi, les dépenses de R&D destinées à la standardisation et à l'innovation marketing devraient être identifiées et suivies dans les budgets de R&D pour les limiter, car nous avons vu qu'elles conduisent à cet appauvrissement. Poser des limites aux formes de management algorithmiques et de surveillance automatisée en prenant appui sur le cadre législatif existant (RGPD par exemple) et sur la négociation collective (Doellgast et al., 2022) est un autre élément important.
- Ces constats appellent également à une analyse fine des motifs et stratégies des entreprises et institutions quant à l'introduction de l'IA, aux modèles de conduites de projets, de conduite des changements organisationnels et de positionnement sur les marchés qu'elles mettent effectivement en œuvre, au-delà des discours et éléments de langage énoncés quant aux objectifs de

développement durable, de performance ou d'augmentation de l'humain poursuivis, ou au caractère « inévitable » de l'introduction de l'IA.

- Ils appellent également à poursuivre des travaux empiriques et situés pour une diversité de SIA et de contexte sur les liens performance, santé, et travail avec une IA. Les travaux sur les usages réels de l'IA en situation, et les liens entre introduction de l'IA, santé et soutenabilité du travail étant négligés dans la plupart des études, sans doute parce qu'ils sont sous valorisés dans les pratiques de recherche et donc sous financés.
- Notre travail pointe aussi les risques d'inégalité dans les usages de l'IA. Nous en avons vu un exemple concernant les experts vs. les novices. A la lumière de ce constat qui reste à approfondir, il convient de s'intéresser à ce type d'inégalités : les documenter et comprendre leurs origines. Même si elle n'apparaît pas dans nos analyses, une attention particulière doit également être portée à un autre type d'inégalités, celles de genre liées aux usages des IA. On connaît tout d'abord la sensibilité des inégalités de genre au pouvoir de marché des employeurs. On connaît également les biais de genre dans les bases de données utilisées pour entraîner les IA et la forte ségrégation professionnelles des métiers associés au développement de ces technologies. L'analyse du corpus des publications scientifiques qui croisent le développement et les usages de l'IA et les 17 objectifs de développement durable de l'ONU montrent en outre que l'on identifie plus souvent des impacts négatifs que positifs des IA sur les inégalités de genre (Bianchini et al., 2023).
- Enfin, cette littérature identifie aussi des tensions entre développement des technologies digitales et écologie, qu'il est essentiel d'explorer car elles restent relativement absentes des travaux.

6.3. Préconisations

Soutenir le développement d'usages soutenables passe selon nous par la mise en place de politiques publiques ambitieuses pour lesquelles notre analyse critique de la littérature permet de définir plusieurs préconisations :

- Mobiliser les dispositifs du dialogue social administré pour soutenir les investissements financiers et humains dans la capacité d'apprentissage des organisations. Par exemple, dans les négociations sur les conditions de travail, l'égalité professionnelle et la Qualité de Vie au Travail, la formation professionnelle ou la Gestion Prévisionnelle des Emplois³⁴. L'accord cadre européen sur la digitalisation, signé par les partenaires sociaux en 2020, propose des pistes de réflexion intéressantes pour orienter un dialogue social constructif sur les usages des technologies digitales

³⁴ Dans le cadre du projet en cours Bridges 5.0, les accords d'entreprise et de branche mis en ligne sur Légifrance sont analysés pour identifier les éléments négociés autour de l'usage des technologies 4.0 (travaux en cours).

(European social partners framework agreement on digitalisation). Il s'agit bien ici de développer le pouvoir de négociation des salariés.

- Mettre en place un observatoire des expérimentations associées aux usages des IA dans des branches ou des secteurs particuliers, par exemple, la fonction publique ou le secteur du digital³⁵. Ces deux secteurs cherchent une maîtrise de leurs coûts et font face à des pénuries de main d'œuvre. L'arbitrage entre IA qui se substitue au, ou qui soutient le travailleur, y est donc très présent et identifier les avantages/inconvénients des deux stratégies et les moyens nécessaires à la mise en œuvre du scénario vertueux y est particulièrement utile. Cet observatoire pourrait également se nourrir des expérimentations et des études qui sont réalisées à l'échelle internationale, dans d'autres pays. Il se doit d'être extrêmement concret en articulant données quantitatives et d'expériences singulières (incluant analyse du travail et conduite des projets) de façon à alimenter le processus de réflexion sur les enjeux de politique publique.
- Construire des modalités d'accompagnement du développement et de l'appropriation des usages des IA, notamment génératives. Les usages des IA génératives dans l'enseignement et la recherche publiques sont particulièrement critiqués pour leur impact sur la capacité des individus à participer à un apprentissage collectif en milieu de travail (éducation) et le développement de savoir nouveaux (recherche).
- Soutenir la recherche interdisciplinaire en s'inscrivant dans des approches situées à travers des appels à projet de recherche en lien avec les perspectives de recherche énoncées plus haut, et que l'on peut regrouper autour de deux axes de recherche comme (1) le développement de l'IA au service des travailleurs en contexte de travail (2) le développement de l'IA au service des objectifs environnementaux.
- Analyser la capacité des cadres réglementaires existants (1) à réguler les entraves à la concurrence sur les marchés des produits et du travail notamment celles liées au développement et aux usages des IA génératives, (2) à imposer des limites au management algorithmique et à la surveillance automatisée, (3) à réduire les inégalités de genre entretenues par le développement de ces technologies.
- Évaluer le soutien public aux pratiques de standardisation et d'innovation marketing via le crédit d'impôt recherche et aux pratiques de recherche purement expérimentales et technodeterministes.
- Promouvoir une « culture » participative et réflexive de l'innovation technologique qui met l'accent (1) sur la nécessité d'intégrer réellement - au-delà des discours - les travailleurs concernés dans les processus de décisions de transformation technico-organisationnelle du travail ; (2) sur la

³⁵ Des recommandations similaires sont prônées par le LaborIA.

nécessité – et donc le temps – de développer une pratique réflexive (Schön, 1987) à même de soutenir la prise de distance critique face aux actions réalisées avec une IA.

- Améliorer les dispositifs de statistique publique permettant de suivre la diffusion des technologies émergentes ainsi que les changements organisationnels qui accompagnent la définition de leurs usages.

Remerciements

Les auteurs tiennent à remercier Arianna Sullivan, Tobias Génies, Charlyne Poncato, Titouan Le Pelley-Fonteny, Marina Launay, Emeline Maire, Jean Larbaigt et Annaëlle Soulet du CRTD du Cnam pour leur aide dans la relecture et la mise en forme de ce document.

Bibliographie

Académie des technologies. (2018). *Renouveau de l'Intelligence artificielle et de l'apprentissage automatique* (N° 9791097579074). http://academie-technologies-prod.s3.amazonaws.com/2018/04/06/13/49/30/183/Rapport_IA_DEF.pdf

Acemoglu, D., Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American economic review*, 108(6), 1488-1542.

Adams-Prassl, J. (2019). What if your boss was an algorithm? Economic Incentives, Legal Challenges, and the Rise of Artificial Intelligence at Work. *Comparative Labor Law & Policy Journal*, 41(1), 123-146.

AI HLEG. (2019). *Ethics Guidelines for Trustworthy AI*. <https://ec.europa.eu/digital>

Alhwaiti, M. (2023). Acceptance of Artificial Intelligence Application in the Post-Covid Era and Its Impact on Faculty Members' Occupational Well-being and Teaching Self Efficacy : A Path Analysis Using the UTAUT 2 Model. *Applied Artificial Intelligence*, 37(1), 2175110. <https://doi.org/10.1080/08839514.2023.2175110>

Anichini, G., & Geffroy, B. (2021). L'intelligence artificielle à l'épreuve des savoirs tacites. Analyse des pratiques d'utilisation d'un outil d'aide à la détection en radiologie. *Sciences sociales et santé*, 39(2), 43-69. <https://doi.org/10.1684/sss.2021.0200>

Arntz, M., Gregory, T., & Zierahn, U. (2016). *The Risk of Automation for Jobs in OECD Countries : A Comparative Analysis* (N° 189). OECD Social, Employment and Migration Working Papers. https://www.oecd-ilibrary.org/social-issues-migration-health/the-risk-of-automation-for-jobs-in-oecd-countries_5j1z9h56dvq7-en

Askenazy, P., & Bach, F. (2019). IA et emploi: Une menace artificielle. *Pouvoirs*, 170(3), 33-41.

Autor, D.H., Levy F., Murnane R. J. (2002). Upstairs, downstairs: Computers and skills on two floors of a large bank. *ILR Review*, 55.3 : 432-447.

Autor, D.H., Levy F., Murnane R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, 118.4 : 1279-1333.

Baiocco, S., Fernández-Macías, E., Rani, U., & Pesole, A. (2022). *The Algorithmic Management of Work and its Implications in Different Contexts* (N° JRC129749; JRC Working Papers Series on Labour, Education and Technology). Joint Research Centre (JRC). <https://data.europa.eu/doi/10.2760/459278>

- Ballantine, M. (1980). Conversing with computers—The dream and the controversy. *Ergonomics*, 23(9), 935-945. <https://doi.org/10.1080/00140138008924802>
- Bainbridge, L. (1983). Ironies of automation. In *Analysis, design and evaluation of man-machine systems* (pp. 129-135). Pergamon.
- Bannon, L. (2011). Reimagining HCI: toward a more human-centered perspective. *Interactions*, 18(4), 50-57. doi:10.1145/1978822.1978833
- Barcellini, F. (2015). *Développer des interventions capacitanes en conduite du changement. Comprendre le travail collectif de conception, agir sur la conception collective du travail*. Habilitation à Diriger des Recherches, Université de Bordeaux.
- Barcellini, F. (2020). Quelles conceptions de la coopération humains - robots collaboratifs? *Activités*, 17(1).
- Barcellini, F. (2022). The design of “future work” in industrial contexts. *Managing Future Challenges for Safety: Demographic Change, Digitalisation and Complexity in the 2030s*. Edited by Hervé Laroche, Corinne Bieder and Jesús Villena-López. Berlin/Heidelberg: Springer, 75-83.
- Barcellini, F., Lefeuvre, A.-G. (2023), L'industrie 4.0 passe (d'abord) par l'humain, *Santé & Travail*, 122, avril.
- Barcellini, F. (2019). Industrie du futur : quelle place pour le travail et ses transformations. In Bourdu E., Lallement M., Veltz P. et Weil T. (dir.), *Le Travail en mouvement*, Colloque de Cerisy, Presses des Mines, avril 2019
- Benbouzid, B., & Cardon, D. (2018). Machines à prédire. *Réseaux*, 211(5), 9-33. <https://doi.org/10.3917/res.211.0009>
- Benbouzid, B., Meneceur, Y. & Smuha, N. (2022). Quatre nuances de régulation de l'intelligence artificielle: Une cartographie des conflits de définition. *Réseaux*, 232-233, pp. 29-64. <https://doi.org.proxybib-pp.cnam.fr/10.3917/res.232.0029>
- Benhamou, S., Janin, L., Bocognano, A., Charrié, J., & Thibault, G. (2018). *Intelligence Artificielle et travail*. http://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/fs-rapport-intelligence-artificielle-28-mars-2018_0.pdf
- Bertrand B. (2023), « La régulation européenne de l'intelligence artificielle », dossier « La recherche en sciences humaines et sociales au cœur de la transition numérique », la Lettre de l'InSHS Droit, n°82, 15 mars. <https://www.inshs.cnrs.fr/fr/cnrsinfo/la-regulation-europeenne-de-lintelligence-artificielle>
- Bodrőzic, Z. and P. S. Adler (2018). The evolution of management models: a neo-Schumpeterian theory, *Administrative Science Quarterly*, 63(1), 85–129.
- Bouabdallah, K., Greenan, N., Villeval, M. C. (1999). Le biais technologique fondements, mesures et tests empiriques. *Revue française d'économie*, 14(1), 171-227.
- Bradshaw, A., Craig CBE, C., Davis, L., Gibson, H., Limon, B., McCarthy, N., Montgomery, J., O'Neil, S., Pickering, M., Vigar, D., & Watson, O. (2018). *The impact of Artificial Intelligence on work*. An

evidence synthesis on implications for individuals, communities, and societies (N° 978-0-85672-626-2). <https://www.thebritishacademy.ac.uk/sites/default/files/AI-and-work-evidence-synthesis.pdf>

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age. Work, progress, and prosperity in a time of brilliant technologies.*

Buchanan, B. G. (2005). A (very) brief history of artificial intelligence. *AI Magazine*, 26(4), 53-60.

Calisto, F. M., Santiago, C., Nunes, N., & Nascimento, J. C. (2022). BreastScreening-AI: Evaluating medical intelligent agents for human-AI interactions. *Artificial Intelligence in Medicine*, 127, 102285. <https://doi.org/10.1016/j.artmed.2022.102285>

Calvino, F., Virgillito, M. E. (2018). The innovation-employment nexus: a critical survey of theory and empirics. *Journal of Economic surveys*, 32(1), 83-117.

Cardon, D., Cointet, J.-P., & Mazières, A. (2018). La revanche des neurones. *Réseaux*, 211(5), 173-220. <https://doi.org/10.3917/res.211.0173>

Caroly, S., & Barcellini, F. (2013). Le développement de l'activité collective. In P. Falzon (Ed) *Ergonomie constructive*. PUF

Casilli, A. A. (2015). Digital Labor: Travail, technologies et conflictualités. In D. Cardon & A. A. Casilli (Éds.), *Qu'est-ce que le digital labor ?* (p. 10-42). Editions de l'INA.

Casilli, A. A. (2019). *En attendant les robots. Enquête sur le travail du clic* (Seuil). <https://www.seuil.com/ouvrage/en-attendant-les-robots-antonio-a-casilli/9782021401882>

Cazes, S. (2023), « Le dialogue social et la négociation collective à l'heure de l'intelligence artificielle », pp. 252-289, in Perspectives de l'emploi de l'OCDE 2023 : Intelligence artificielle et marché du travail, Éditions OCDE, Paris, <https://doi.org/10.1787/aae5dba0-fr>

CEPEJ. (2019). *Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement*. <https://rm.coe.int/charte-ethique-fr-pour-publication-4-decembre-2018/16808f699b>

Chen, Z. (2022). Collaboration among recruiters and artificial intelligence: Removing human prejudices in employment. *Cognition, Technology & Work*. <https://doi.org/10.1007/s10111-022-00716-0>

Choudhury, A., Asan, O., & Medow, J. E. (2022). Effect of risk, expectancy, and trust on clinicians' intent to use an artificial intelligence system—Blood Utilization Calculator. *Applied Ergonomics*, 101, 103708. <https://doi.org/10.1016/j.apergo.2022.103708>

Cirillo, V., Rinaldini, M., Staccioli, J., & Virgillito, M. E. (2021). Technology vs. workers: the case of Italy's Industry 4.0 factories. *Structural change and economic dynamics*, 56, 166-183.

CNIL. (2017). *Comment Permettre À L'Homme De Garder La Main ?* (N° 0022-3077 (Print)\n0022-3077 (Linking)). <https://doi.org/10.1152/jn.01190.2003\n01190.2003>

COE. (2017). *Automatisation, numérisation et emploi. Tome 1 : Les impacts sur le volume, la structure et la localisation de l'emploi*. <https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/coe->

rapport-tome-1-automatisation-numerisation-emploi-janvier-2017.pdf

Condé, J., & Ferguson, Y. (2023). *Usages et impacts de l'IA sur le travail, au prisme des décideurs*. LaborIA Explorer.

Coutrot, T., & Perez, C. (2022). *Redonner du sens au travail: Une aspiration révolutionnaire*. Seuil.

Crawford, K., Cattabriga, A., & Joler, V. (2023). A Perspective on AI and Data in Design: Interview With Kate Crawford. *Diid Disegno Industriale Industrial Design*, (80).

Curci Y., Greenan N., Napolitano S. (2024). Innovating for the good or for the bad. An EU-wide analysis of the impact of technological transformation on the labour market, Miméo Beyond 4.0.

Daniellou, F. (1986). *L'Opérateur, la Vanne, l'Ecran. L'Ergonomie des Salle de Controle*. Paris: Editions de l'Anact.

Daniellou F. (1992). *Le statut de la pratique et des connaissances dans l'intervention ergonomique en conception*. Thèse pour l'Habilitation à diriger des recherches, Toulouse.

Darses, F., & Reuzeau, F. (2004). 24. Participation des utilisateurs à la conception des systèmes et dispositifs de travail. In P. Falzon (Éd.), *Ergonomie* (p. 405-420). Presses universitaires de France «Hors collection». <https://doi.org/10.3917/puf.falzo.2004.01.0405>

Daugherty, P. R., & Wilson, J. (2018). *Humans + Machine : Reimagining work in the age of AI*. Harvard Business Review Press.

Davenport, T. H., & Kirby, J. (2015). Beyond automation. *Harvard Business Review*, 93(6), 58-65.

De Ganay, C., & Gillot, D. (2017). *Pour une intelligence artificielle maîtrisée, utile et démystifiée*. <https://www.senat.fr/rap/r16-464-1/r16-464-11.pdf>

De Stefano, V., & Taes, S. (2021). Management algorithmique et négociation collective. *Notes de prospective*, 10.

De Sefano V. & Doellgast V. (2023), « Introduction to the Transfer spécial issue. Regulating AI at work: labour relations, automation, and algorithmic management », *Transfer*, vol. 29, issue 1, February, pp. 9-20. <https://doi.org/10.1177/10242589231157656>

De Terssac, G., Soubie, J., & Neveu, J. (1988). Systèmes experts et transferts d'expertise. *Sociologie du travail*, 3, 461-477.

de La Rochefoucauld, M. (2020). L'IA et la reconfiguration de la fonction RH. *Management & Data Science*. <https://doi.org/10.36863/mds.a.10973>

Delgoulet, C., & Vidal-Gomel, C. (2013). Le développement des compétences: Une condition pour la construction de la santé et de la performance au travail. In P. Falzon, *Ergonomie constructive* (p. 17-32). Presses Universitaires de France. <https://doi.org/10.3917/puf.falzo.2013.01.0017>

DiMaggio, Paul J. Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, 48(2), 147–160.

- Doellgast, V., Wagner I., et O'Brady S. (2023). Negotiating limits on algorithmic management in digitalised services: cases from Germany and Norway, *Transfer: European Review of Labour and Research*, 29.1 : 105-120.
- Dubus, N. (1995). Le transfert de la connaissance pour la réalisation d'un système expert. *Deuxième Rencontres de Théo Quant*, 157-165.
- Engelbart, D. C. (1962). *Augmenting human intellect : A conceptual framework*.
- Ferguson, Y. (2019). 1. Ce que l'intelligence artificielle fait de l'homme au travail. Visite sociologique d'une entreprise. In F. Dubet (Éd.), *Les mutations du travail* (p. 23-42). La Découverte. <https://doi.org/10.3917/dec.dubet.2019.01.0023>
- Fischer, F. (2017). *Les enjeux de mise en oeuvre opérationnelle de l'intelligence artificielle dans les grandes entreprises*. Cigref. <https://www.cigref.fr/wp/wp-content/uploads/2017/10/CIGREF-Cercle-IA-2017-Mise-en-oeuvre-operationnelle-IA-en-Entreprises.pdf>
- Fischer, F., & de Broca, X. (2018). *L'intelligence artificielle en entreprise : Stratégies, gouvernances et challenges de la data intelligence*. Cigref.
- Flandrin, P., Wuidar, S., Hellemans, C., & van de Leemput, C. (2022). *Analyser l'activité à la réception pour réinterroger les choix d'organisation du travail et de conception d'un agent conversationnel. Le cas d'un hôtel indépendant « smart »*. 56ème congrès SELF.
- Frey, C. B., & Osborne, M. A. (2013). *The Future of Employment : How susceptible are jobs to computerisation?* https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf
- Frey, C. B., Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation?. *Technological forecasting and social change*, 114, 254-280.
- Gamkrelidze, T. (2022). *Des discours aux réalités de la conception, du déploiement et des usages des systèmes d'Intelligence Artificielle dans les situations de travail*.
- Gamkrelidze, T., Zouinar, M., & Barcellini, F. (2021). The "Old" Issues of the "New" Artificial Intelligence Systems in Professional Activities. In M.-E. Bobillier Chaumon (Éd.), *Digital Transformations in the Challenge of Activity and Work : Understanding and Supporting Technological Changes* (p. 71-86). Wiley-ISTE.
- Gaudart, C., Lhuillier, D., Molinié, A. F., & Waser, A. M. (2019). Santé fragilisée et construction d'un travail soutenable. *Psychologie du Travail et des Organisations*, 25(1), 1-5.
- Galey, L., Terquem, V., & Barcellini, F. (2022). A Social Design Approach: Enhancement of Local Social Dialogue on the Transformation of Work by Digital Technology. *Relations industrielles/Industrial Relations*, 77(3).
- Giblas, D., Godon, A.-S., Fargeas, M., Durantou, S., Gard, J.-C., Audier, A., Caye, J.-M., & Buffard, P.-E. (2018). *Intelligence artificielle et capital humain. Quels défis pour les entreprises ?*

- Giniaux-Kats, J., Campanini, E., Dufour, M., Forestier, F., Maury, P., & Teynier, S. (2021). Reprendre le contrôle de l'intelligence artificielle dans l'emploi. *L'Economie politique*, 92(4), 44-53.
- Gollac, M., Guyot, S., Volkoff, S. (2008). A propos du travail soutenable. *Centre d'étude de l'emploi, Rapport de recherche n°48*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. The MIT Press.
- Goos, M., and Manning A. (2007). Lousy and lovely jobs: The rising polarization of work in Britain. *The review of economics and statistics*, 89.1 (2007): 118-133.
- Gras Gentiletti, M., Bourmaud, G., Fréjus, M., & Decortis, F. (2022). Concevoir pour des activités instrumentées par des chatbots: Apports d'une approche de l'activité située et médiatisée pour la conception. *Activites*, 19(1). <https://doi.org/10.4000/activites.7428>
- Greenan N., Napolitano S., (2023a), Digital technologies, learning capacity of the organization and innovation: EU-wide empirical evidence from a combined dataset, *Industrial and Corporate Change*, dtad064, <https://doi-org.proxybib-pp.cnam.fr/10.1093/icc/dtad064>
- Greenan N, Napolitano S. (2023b), Investir dans la capacité d'apprentissage de l'organisation pour la double transition digitale et écologique, in Amossé et al. *Que sait-on du travail ?*, SciencesPo les presses, 209-221.
- Gries, T., Naudé, W. (2022). Modelling artificial intelligence in economics. *Journal for labour market research*, 56(1), 12
- Griliches, Z. (1969). Capital-skill complementarity. *The review of Economics and Statistics*, 51 :4, 465-468.
- Guérin, F., Pueyo, V., Béguin, P., Garrigou, A., Hubault, F., Maline, J., & Morlet, T. (2021). *Concevoir le travail, le défi de l'ergonomie* (Octarès). <https://www.octares.com/travail-et-activite-humaine/279-concevoir-le-travail-le-defi-de-lergonomie.html>
- Haslberger, Matthias and Gingrich, Jane and Bhatia, Jasmine, No Great Equalizer: Experimental Evidence on AI in the UK Labor Market (October 6, 2023). Available at SSRN: <https://ssrn.com/abstract=4594466> or <http://dx.doi.org/10.2139/ssrn.4594466>
- Hagemann, V., Rieth, M., Suresh, A. and Kirchner, F. (2023). Human-AI teams—Challenges for a team-centered AI at work. *Front. Artif. Intell.* 6:1252897. doi: 10.3389/frai.2023.1252897
- Henry, K. E., Kornfield, R., Sridharan, A., Linton, R. C., Groh, C., Wang, T., Wu, A., Mutlu, B., & Saria, S. (2022). Human–machine teaming is key to AI adoption: Clinicians' experiences with a deployed machine learning system. *Npj Digital Medicine*, 5(1), 97. <https://doi.org/10.1038/s41746-022-00597-7>
- Hicks, J. (1932) *The Theory of Wages*. Macmillan, London.
- Hoc, J.-M. (2000). From human-machine interaction to human-machine cooperation. *Ergonomics*, 43(7), 833-843.

- Hötte, K., Somers, M., & Theodorakopoulos, A. (2023). Technology and jobs: A systematic literature review. *Technological Forecasting and Social Change*, 194, 122750.
- IEEE. (2019). *IEEE Position Statement. Artificial Intelligence*. <https://globalpolicy.ieee.org/wp-content/uploads/2019/06/IEEE18029.pdf>
- Institut de Recherche en Informatique de Toulouse. (2001). *L'intelligence artificielle, mais enfin de quoi s'agit-il ?*
- Inrs (2022), *L'intelligence artificielle au service de la santé et de la sécurité au travail, enjeux et perspectives à l'horizon 2035*, Collection Prospectives en santé et sécurité au travail.
- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577-586. <https://doi.org/10.1016/j.bushor.2018.03.007>
- Jarrahi, M. H., Möhlmann, M., & Lee, M. K. (2023). Algorithm management: The Role of AI in Managing Workforces. *MIT Sloan Management Review*, 64(3), 1-5.
- Jarrahi, M. H., Newlands, G., Lee, M. K., Wolf, C. T., Kinder, E., & Sutherland, W. (2021). Algorithmic management in a work context. *Big Data & Society*, 8(2), 1-14. <https://doi.org/10.1177/20539517211020332>
- Johnson, M., & Vera, A. H. (2019). No Ai is an island : The case for teaming intelligence. *AI Magazine*, 40(1), 16-28. <https://doi.org/10.1609/aimag.v40i1.2842>
- Julia, L. (2019). *L'intelligence artificielle n'existe pas*. Éditions First. <https://doi.org/10.3917/pes.385.0049>
- Krämer, C. et S. Cazes (2022), « Shaping the transition : Artificial intelligence and social dialogue », *Documents de travail de l'OCDE sur les questions sociales, l'emploi et les migrations*, n° 279, Éditions OCDE, Paris, <https://doi.org/10.1787/f097c48a-en>
- Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at Work: The New Contested Terrain of Control. *Academy of Management Annals*, 14(1), 366-410. <https://doi.org/10.5465/annals.2018.0174>
- LawGeex. (2018). *Comparing the Performance of Artificial Intelligence to Human Lawyers in the Review of Standard Business Contracts*. LawGeex.
- Lamannis M. (2023), *Collective bargaining in the platform economy*, The European Trade Union Institute (ETUI), <https://www.etui.org/publications/collective-bargaining-platform-economy>
- LeCun, Y. (2019). *Quand la machine apprend. La révolution des neurones artificiels et de l'apprentissage profond* (Odile Jaco).
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436-444. <https://doi.org/10.1038/nature14539>
- Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers. *Proceedings of the 33rd Annual ACM*

- Conference on Human Factors in Computing Systems*, 1603-1612.
<https://doi.org/10.1145/2702123.2702548>
- Legg, S., & Hutter, M. (2007). Universal intelligence : A definition of machine intelligence. *Minds and Machines*, 17(4), 391-444. <https://doi.org/10.1007/s11023-007-9079-x>
- Lévy, C. (2019). Les outils big data dans les RH. Du mythe à la pratique. *Politiques de communication*, 12(1), 45-69. <https://doi.org/10.3917/pdc.012.0045>
- Licklider, J. C. R. (1960). Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, HFE-1, 4-11.
- Licoppe, C., & Dumoulin, L. (2019). Le travail des juges et les algorithmes de traitement de la jurisprudence. Premières analyses d'une expérimentation de «justice prédictive» en France. *Droit et société*, 103(3), 535-554. <https://doi.org/10.3917/drs1.103.0535>
- Maline, J., & Pretto, J. (1994). *Simuler le travail: une aide à la conduite de projet*. Agence nationale pour l'amélioration des conditions de travail.
- Martin A. (2023), Martin. L'intelligence artificielle peut-elle être saisie par le droit de l'Union européenne ? Conférence Nationale en Intelligence Artificielle 2022 (CNIA 2022), Juin, Saint-Etienne, France. hal-03866148.
- Mcafee, A., & Brynjolfsson, E. (2017). *Machine Platform Crowd : Harnessing our Digital Future*. W.W. Norton and Company.
- McCarthy, J. (2007). From here to human-level AI. *Artificial Intelligence*, 171(18), 1174-1182.
<https://doi.org/10.1016/j.artint.2007.10.009>
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A proposal for the Dartmouth summer research project on artificial intelligence*.
- McIntosh, S. (2018). AI and automation: The benefits for business and industry. Sage.
<https://www.sage.com/en-gb/blog/ai-and-automation-business/>
- Mehrizi, M. H. R., Mol, F., Peter, M., Ranschaert, E., Santos, D. P. D., Shahidi, R., Fatehi, M., & Dratsch, T. (2022). *How are radiologists' decisions impacted by AI suggestions? Moderating effect of explainability inputs and attitudinal priming in examining mammograms* [Preprint]. In Review.
<https://doi.org/10.21203/rs.3.rs-2331488/v1>
- Menant, L. (2021). *L'acceptabilité de l'intelligence artificielle appliquée aux systèmes d'information des ressources humaines : Cas d'un agent conversationnel*.
- Meneceur, Y. (2020). *L'intelligence artificielle en procès. Plaidoyer pour une réglementation internationale et européenne*. Lefebvre Sarrut Belgium.
- Mignot, L., & Schultz, É. (2022). Les innovations d'intelligence artificielle en radiologie à l'épreuve des régulations du système de santé. *Réseaux*, 232-233(2), 65-97. <https://doi.org/10.3917/res.232.0065>

- Milanez, A. (2023), "The impact of AI on the workplace: Evidence from OECD case studies of AI implementation", *Documents de travail de l'OCDE sur les questions sociales, l'emploi et les migrations*, n° 289, Éditions OCDE, Paris, <https://doi.org/10.1787/2247ce58-en>.
- Millot, P. (1999). Systemes Homme-Machine et automatique. *Journées Doctorales d'Automatique JDA'99*.
- Millot, P., & Boy, G. A. (2012). Human-machine cooperation: A solution for life-critical Systems? *Work*, 41, 4552-4559. <https://doi.org/10.3233/WOR-2012-0033-4552>
- Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The structure of "unstructured" decision processes. *Administrative science quarterly*, 246-275.
- Molina O., Butollo F., Makó C., Godino A., Holtgrewe U., Illsoe A., Junte S., Larsen, T. P., Illésy M., Pap J. & Wotschack P. (2023), « It takes two to code: a comparative analysis of collective bargaining and artificial intelligence », *Transfer*, vol. 29, issue 1, February, pp. 1-8, <https://doi.org/10.1177/10242589231156515>.
- Mollen, A., & Hondrich, L. (2023). From risk mitigation to employee action along the Machine Learning Pipeline. *Hans-Böckler-Stiftung*, 278.
- Moore, P. V. (2019). *La SST et l'avenir du travail: Avantages et risques des outils d'intelligence artificielle sur les lieux de travail*. European Agency for Safety and Health at Work.
- Morozov, E. (2013). *To Save Everything, Click Here. The Folly of Technological Solutionism*. PublicAffairs.
- Navarro, D. F., Kocaballi, A. B., Dras, M., & Berkovsky, S. (2022). Collaboration, not Confrontation: Understanding General Practitioners' Attitudes Towards Natural Language and Text Automation in Clinical Practice. *ACM Transactions on Computer-Human Interaction*, 3569893. <https://doi.org/10.1145/3569893>
- Nerce, J., Marie, S., & Coggia, P. (2018). *(Re)Launch, Scale & Lean AI with Cloud* (N° 9788578110796). <https://doi.org/10.1017/CBO9781107415324.004>
- Norman, D. A., & Draper, S. W. (1986). *User Centered System Design: New Perspectives on Human computer Interaction*. Hillsdale: L. Erlbaum Associates.
- Norman, D. (1993). Les artefacts cognitifs. In *Raisons Pratiques* (Vol. 4, p. 15-34).
- Norman, D. (2017). Design, business models, and human-technology teamwork. *Research-Technology Management*, 60(1), 26-30.
- Ochanine, D. (1978). Le rôle des images opératives dans la régulation des activités de travail. *Psychologie et éducation*, 2(2), 63-72.
- Oppenheimer, J., Lüken, S., Hamm, B., & Niehues, S. M. (2023). A Prospective Approach to Integration of AI Fracture Detection Software in Radiographs into Clinical Workflow. *Life*, 13(1), 223. <https://doi.org/10.3390/life13010223>

- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253.
- Park, H. J., Kim, S. M., La Yun, B., Jang, M., Kim, B., Jang, J. Y., Lee, J. Y., & Lee, S. H. (2019). A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of breast masses on ultrasound : Added value for the inexperienced breast radiologist. *Medicine*, 98(3), e14146. <https://doi.org/10.1097/MD.00000000000014146>
- Perez C (2010) Technological revolutions and techno-economic paradigms. *Cambridge Journal of Economics*, 34(1), 185–202.
- Ponce del Castillo A. (2020), *Labour in the Age of AI: Why Regulation is Needed to Protect Workers*, ETUI, Bruxelles.
- Poullet Y. (2023), « IA et santé - A la lumière de quelques textes réglementaires. L'intelligence artificielle dans le secteur médical : les défis du droit face à la santé « intelligente » ? », in M-C. Piatti & ! F. Violet (dir.), *Pour des Intelligences artificielles au service du corps vulnérable*. Actes de la Journée d'étude du 3 décembre 2021., Editions des archives contemporaines, France, ISBN : 9782813004598, pp. 47-112, doi : <https://doi.org/10.17184/eac.5911>
- Rabardel, P. (1995). *Les hommes et les technologies. Une approche cognitive des instruments contemporains*. Armand Colin. <https://hal.archives-ouvertes.fr/hal-01017462>
- Rajabiyazdi, F., Jamieson, G. A., & Guanolusia, D. Q. (2022). An Empirical Study on Automation Transparency (i.e., seeing-into) of an Automated Decision Aid System for Condition-Based Maintenance. In N. L. Black, W. P. Neumann, & I. Noy (Éds.), *Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021)* (Vol. 223, p. 675-682). Springer International Publishing. https://doi.org/10.1007/978-3-030-74614-8_84
- Reverberi, C., Rigon, T., Solari, A., Hassan, C., Cherubini, P., GI Genius CADx Study Group, Antonelli, G., Awadie, H., Bernhofer, S., Carballal, S., Dinis-Ribeiro, M., Fernández-Clotett, A., Esparrach, G. F., Gralnek, I., Higasa, Y., Hirabayashi, T., Hirai, T., Iwatate, M., Kawano, M., ... Cherubini, A. (2022). Experimental evidence of effective human–AI collaboration in medical decision-making. *Scientific Reports*, 12(1), 14952. <https://doi.org/10.1038/s41598-022-18751-2>
- Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies*, 1(1), 33-36.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1, 206-215. <https://doi.org/10.1038/s42256-019-0048-x>
- Russel, S., & Norving, P. (2010). *Artificial intelligence. A modern approach*. Pearson Education, Inc. <https://doi.org/10.1017/S0269888900007724>
- Russo, M., Pavone, P., & Cetrulo, A. (2019). Conflict and participation in bargaining at company level: The Lamborghini case. *Economia & lavoro*, 53(1), 53-74.

Sage. (2020). *Artificial Intelligence in business guide 2020. How to lift productivity and profit in the age of automation.*

https://search.proquest.com/docview/1900298116?accountid=10297%0Ahttp://resolver.ebscohost.com/openurl?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQ%3Aabiglobal&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&rft.genre=article&rft.jtitle=St

Salembier, P., & Pavard, B. (2004). Analyse et modélisation des activités coopératives situées. *Activités*, 1(1), 87-99.

Searle, J. (1984). *Minds, brains and science*. Harvard University Press.

Schoonderwoerd, T. A.J., Jorritsma, W., Neerincx, M. A. and van den Bosch, K. (2021) Human-centered XAI: Developing design patterns for explanations of clinical decision support systems, *International Journal of Human-Computer Studies*, Volume 154.

Schmager, S., Pappas, Ilias, and Vassilakopoulou, P. (2023). Defining Human-Centered AI: A Comprehensive Review of HCAI Literature. Conference: 15th Mediterranean Conference on Information Systems (MCIS) and the 6th Middle East & North Africa Conference on digital Information Systems (MENACIS), Madrid, Spain.

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass.

Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems*, 10(4), 1-31. <https://doi.org/10.1145/3419764>

Shneiderman, B. (2021). Human-Centered AI. *Issues in Science and Technology* 37, no. 2 (Winter 2021): 56–61.

Shoham, Y., Perrault, R., Brynjolfsson, E., Clark, J., Manyika, J., Niebles, J. C., Lyons, T., Etchemendy, J., Grosz, B., & Bauer, Z. (2018). *Artificial Intelligence Index Annual Report* (p. 1-94). AI Index Steering Committee, Human-Centered AI Initiative, Stanford University.

Sirén-Heikel, S., Kjellman, M., & Lindén, C. (2023). At the crossroads of logics: Automating newswork with artificial intelligence — (Re)defining journalistic logics from the perspective of technologists. *Journal of the Association for Information Science and Technology*, 74(3), 354-366. <https://doi.org/10.1002/asi.24656>

Suchman, L. A. (1987). *Plans and Situated Actions. The Problem of Human-Machine Communication*. Cambridge University Press.

Sujan, M., Furniss, D., Hawkins, R. D., & Habli, I. (2020). Human Factors of Using Artificial Intelligence in Healthcare: Challenges That Stretch Across Industries. *Safety-Critical Systems Symposium*.

Timmer, I., & Rietveld, R. (2019). Rule-Based Systems for Decision Support and Decision-Making in Dutch Legal Practice. A Brief Overview of Applications and Implications. *Droit et Société*, 103(3),

517-534. <https://doi.org/10.3917/drs1.103.0517>

Tolan, S., Pesole, A., Martínez-Plumed, F., Fernández-Macías, E., Hernández-Orallo, J., & Gómez, E. (2021). Measuring the Occupational Impact of AI: Tasks, Cognitive Abilities and AI Benchmarks. *Journal of Artificial Intelligence Research*, 71, 191-236. <https://doi.org/10.1613/jair.1.12647>

Trapp, R. (1985). *Impacts of artificial intelligence*. ELSEVIER SCIENCE PUBLISHERS B.V.

Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., Janda, M., Lallas, A., Longo, C., Malvehy, J., Paoli, J., Puig, S., Rosendahl, C., Soyer, H. P., Zalaudek, I., & Kittler, H. (2020). Human-computer collaboration for skin cancer recognition. *Nature Medicine*, 26(8), 1229-1234. <https://doi.org/10.1038/s41591-020-0942-0>

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433-460.

Van Belleghem, L. (2018). La simulation de l'activité en conception ergonomique: acquis et perspectives. *Activités*, 15(15-1).

Van Den Bosch, K., & Bronkhorst, A. (2018). Human-AI Cooperation to Benefit Military Decision Making. In *STO Nato* (p. 1-12).

Velkovska, J., & Relieu, M. (2021). Pour une conception «située» de l'intelligence artificielle. Des interactions hybrides aux configurations socio-techniques. *Reseaux*, 229(5), 215-229. <https://doi.org/10.3917/res.229.0215>

Villani, C. (2018). *Donner un sens à l'intelligence artificielle* (N° 978-2-11-145708-9). https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

Vivarelli, M. (2014) Innovation, employment and skills in advanced and developing countries: a survey of economic literature. *Journal of Economic Issues*, 48(1): 123–154.

Volkoff, S. (2005). *L'ergonomie et les chiffres de la santé au travail: ressources, tensions et pièges*. Toulouse, Octarès.

Wang, X., Lin, X., & Shao, B. (2023). Artificial intelligence changes the way we work: A close look at innovating with chatbots. *Journal of the Association for Information Science and Technology*, 74(3), 339-353. <https://doi.org/10.1002/asi.24621>

Williams, C. (1983). A brief introduction to artificial intelligence. *Proceedings OCEANS'83*, 94-99.

Willis, M., & Jarrahi, M. H. (2019). Automating Documentation: A Critical Perspective into the Role of Artificial Intelligence in Clinical Documentation. *Information in Contemporary Society: 14th International Conference, IConference 2019, 11420*, 200-209. https://doi.org/10.1007/978-3-030-15742-5_19

Wilson, H. J., Daugherty, P. R., & Morini-Bianzino, N. (2017). The jobs that Artificial Intelligence will create. *Mit Sloan Management Review*, 58(4), 14-16. <https://doi.org/10.1016/j.trf.2006.08.002>

Woods, D. D. (1996). Decomposing Automation: Apparent Simplicity, Real Complexity. In R. Parasuraman & M. Mouloua (Éds.), *Automation and Human Performance: Theory and Applications*

(Erlbaum, p. 3-17).

Woods, D. D., Roth, E. M., & Bennett, K. B. (1990). Explorations in Joint Human-Machine Cognitive systems. In *Human-Machine cognitive systems* (p. 123-158). <https://doi.org/10.1108/jd.2008.27864fae.001>

Xu, W. (2019). Toward human-centered AI: A perspective from human-computer interaction. *Interactions*, 26(4), 42-46. <https://doi.org/10.1145/3328485>

Xu, W., Dainoff, M., Ge, L., & Gao, Z. (2021). *From Human-Computer Interaction to Human-AI Interaction : New Challenges and Opportunities for Enabling Human-Centered AI*.

Zouinar, M. (2020). Evolutions de l'Intelligence Artificielle et travail: Quels enjeux pour l'activité humaine et la relation Humain-Machine? *Activités*, 17(1).

Zuboff, S. (2019). *The Age of Surveillance Capitalism : The Fight for a Human Future at the New Frontier of Power*. Profile Books.

Annexes

Annexe 1 Méthodologie de recherche et d'analyse systématique des études empiriques

L'analyse que nous avons mise en œuvre met l'accent sur des études empiriques portant sur le développement, les usages et le déploiement des systèmes d'IA. Les études sont considérées comme empiriques quand elles sont fondées sur des observations directes, des expériences, des enquêtes ou d'autres méthodes de collecte des données concrètes et vérifiables dans des situations réelles.

Une première étape a consisté à réaliser une recherche bibliographique systématique dans des revues du champ de l'ergonomie, de l'interaction humains-machine et en informatique spécialisée en IA avec des mots-clés et leurs combinaisons (en anglais et en français selon la langue de la revue), a priori sans limitation de période (qui s'étend généralement des années 1980 à l'année 2023) :

Tableau 6. Combinaisons des mots-clés utilisées pour la recherche bibliographique

IA OU intelligence artificielle	AI OR Artificial intelligence
IA OU Intelligence artificielle ET travail	AI OR Artificial intelligence AND work
IA OU Intelligence artificielle ET qualité du travail	AI OR Artificial intelligence AND work quality
IA OU Intelligence artificielle ET santé au travail	AI OR Artificial intelligence AND occupational health
IA OU Intelligence artificielle ET futur du travail	AI OR Artificial intelligence AND futur of work
Robots ET travail	Robots AND work
Robots ET qualité du travail	Robots AND work quality
Robots ET santé au travail	Robots AND occupational health
Robots ET futur du travail	Robots AND future of work
Système expert ET travail	Expert system AND work
Système expert ET qualité du travail	Expert system AND work quality
Système expert ET santé au travail	Expert system AND occupational health
Système expert ET futur du travail	Expert system AND future of work
Algorithme ET travail	Algorithm AND work

Algorithme ET qualité du travail	Algorithm AND work quality
Algorithme ET santé au travail	Algorithm AND occupational health
Algorithme ET futur du travail	Algorithm AND future of work

Les premières recherches mobilisant ces mots-clés ont été effectuées dans les revues suivantes : ACM Transactions on Computer-Human Interaction (TOCHI), ACM Transactions on Information Systems, Activités, Applied Artificial Intelligence, Applied Ergonomics, British Journal of Management, Cognition, Technology & Work, Ergonomics, Human Factors, International Journal of Industrial Ergonomics, Journal of the Association for Information Science & Technology, Journal of Artificial Intelligence Research, Pistes-Perspectives Interdisciplinaires sur le Travail et la Santé, Le travail humain.

Tableau 7. Exemple de nombre d'articles recherché par revue et avec des mots-clés

	ACM Transactions on Computer-Human Interaction (TOCHI)	ACM Transactions on Information Systems	Activités
IA OU intelligence artificielle	397	317	23
IA OU Intelligence artificielle ET travail	396	317	20
IA OU Intelligence artificielle ET qualité du travail	0	2	0
IA OU Intelligence artificielle ET santé au travail	0	0	2
IA OU Intelligence artificielle ET futur du travail	0	0	1
Robots ET travail	131	55	16
Robots ET qualité du travail	0	0	2
Robots ET santé au travail	0	0	4
Robots ET futur du travail	0	0	0
Système expert ET travail	6	29	4
Système expert ET qualité du travail	0	1	0
Système expert ET santé au travail	0	0	0

Système expert ET futur du travail	0	0	0
Algorithme ET travail	373	853	22
Algorithme ET qualité du travail	1	2	2
Algorithme ET santé au travail	0	0	2
Algorithme ET futur du travail	0	0	0

Bien que ces recherches aboutissent à des milliers d'articles, notamment avec des mots-clés « Intelligence artificielle », peu d'entre eux sont réellement pertinents pour étudier les questions des transformations du travail. Effectivement, la littérature et la recherche sur l'IA sont abondantes et ne cessent d'augmenter. Toutefois, soit la plupart de ces publications traitent les questions générales et techniques du développement des algorithmes de l'IA, soit elles ont tendance à s'intéresser aux questions d'interaction Humain-Machine, d'explicabilité et de confiance ou encore d'éthique plus qu'à la question du travail de façon plus large et telle qu'elle nous intéresse dans ce rapport. De plus, certains mots-clés comme la qualité du travail ou la santé au travail donnent très peu, voire aucun résultat selon les revues. Cela peut donner une indication sur un manque de prise en compte de ces thématiques dans les études relatives à l'IA et ses conséquences.

Une quarantaine d'articles, traitant la question des usages et des effets par exemple sur la performance en référence à l'IA, ont été identifiés et consultés, en particulier dans les revues suivantes : ACM Transactions on Computer-Human Interaction (TOCHI), Activités, Applied Artificial Intelligence, Applied Ergonomics, British Journal of Management, Cognition, Technology & Work, Ergonomics, Human Factors, Journal of the Association for Information Science & Technology.

Nous avons complété cette première vague de recherche systématique par une vague de recherche plus globale et diversifiée avec des bases de données scientifiques comme Cairn, Sage, Science Direct, OpenEdition Journals, Persée, Wiley et avec le moteur de recherche de Google Scholar. Les principales combinaisons des mots-clés utilisés pour ces recherches élargies étaient « Intelligence Artificielle » et « Travail ». L'objectif était d'accéder à plus d'articles potentiellement pertinents pour l'analyse des transformations du travail. Ces recherches ont permis de compléter les articles déjà collectés.

Nous avons retenu 27 articles empiriques – les autres articles et ressources ont été utilisés afin de réaliser cette analyse en s'intéressant aux éléments suivants, mentionnés ou non dans ces articles et la manière dont ils sont abordés :

- Le secteur d'activité ;
- Le type d'application de l'IA, ses fonctionnalités ou des scénarios d'utilisation ;
- Les utilisateurs travailleurs concernés ;
- Les motivations, les objectifs et les promesses de l'introduction d'un système d'IA (si tel est le cas) et/ou les objectifs de l'étude ;

- La démarche de conduite de projet ;
- Les méthodes et les évaluations réalisées ;
- Les principaux résultats ;
- Le type de situation considérée pour produire des connaissances : situation expérimentale en « laboratoire » avec ou sans professionnels du domaine, situation expérimentale dans un environnement réel de travail, situation opérationnelle. Ceci est essentiel notamment pour qualifier la portée des résultats.

Le tableau détaillé du contenu de ces éléments pour chaque article est présenté ci-dessous.

Tableau 8. Articles mobilisés pour construire le rapport

		Situation expérimentale en « laboratoire » avec professionnels du domaine	Situation expérimentale en « laboratoire » sans professionnels du domaine	Situation à base des scénarios et/ou Points de vue des acteurs	Situation expérimentale dans un environnement réel de travail	Situation opérationnelle
Professionnels de santé	Médecins généralistes			<p>Collaboration, not Confrontation: Understanding General Practitioners' Attitudes Towards Natural Language and Text Automation in Clinical Practice (Navarro et al., 2022)</p> <p>Automating Documentation: A critical perspective into the role of artificial intelligence in clinical documentation (Willis & Jarrahi, 2019)</p>		
	Médecins spécialistes (radiologues...)	<p>BreastScreening-AI: Evaluating medical intelligent agents for human-AI interactions (Calisto et al., 2022)</p> <p>Experimental evidence of effective human-AI collaboration in medical decision-making (Reverberi et al., 2022)</p> <p>Human-computer collaboration for skin cancer recognition (Tschandl et al., 2020)</p> <p>How are radiologists' decisions impacted by AI suggestions?</p>		<p>Les innovations d'intelligence artificielle en radiologie à l'épreuve des régulations du système de santé (Mignot & Schultz, 2022)</p>	<p>A Prospective Approach to Integration of AI Fracture Detection Software in Radiographs into Clinical Workflow (Oppenheimer et al., 2023)</p>	<p>L'intelligence artificielle à l'épreuve des savoirs tacites. Analyse des pratiques d'utilisation d'un outil d'aide à la détection en radiologie (Anichini & Geffroy, 2021)</p>

		<p>Moderating effect of explainability inputs and attitudinal priming in examining mammograms (Mehrizi et al., 2022)</p> <p>A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of breast masses on ultrasound Added value for the inexperienced breast radiologist (Park et al., 2019)</p>				
	Divers métiers de santé			Human Factors of Using Artificial Intelligence in Healthcare: Challenges That Stretch Across Industries (Sujan et al., 2020)		<p>Effect of risk, expectancy, and trust on clinicians' intent to use an artificial intelligence system -- Blood Utilization Calculator (Choudhury et al., 2022)</p> <p>Human-machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system (Henry et al., 2022)</p>
Professionnels du droit	Avocats	Comparing the Performance of Artificial Intelligence to Human Lawyers in the Review of Standard Business Contracts (LawGeex, 2018)				
	Juges				Le travail des juges et les algorithmes de traitement de la jurisprudence. Premières analyses d'une expérimentation de « justice prédictive » en	

					France (Licoppe & Dumoulin, 2019)	
	Divers métiers juridiques			Rule-Based Systems for Decision Support and Decision Making in Dutch Legal Practice. A Brief Overview of Applications and Implications (Timmer & Rietveld, 2019)		Concevoir pour des activités instrumentées par des chatbots Apports d'une approche de l'activité située et médiatisée pour la conception (Gras Gentiletti et al., 2022)
Professionnels de RH				Collaboration among recruiters and artificial intelligence: removing human prejudices in employment (Chen, 2022) L'IA et la reconfiguration de la fonction RH (de La Rochefoucauld, 2020)		Les outils big data dans les RH. Du mythe à la pratique (Lévy, 2019) Concevoir pour des activités instrumentées par des chatbots Apports d'une approche de l'activité située et médiatisée pour la conception (Gras Gentiletti et al., 2022) L'acceptabilité de l'intelligence artificielle appliquée aux systèmes d'information des ressources humaines: cas d'un agent conversationnel (Menant, 2021)
Journalistes				At the crossroads of logics: Automating newswork with artificial intelligence—(Re)defining journalistic logics from the perspective of technologists (Sirén-Heikel et al., 2023)		

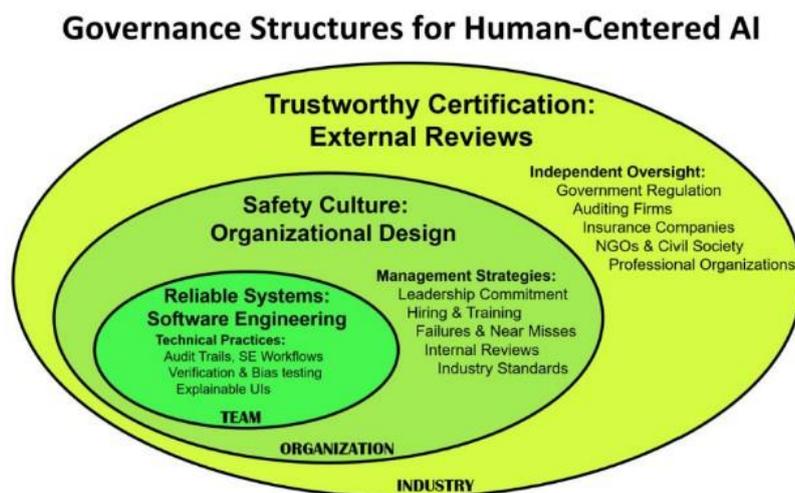
Enseignants - Chercheurs		Acceptance of Artificial Intelligence Application in the PostCovid Era and Its Impact on Faculty Members' Occupational Well-being and Teaching Self Efficacy: A Path Analysis Using the UTAUT 2 Model (Alhwaiti, 2023)				Concevoir pour des activités instrumentées par des chatbots Apports d'une approche de l'activité située et médiatisée pour la conception (Gras Gentiletti et al., 2022)
Divers employés de bureau, de service client, de marketing...				Ce que l'intelligence artificielle fait de l'homme au travail. Visite sociologique d'une entreprise (Ferguson, 2019)		Analyser l'activité à la réception pour réinterroger les choix d'organisation du travail et de conception d'un agent conversationnel. Le cas d'un hôtel indépendant « smart » (Flandrin et al., 2022) Artificial intelligence changes the way we work: A close look at innovating with chatbots (Wang et al., 2023)
Participants non professionnels			An Empirical Study on Automation Transparency (i.e., seeing-into) of an Automated Decision Aid System for Condition-based Maintenance (Rajabiyazdi et al., 2022)			

Annexe 2. Méthodologies de conception d'IA « centrées humain »³⁶

(Shneiderman, 2020) propose les mesures pratiques pour une gouvernance de l'IA centrée humain avec trois niveaux (Figure 6) :

- des systèmes fiables basés sur des pratiques d'ingénierie logicielle solides, des « bonnes pratiques »,
- une culture de la sécurité grâce à des stratégies de gestion d'entreprise éprouvées,
- une certification digne de confiance par une surveillance indépendante.

Figure 6. Les mesures de gouvernance pour l'IA centrée humain (Shneiderman, 2020)



(Shneiderman, 2020) soutient l'idée qu'il y a eu un changement de paradigme : si dans le passé, les recherches en IA ont été centrées davantage sur la construction des machines autonomes plutôt que sur le contrôle humain des interfaces, aujourd'hui, l'approche *Human-Centered AI (HCAI)* place les utilisateurs humains au cœur de la réflexion sur la conception, en particulier sur la conception de l'expérience utilisateur. Ainsi, l'approche HCAI serait centrée sur les systèmes fiables, sûrs et dignes de confiance (Shneiderman, 2020).

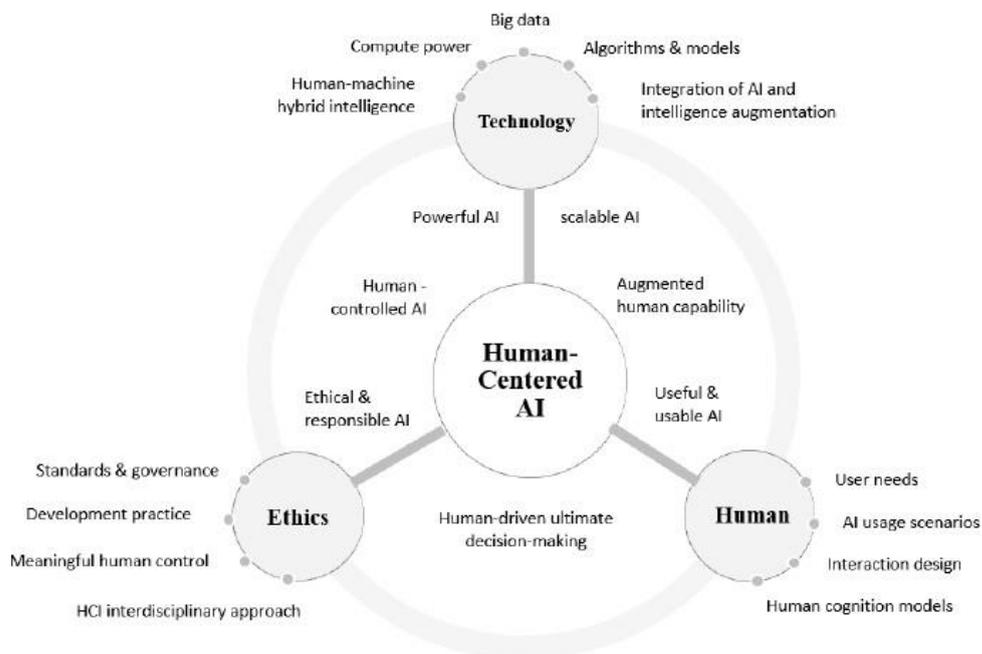
(Xu, 2019; Xu et al., 2021) proposent également de développer cette approche comprehensive « Human-Centered AI » (HAI ou HCAI)³⁷. Les principes de base de cette approche suivent trois objectifs : refléter techniquement les caractéristiques de l'intelligence humaine pour mieux la comprendre, soutenir plutôt que remplacer les capacités humaines (pour répondre aux enjeux éthiques)

³⁶ Cette section est entièrement issue de la thèse de Gamkrelidze (2022).

³⁷ HAI est aussi un institut de recherche créé récemment par l'université de Stanford, Berkley et MIT.

et porter une attention particulière aux impacts de l'IA sur les humains (concevoir les systèmes d'IA explicables, utiles et utilisables). Ainsi, cette approche inclut trois éléments de base interdépendants (Xu et al., 2021) : technologie, humain et éthique (Figure 7). L'entité « humain » renvoie à la démarche qui se base sur des besoins des humains pour l'identification des scénarios d'utilisation des systèmes d'IA. L'objectif est de développer des systèmes d'IA utiles et utilisables avec les humains qui doivent garder le rôle de décideur. L'entité « technologie » détaille la combinaison des technologies d'IA, de l'intelligence humaine augmentée par la technologie et l'intelligence hybride du système humain-machine. L'objectif poursuivi dans cette entité est le développement des systèmes d'IA évolutifs qui augmentent les capacités humaines au lieu de les remplacer et qui sont contrôlés par l'humain. Enfin, la dernière entité « éthique » doit permettre de garantir l'équité, la responsabilité et le pouvoir décisionnel de l'humain pour une conception d'une IA éthique et responsable.

Figure 7. The Human-Centered AI (HCAI) Framework (Xu et al., 2021)



Inspirée des approches de l'interaction Humain-Machine (IHM) (notamment, l'expérience utilisateur ou encore la conception centrée utilisateur), la conception de l'IA selon l'approche HCAI est surtout pensée avec ses critères d'explicabilité, d'utilité et d'utilisabilité.

L'approche HCAI se caractérise pour (Xu et al., 2021) par :

- Une place centrale de l'humain qui consiste au fait de veiller à ce que les humains soient maintenus au centre des systèmes d'IA hybrides humain-machine, de garantir le contrôle humain sur les systèmes d'IA utiles, utilisables et explicables mais aussi la conception des systèmes éthiques et humanistes.

- La complémentarité entre l'intelligence humaine et l'intelligence de la machine : utiliser la complémentarité pour compenser les déficiences du développement de l'IA censée simuler l'intelligence humaine. L'objectif est d'améliorer l'intelligence des systèmes d'IA à l'aide de « l'intelligence hybride homme-machine » (Johnson & Vera, 2019).
- L'interdépendance de l'humain, de la technologie et de l'éthique : pour concevoir des systèmes d'IA centrés sur l'humain (donc utile et utilisable), il faut prendre en compte l'impact de l'IA sur les humains, par exemple les enjeux éthiques.
- Une conception systématique : cette conception suppose de développer la complémentarité de l'intelligence humaine et de l'IA dans une perspective du système humain-machine-environnement. La conception des systèmes d'IA n'est pas envisagée comme étant uniquement un projet technologique. Elle nécessite donc la mobilisation de diverses disciplines.

En se basant sur cette approche HCAI, l'auteur propose sa mise en œuvre à travers un domaine interdisciplinaire émergent qu'il qualifie de « Humain-AI Interaction » (HAI) (Xu et al., 2021). Ce domaine HAI comprend l'étude de l'interaction et la collaboration humain-IA, l'augmentation de l'intelligence humaine, l'intelligence hybride homme-machine, l'IA explicable, l'autonomie contrôlée par l'homme, la conception éthique et l'interaction intelligente, en mobilisant des méthodes de diverses disciplines mais notamment celles de l'IHM, avec l'objectif de promouvoir et de développer des systèmes d'IA centrés sur l'humain. »

Résumé

Ce rapport propose une analyse critique et empirique des promesses et craintes associées aux transformations du travail et de l'emploi liées à l'IA. Il propose un cadrage de ce que l'on appelle Intelligence Artificielle, de ses limites techniques (par exemple opacité) et le revers de ses promesses (par exemple mirage de la collaboration humain-technologie). Nous montrons que les approches économiques dominantes qui évaluent les conséquences des IA sur l'emploi prennent appui : sur un modèle de la production où les choix d'organisation du travail ne sont pas considérés, et sur un modèle du travail pensé comme une liste de tâches éloignée du réel du travail. Dans ce cadre, les IA viennent automatiser les tâches existantes avec des effets destructeurs sur l'emploi, et certains de ces effets sont invisibilisés (délocalisation, sous-traitance, départs naturels, réduction des effectifs déjà engagée). Par ailleurs, les travaux sur la régulation de l'IA soulignent la nécessité de penser une régulation d'ensemble pour pallier les incohérences et les lacunes des instruments juridiques actuels. De fait, malgré un nombre croissant de publications portant sur l'IA, les travaux font relativement peu de place aux évolutions du travail et de l'emploi en lien avec l'IA explorés des approches systémiques et situées. Beaucoup d'études restent expérimentales, et elles s'inscrivent encore trop dans une approche déterministe de la technologie qui néglige le rôle des choix organisationnels, de la conduite de projet, et de la place des activités et de l'expérience des professionnels qui ont pourtant un rôle central dans le développement des usages de l'IA. Enfin, beaucoup de publications sont consacrées uniquement à certains secteurs professionnels (par exemple la médecine) ou fonctionnalités d'IA emblématiques (par exemple, aide au diagnostic ou recherche et traitement d'informations, génération de contenu). Du côté des transformations du travail, les travaux peignent un tableau nuancé des apports réels de l'IA – ses transformations sont dépendantes des choix organisationnels et si les études expérimentales montrent des bénéfices sur la performance, ces résultats restent dépendant de l'expertise des professionnels et cantonnés à certaines tâches (par exemple les tâches de diagnostic ou de recherche d'informations). Il existe des effets négatifs de l'IA dans certains contextes réels de travail (par exemple charge de travail accrue, sens du travail affecté, entrave à la coopération et au développement de l'expérience professionnelle) et l'opacité des systèmes d'IA fait peser un risque pour la fiabilité des actions réalisées avec une IA. Par ailleurs, une place résiduelle est accordée aux professionnels et à leurs expériences dans la conduite de projets, qui sont de fait insuffisamment pensées, et la possibilité de réguler les conséquences des SIA sur le travail et l'emploi dépend beaucoup du modèle de relations professionnelles. Forts de cette analyse, nous recommandons de développer 4 piliers pour un usage soutenable de l'IA dans la sphère professionnelle : (1) la capacité d'apprentissage des organisations, (2) un dialogue social renouvelé, (3) des conduites de projet donnant une valeur à l'activité et l'expérience des professionnel et (4) la documentation d'expérimentation en situation réelle.

Mots-clés : Intelligence Artificielle, emploi, travail, soutenabilité, économie, ergonomie

