



HAL
open science

L'IA arrive et nos langues doivent prendre la vague

Mélanie Jouitteau

► **To cite this version:**

| Mélanie Jouitteau. L'IA arrive et nos langues doivent prendre la vague. 2024. hal-04464991v1

HAL Id: hal-04464991

<https://hal.science/hal-04464991v1>

Submitted on 19 Feb 2024 (v1), last revised 20 Nov 2024 (v2)

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

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



Distributed under a Creative Commons Attribution 4.0 International License

L'IA arrive et nos langues doivent prendre la vague

Mélanie Jouitteau
IKER, CNRS
melanie.jouitteau@iker.cnrs.fr¹

traduction d'un article à paraître dans sa version anglaise et galloise dans *Language and Technology in Wales Volume II*, University of Bangor, en CC-BY.

Mon propos est simple. Nos langues doivent intégrer l'IA pour survivre précisément au déploiement qui vient des applications sur données personnelles.

L'IA, comme système utilisant des modèles réglés automatiquement sur des données vastes, entre autres linguistiques, est déjà installée dans nos pratiques dès lors qu'on utilise de la saisie prédictive ou de la reconnaissance vocale sur smartphone. Ce n'est que le début de son intégration sociale. Je compare son intégration sociale à celle des téléphones portables. Je décris en 12 points comment cette révolution technologique précédente s'était installée dans nos modes de vies, et j'explique pour chacune de ces étapes comment celle de l'IA est comparable. Les étapes actuelles de développement de l'IA préparent le déploiement des applications sur données personnelles. J'explique précisément pourquoi les applications d'IA sur données personnelles présentent, de mon point de vue, le danger d'une réduction de masse de la diversité linguistique mondiale à très court terme. Si l'intégration rapide de l'IA est probablement inévitable, ce danger précis de l'effet de réduction linguistique est, lui, évitable en rendant compatibles nos langues avec ces modèles. J'appelle à une mobilisation publique, industrielle, scientifique et sociétale pour que les langues humaines passent la vague.

Avez-vous un téléphone portable ?

Je suis de la génération X qui a vu l'arrivée de ces portables et smartphones, et je suis frappée par la ressemblance entre ce que les gens disaient des portables il y a 25 ans, et ce que j'entends sur l'IA aujourd'hui. Il me semble que le corps social réagit exactement de la même manière à une révolution technologique similaire sous bien des aspects. Dans les années 90, des personnes qui ont pu être nettement réticentes aux portables en ont finalement aujourd'hui un dans la poche. J'ai listé les raisons différentes de leur volte-face en ordre chronologique de la pénétration de la technologie dans la société, qui va de l'ère pionnière à l'intégration avancée et aux points de non-retour. Souvenons-nous ensemble. Ceux qui ont vécu ce basculement, surtout ceux qui résistaient aux portables, se rappelleront chacun quel fut leur point de bascule personnel. Ceux qui observent actuellement le développement de l'IA pourront voir en parallèle à quelle étape nous en sommes.

Au commencement, c'est toujours...

1. ****L'ère des happy few : ****

1 Je remercie pour leurs relectures et discussions critiques Rayan Ziane, Loic Grobol, Reun Bideault, Delyth Prys et Milan Rezac. Toute mésinterprétation éventuelle reste mienne.

Quelques enthousiastes au pouvoir économique et culturel technologique important adoptent rapidement la technologie nouvelle, qui y gagne une valeur symbolique de statut social élevé. Si la technologie reste inabordable pour le plus grand nombre, l'ère des happy few s'installe en longueur. Dans le cas contraire, on dit que la technologie se « démocratise ». Cela n'a rien de démocratique, cela veut juste dire que l'objet n'est plus l'exclusivité d'une élite. Il n'y a pas d'élections par l'ensemble de la population, pas de référendum sur l'adoption d'une nouveauté technologique. Les marchés mènent le jeu. Ce fut indéniablement le cas des portables. C'est largement le cas de l'IA depuis ChatGPT, et ce d'autant plus que le chatbot est d'accès gratuit pour une partie de ses services, ce qui n'était pas vrai des portables.

2. ****Exigences professionnelles : ****

La technologie vient soulager quelques tâches professionnelles, et pénètre par ces métiers. L'utilisation du portable a permis une joignabilité de plus en plus permanente pour les tâches liées au travail, joignabilité qui a été appréciée, puis requise, dans de nombreux emplois et professions, en commençant par les professions médicales comme les médecins de garde. Résister aux portables a signifié quitter ces métiers, dans un contexte de chômage de masse. L'IA offre déjà un avantage professionnel dans de nombreux métiers (métiers du code ou du texte, services clients, etc.). Son efficacité y prépare elle-même le contexte du chômage de masse.

3. ****Réduction de dysfonctionnements lourds : ****

« C'est dangereux et ça ne marche pas! », disait-on communément des portables. Effectivement, au début, la portabilité était risible car les portables ne captaient que sur des petits périmètres, puis uniquement dans des grands centres urbains. Le groupe d'utilisateurs a cependant grandi avec son réseau, malgré les soupçons lourds répétés sur les dangers physiques des ondes émises par les premiers modèles. Il est important de souligner que l'outil n'a jamais éradiqué ces deux problèmes initiaux : (1) des utilisateurs sensibles aux ondes s'exilent toujours dans les rares zones blanches pour leur propre santé sans que la société ne s'en émeuve particulièrement, et (2) votre téléphone ne capte toujours pas partout non plus. Vous voulez cependant ce portable car il capte « assez ». De façon comparable, en 2024, l'IA produit des réponses parfois aberrantes, et il n'est pas sur à ce jour que l'on puisse jamais se débarrasser de ces « hallucinations ». Les développeurs de l'IA savent qu'ils n'ont pas besoin de livrer une technologie parfaite. Ils vont juste chercher à ce que cela vous serve « assez ». Cela fonctionne déjà « assez » pour que la planète entière nourrisse les interfaces comme chatGPT de masses de données personnelles et industrielles, que les capitaux affluent pour réduire les dysfonctionnements.

4. ****Peurs et situations d'urgence : ****

La peur est un puissant levier de changement social. Les communications d'urgence et l'accès à des informations importantes en nomade ont vaincu des réticences parfois profondes. Les avantages potentiels en cas d'urgence, tels que l'accès immédiat à de l'aide ou à des informations vitales, peuvent primer sur beaucoup de choses, pour la sécurité de soi ou de ceux que l'on aime, surtout les plus fragiles d'entre-nous.

Il n'y a pas lieu de douter que les applications d'IA sont présentées en bienfaitrices de notre santé et de notre sécurité comme le font les montres connectées aux paramètres vitaux, avec une attention toute particulière aux enfants, aux personnes âgées ou en situation de handicap. Cette dernière population est d'ailleurs régulièrement en première ligne des tests d'intégration technologiques les plus près du corps, les implants, ce qui peut vaincre très prosaïquement pour tous des tabous sociétaux très profonds sur l'intégration physique des technologies.

5. ****Promesse de plus de contact humain : ****

Les humains sont passionnés de parler aux autres humains, et les nouvelles technologies entrent par là. Les smartphones ont apporté une dimension d'instantanéité à notre communication avec la famille et les amis par les messageries et les réseaux sociaux, avec une promesse de plus de contact humain, sur un tempo accéléré, un rythme et une constance insoutenables pour un téléphone mural. La FOMO (*Fear of Missing Out*) a accentué cet effet. Pour les plus jeunes, c'était la peur de rater les événements sociaux et les mises à jour des réseaux amicaux. Pour leurs parents, la peur de perdre le contact avec eux. En contexte de chômage de masse, dans ces années d'avant l'internet globalisé, chez les adultes c'est l'offre d'emploi par contact téléphonique qui devenait inratable, avant même l'omniprésence des applis de rencontre. Du côté de l'IA, loin de vouloir nous éloigner des humains que nous aimons, c'est un nouveau mode d'accès aux humains qui nous est proposé. Les chat-bots de l'IA livrent du langage humain généré à la demande, sans fatigue ni réticences. Ils peuvent produire de la matière spécifiquement humaine, le langage, plus que n'importe quel humain réel sur terre, à un rythme et une constance insoutenables pour un humain réel.

C'est cette promesse de plus de contacts humains qui nous séduit dans ces technologies. Et c'est aussi la raison pour laquelle les technologies captent autant de nos données personnelles car, massivement, nous les appelons au cœur de nos relations les plus intimes. Certes, nos concitoyen·ne·s sont de mieux en mieux éduqué.e.s à la prudence vis-à-vis de leurs données personnelles, et d'aucun concluent que l'avancée de l'intégration des technologies instruit en même temps une résistance sociale à leur égard. Je pense que c'est confondre la résistance à l'installation de technologies et la modification de nos cultures humaines en fonction de cette technologie. Loin de constituer une résistance, cette adaptation culturelle favorise plutôt son intégration paisible. Elle prépare la faisabilité des applis sur données personnelles.

Les applications d'IA à qui nous allons vouloir offrir nos données intimes et personnelles
C'est à cette étape que les locuteurs de langues non-compatibles avec l'IA commencent à être pénalisés.

6. ****Commodité et efficacité : ****

« J'aime pas ces trucs, mais j'aime pas la paperasse non plus ». Certains d'entre nous ont revendiqué la paresse pour passer au portable, car des tâches ennuyeuses pouvaient y être prises en charge de façon plus efficace. Les smartphones ont en effet offert une large gamme d'applications et de services qui rationalisent les activités quotidiennes, de la navigation spatiale avec le GPS, la navigation temporelle avec les agendas intégrés, et jusqu'aux opérations bancaires.

Les applications d'IA qui s'alimenteront de vos données personnelles sont justement conçues spécifiquement pour combler ce souhait de gestion intégrée d'une myriade de tâches fastidieuses. Dans le domaine de la santé, l'accès à vos données personnelles à la source promet de pouvoir croiser vos données corporelles en temps réel, votre alimentation, vos sources de stress quantifiées et analysées, votre profil génétique et le temps qu'il fait pour fournir à votre médecin des pré-diagnosics que son IA à elle pourra analyser. Comment ne pas vouloir cette efficacité ?

7. ****Centralisation de l'outillage : ****

L'intégration de divers services et fonctionnalités dans les smartphones, tels que le réveil, l'appareil photo, le courrier électronique, la messagerie, les médias sociaux, sans parler des horaires des

marées, a donné l'impression de se libérer d'un fatras d'outils, tous finalement remplaçables par le portable.

Les applications d'IA sur données personnelles vont proposer de se centrer sur vous, de se modeler sur vos intérêts propres, documentés sur vous-mêmes, et de gérer selon votre style et votre rythme cette myriade d'applis. Vous libérer, en somme, du fatras d'applis... de votre portable. Je ne vous vois pas résister.

8. ****Accès à l'information :****

Couplé avec internet, la grande quantité d'informations disponibles à portée de main via les smartphones a été à lui seul un puissant facteur de motivation. Des personnes initialement résistantes aux portables ont pu peser leurs réticences contre l'intérêt d'avoir un accès instantané aux actualités, aux connaissances et à d'autres ressources, ce qui les a finalement amenées à adopter les smartphones.

Nous connaissons déjà le filtrage par grandes compagnies ou gouvernements, avec des algorithmes qui filtrent vos canaux d'information en amont, comme les filtres Facebook actuels. Une fois alimentée de données personnelles, l'IA promet de vous donner la main : vous pourrez filtrer l'internet entier, passé et présent selon vos propres sources de confiance, selon vos intérêts propres et non plus les filtres posés par d'autres personnes, politiques ou publicitaires (qui persistent en amont). Une phobie des serpents ? Vous pouvez ne plus jamais en voir par hasard en ligne. Un tabou culturel peu respecté par la société ? Vous pouvez surfer tranquille, votre assistant les écarte. Votre accès personnel au monde de l'information est facilité, modelé par vous-même.

Pertes de contrôle, points de non-retour, et restriction du choix individuel

9. ****Souveraineté des données, sécurité:****

Au fil du temps, les smartphones ont intégré l'authentification biométrique et le cryptage. Les progrès des paramètres de confidentialité, des autorisations d'application et la possibilité de contrôler l'accès aux données sont apparus. Les personnes qui hésitaient au départ en raison de problèmes de confidentialité se sont senties plus confiantes quant à la protection de leurs informations personnelles. Malgré le fait, entre autres, que leur géolocalisation devenait permanente. Malgré le fait qu'ils voyageaient au quotidien ou à l'autre bout du globe avec leurs données les plus bancaires ou intimes dans des « coffres fracturables » (et que certaines douanes ont tout à fait le droit de copier intégralement). Nous savons donc que si nous voulons une technologie, notre souci de sécurité peut être saturé par quelques réassurances parcellaires, nous laissant négliger les plus évidentes règles de sécurité personnelles, négliger des fuites des plus problématiques en terme de libertés fondamentales des personnes. Cela commence à ressembler au paradigme de l'addiction.

10. ****Intégration sociale :****

À mesure que les smartphones sont devenus omniprésents, la pression sociale s'est accrue. Amis, famille ou collègues peuvent fortement compter sur les smartphones pour la communication et la coordination, ce qui rend difficile pour les individus de s'abstenir sans se sentir isolés ou exclus. A ce stade, le rapport humain comprend alors nécessairement la technologie. Elle ne le prolonge plus, mais participe à le constituer. Le rapport humain est handicapé en l'absence de la technologie, démunie de l'appareil culturel pré-technologique.

11. ****Ubiquité totalisante des services :****

De nombreux services essentiels, tels que les transports, les achats, les billetteries et les services bancaires, sont désormais hautement intégrés aux applications pour smartphones. Une fois le basculement numérique effectué, les fournisseurs de service commencent à trouver coûteux pour eux les consommateurs non-conformes. Leur droit d'accès aux services est alors lentement mis en cause, avec une aggravation des fractures sociales. A cette étape, le médecin généraliste n'est plus formé pour soigner des personnes dont les données personnelles ne sont pas pré-traitées par une IA.

12. **Points de non-retour culturels et cognitifs : **

Les usages sociaux ont évolué en intégrant la présence du smartphone. Il est devenu commun, ou banalement poli, de confirmer un rdv en s'y rendant, d'en déplacer le lieu au dernier moment par préférence, ou de prévenir d'un retard même minime. Ces usages, intégrés aux rites sociaux, sont impossibles à respecter en l'absence de l'outil. Les générations qui ont vécu cette mutation peuvent se souvenir que nous pouvions nous donner des rendez-vous des mois à l'avance, et que la banale politesse était d'y apparaître sans confirmation nécessaire, exactement en lieu et heure (bien obligé.e.s). Notre capacité collective à retenir des numéros de téléphone par cœur s'est aussi évaporée dès que nous avons cédé cette tâche de mémoire à l'outil. Ce sont des points de non-retour cognitifs. Une résistance organisée demanderait alors de pouvoir s'organiser sans ces outils, ce qui devient de moins en moins accessible. La seule marge de manœuvre pour un individu est de rejeter l'outil et d'assumer l'isolation sociale que cela représente.

Cet article n'est pas une prophétie, c'est une prospective.

Dans cette histoire de notre intégration de la révolution technologique des portables, voyez-vous vraiment une différence fondamentale de contexte ou de nature technologique qui ferait la différence avec l'IA? De mon point de vue, début 2024, je pense que nous dépassons largement la phase 3. Les grandes industries préparent nettement les points suivants, et je ne vois rien qui stopperait cette évolution d'intégration de la technologie, d'une manière qui rappelle une autre révolution technologique dont j'ai déjà observé l'intégration à l'échelle de ma vie d'adulte. Si comme moi vous ne voyez pas non plus de différence fondamentale entre les pénétrabilités sociales de ces deux technologies, les portables et l'IA, il me semble raisonnable de conclure ensemble que la résistance à l'IA n'aura pas lieu. La vague arrivera rapidement dans nos réalités les plus intimes, touchant jusqu'à nos corps et nos réflexes cognitifs. Nos résistances individuelles éventuelles ne seront que des adaptations culturelles à une réalité nouvelle. Cette évolution pose et posera de multiples problèmes, que le champ scientifique est mal préparé à appréhender (Bowman 2023). Les technologies d'IA nécessitent des capacités matérielles importantes pour le traitement des données massives. Une poignée d'entreprises prises en compétition entre elles prend actuellement des décisions de grand impact sur nos vies et notre futur en tant qu'espèce, sans aucune délégation démocratique, et sans avoir les moyens en interne de peser efficacement les effets sociétaux multiples de leurs actes. Ils n'ont pas embauché beaucoup dans les sciences humaines.

En tant que linguiste, c'est le danger de réduction drastique de la diversité linguistique qui m'apparaît nettement. Pour illustrer avec un cas concret, l'État français a une seule langue officielle, le français, qui sera sans aucun doute intégré aux technologies de l'IA. Les plus de cent autres langues, non-officielles, non-protégées, des citoyen·ne·s de l'État français ne sont actuellement pas du tout équipées pour passer la barre des outils d'IA. La plupart d'entre elles sont dans un état de précarité numérique certaine, qu'on peine encore même à cartographier (Jouitteau, Kahane & Grobol 2023-present).

Il existe une croyance, répandue dans nos sociétés en dehors de quelques milieux influents mais isolés, que la résistance à l'intelligence artificielle va être forte dans la société, et que les prospectives de révolution technologique se trompent en ne prenant pas suffisamment en compte

cette résistance qu'elle va déclencher. En Bretagne, je rencontre cette croyance dans des générations très différentes et des milieux sociaux eux aussi très différents, des Monts d'Arrée aux milieux académiques high-tech des mégapoles, comme dans... les milieux high-tech des Monts d'Arrée. Cette croyance, je la respecte en tant que telle, comme croyance. Je respecte les émotions qu'elle protège, et les gens qui la ressentent. Je la pense cependant erronée dans les faits. Je pense aussi qu'elle est dangereuse quand elle ralentit les mesures que nous pouvons encore prendre pour nous adapter à cette mutation qui vient. Cette mutation est objectivement inéluctable. La croyance en une résistance sociétale forte à l'IA ralentit la prise de conscience des locuteurs sur les dangers particuliers que va leur poser le développement de l'IA si les nouveaux outils ne peuvent pas fonctionner dans leurs langues. Elle ralentit la mise en place de mesures d'urgence pour parer à ces dangers. Si ces langues ne sont pas traitables automatiquement, de façon à intégrer ces nouveaux outils, elles feront face à une réduction de pratique encore plus radicale que celle qu'elles ont accusé au vingtième siècle. La plupart disparaîtront.

Toute personne connaissant un peu l'histoire des langues minorisées au vingtième siècle devrait être actuellement en état d'alerte maximale. Voici pourquoi. Nous savons que les langues de locuteurs vivants disparaissent lorsque les parents pensent que ces langues sont dangereuses pour leurs enfants. Les langues non-IA compatibles le seront. Nous avons vu qu'il va y avoir des applications d'IA à qui nous allons vouloir, nous-mêmes, offrir nos données intimes et personnelles. Ce sont les points 6 à 8 illustrés plus haut. Si les petites langues font dysfonctionner ces applis, nous serons acculés à favoriser dans nos pratiques les langues qui y seront intégrées. C'est déjà le cas en pratique quand on dicte simplement un sms en français pour profiter de son écriture automatique plutôt que de taper ce sms en breton. Mais l'échelle de l'évitement, là, va être immense, inégalé. Lorsque la société entrera les points de non-retour, les points entre 9 et 12 de l'intégration totalisante et irréversible, la pression d'élimination augmentera. La non-conformité n'aura d'espace que dans l'exclusion sociale, puisque ces outils s'installent au cœur des liens entre humains. Quand ces applications seront essentielles à l'accès bancaire, à la justice, à l'enseignement, aux rencontres amoureuses ou encore aux services de santé, nous chercherons à protéger le futur de nos enfants en leur interdisant, nous-mêmes, à regret, la pratique de nos langues non-IA compatibles qui seront objectivement transformées en handicap social lourd, dans un grand mouvement mondial vers le monolinguisme. L'histoire dira que nous avons consenti.

Il reste entre 6.000 et 7.000 langues dans le monde. C'est notre trésor collectif, il a été produit par nos corps humains, par des milliards de cerveaux humains en interaction entre eux sur des milliers de générations. La disparition d'une langue produit des blessures profondes sur plusieurs générations, que l'on sait mal soigner. Nos générations vivantes ont l'occasion de se mobiliser pour que cette diversité ne soit pas anéantie, et que l'IA soit une chance pour le multilinguisme et la préservation de cette incroyable diversité, ce qu'elle a sans aucun doute le pouvoir d'être pour les langues rendues compatibles avec cette technologie.

Nous avons besoin d'un plan de sauvetage mondial, d'une mobilisation publique, scientifique et sociétale vivace, d'un plan de sauvetage collectifs pour nos trésors linguistiques. En pratique, les langues ont besoin d'un kit de données numériques de langue écrite et orale, mises en forme correcte pour être traitées numériquement (Abney & Bird 2010, Jouitteau 2023). Communautés parlantes, il est aujourd'hui crucial et urgent de constituer et rendre accessibles et ouverts nos corpus de langue écrite et orale, aux scientifiques comme aux industriels, avec des copyright ouverts. Vous pouvez penser que l'urgence vient de leurs actes au départ, et être très en colère. Cependant, les empêcher d'accéder au matériel d'intégration IA de vos langues constitue une résistance de dimension uniquement morale, sans efficacité à changer la donne. La résistance à l'IA ne viendra pas. C'est prendre les langues en otage qui scellerait leur sort. Les nageurs le savent, quand la vague est trop grosse, il faut plonger. Pas rester à dire que la vague ne devrait pas être là.

Les politiques de soutien linguistique doivent repenser leur arsenal de soutien en les organisant autour des corpus nécessaires. Elles doivent favoriser l'émergence de cellules transdisciplinaires de collaboration entre les gens qui savent coder, éduquer l'IA aux langues à corpus restreint, et les gens qui savent parler aux gens, récolter la matière langage et l'enrichir avec précision d'annotations standard. Les communautés parlantes doivent pouvoir être mises en capacité de construire elles-mêmes les ressources numériques nécessaires à leur inclusion dans les nouveaux outils. Quant aux communautés scientifiques qui produisent, traitent et travaillent les corpus, très très chers collègues, linguistes descriptivistes et formels, travailleurs de terrain, linguistes de corpus et d'élicitation, annotateurs de l'ombre et ingénieurs des architectures de données, nous avons une occasion historique de sauvegarder pour les sociétés humaines l'objet de notre recherche. C'est le moment d'être à la hauteur, et de leur montrer ce qu'on sait faire.

Bibliographie

- Abney, Steven et Steven Bird. 2010. [The Human Language Project : Building a Universal Corpus of the World's Languages](#), *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Sweden. Association for Computational Linguistics, (ACL), 88-97.
- Bowman, Samuel R. 2023. 'Eight Things to Know about Large Language Models', <https://arxiv.org/abs/2304.00612>, (accédé le 01 janvier 2024).
- Jouitteau, Mélanie. 2023. 'Guide de survie des langues minorisées à l'heure de l'intelligence artificielle : Appel aux communautés parlantes', *Lapurdum*, numéro spécial 6. <https://hal.science/hal-04090195v2>
- Jouitteau, Mélanie, Sylvain Kahane & Loic Grobol. 2023-présent. *Entrelangues*, seconde édition, IKER & Modyco, CNRS. <https://entrelangues.modyco.fr/>