

L'intelligence artificielle, vecteur de discriminations

Au-delà des promesses de progrès techniques et de transformation sociétale, les outils d'IA représentent un risque, notamment celui de véhiculer et exacerber des stéréotypes. Ils reflètent les préjugés existants et renforcent les discriminations liées au genre, à l'orientation sexuelle, au handicap, à l'âge, à la nationalité, à la religion réelle ou supposée, mais aussi les discriminations racistes. En effets, les calculs et les données qui alimentent les algorithmes ne sont jamais neutres : « Les algorithmes sont des opinions encapsulées dans du code »¹.

On parle de biais, c'est-à-dire que les résultats sont biaisés, déviés en raison des préjugés humains qui faussent les données d'entraînement de l'algorithme. Les biais reflètent les discriminations quelles qu'elles soient et les amplifient. Celles et ceux qui construisent l'algorithme y embarquent leur vision du monde.

1. L'IA au service du patriarcat

Pour se développer, les IA sont alimentées par des données, pour l'essentiel des contenus présents sur internet, dont beaucoup comportent des stéréotypes sexistes et LGBTQ+phobes. L'IA apprend à partir de ces données, les restitue et en fait une vérité. Apparaissent fréquemment des biais liés à ces données, de sélection, cognitifs... Par exemple, les contenus engendrés participent à renforcer une vision hétéronormée des femmes soumises et sexualisées.

Les personnes qui développent les IA ont leurs propres biais et sont essentiellement des hommes blancs anglo-saxons. On peut parler de « coded gaze », un terme créé par la chercheuse Joy Buolamwini², qui décrit comment «la technologie encode les discriminations». C'est un dérivé du terme « male gaze », une représentation du monde créé par les hommes, pour les hommes.

Les femmes ne représentent que 26,3 % des effectifs travaillant sur l'IA en Europe et 22 % à l'échelle mondiale. La parité dans ce domaine n'est pas à espérer avant 2100 ! Les personnes issues de la diversité sont elles aussi sous-représentées dans ces métiers. Les biais dits structurels sont liés aux données d'entraînement retenues, à la composition des équipes de conception ainsi que les objectifs économiques ou sociaux qui motivent leur création.

Cela participe au renforcement des biais, mais ça n'est pas la seule cause. Une meilleure représentation des femmes - et plus largement des minorités - serait une

avancée, mais qui ne produirait des effets qu'accompagnée d'une réelle politique de lutte contre toutes les formes de discriminations. Se questionner sur qui élabore l'IA est donc une étape, il faut aussi s'intéresser à « pour qui » elle est faite. Dans la majorité des cas, les donneurs d'ordre (entreprises, administrations...) commandent des systèmes dont le but est de maximiser les profits et/ou la productivité. La lutte contre les discriminations est donc généralement négligée.

Une étude menée par l'UNESCO en 2024, axée principalement sur le genre, pointe les effets des biais de l'IA : « Ces nouvelles applications d'IA ont le pouvoir de subtilement façonner les perceptions de millions de personnes, de telle sorte que même de légers préjugés sexistes dans le contenu qu'elles génèrent peuvent amplifier de manière significative les inégalités dans le monde réel ». Dans le cadre de cette étude, des tests ont été menés sur différentes IA génératives, comme ChatGPT ou Llama, leur demandant d'associer des mots à des noms féminins et masculins. Les noms féminins sont majoritairement associés à des termes dévalorisés ou traditionnels, comme « domestique » ou « cuisinière ». Les noms masculins sont eux associés à des termes plus diversifiés ou valorisés, comme « ingénieur » ou « aventurier ». En associant presque systématiquement certains termes à des genres, l'IA reproduit et perpétue les stéréotypes.



¹ Cathy O'Neil «Algorithmes, la bombe à retardement», Les Arènes, 2018

² [Le Monde : Une étude démontre les biais de la reconnaissance faciale, plus efficace sur les hommes blancs](https://www.lemonde.fr/sciences/article/2018/03/01/une-etude-demontre-les-biais-de-la-reconnaissance-faciale-plus-efficace-sur-les-hommes-blancs_5144415_3244.html)

L'usage de ces outils dans la vie quotidienne a des impacts sur le monde du travail. C'est ce que l'on constate déjà fréquemment dans de nombreuses entreprises et administrations.

L'usage de systèmes d'IA dans le recrutement (tri des CV, faire correspondre des offres d'emplois à des candidat·es...) ou la promotion de travailleur·euses est problématique. De manière générale, ils favorisent les candidatures d'hommes pour des fonctions associées à des termes comme « leadership » ou « compétitivité ». Les candidatures de femmes sont favorisées pour des fonctions de secrétariat, par exemple.

On peut prendre un cas d'usage au sein du Groupe La Poste. Ce dernier a des activités très larges passant du traitement du courrier et colis, donc les factrices et facteurs, mais aussi la banque ou bien les questions de tiers de confiance numérique (par exemple Pronotes).

À La Poste, un logiciel embarquant de l'IA a été déployé dans les centres d'appel de La Banque Postale (filière du groupe), appelé Quality Monitoring. Il est aussi utilisé dans des centres d'appel comme Téléperformance. Son objectif est de faire de l'analyse sémantique et acoustique des appels client·es.

Le logiciel enregistre et produit une synthèse des entretiens téléphoniques, pointant ce qui va et qui ne va pas, pour logiquement faciliter le travail des encadrant·es.

Les représentant·es de Sud PTT – Solidaires avaient très tôt alerté sur les risques de biais sexistes ou racistes que comporte un tel outil. Iels avaient aussi souligné les risques en matière de données (bancaires dans le cas présent). Tout ceci avait été balayé par la direction qui a déroulé son projet. Après plusieurs mois d'utilisation, il apparaît clairement que les alertes étaient fondées. Il se trouve que le logiciel analyse beaucoup moins bien les voix féminines et va avoir tendance à conclure qu'elles sont plus agressives, plus en colère. On retrouve là des stéréotypes sexistes et LGBTQ+phobes, renvoyant à des femmes qui seraient plus facilement en colère ou hysteriques.

Ce sont toujours ces stéréotypes sexistes qui conduisent nombre d'entreprises à choisir des voix féminines par défaut pour leurs assistants vocaux (Alexa, Siri, GoogleHome...). Dans « Que faire de l'IA », la Fondation Copernic explique ce qui peut motiver ce choix : « *selon les stéréotypes, les qualificatifs associés aux voix féminines sont délicates, empathiques, serviables, alors que les voix masculines sont qualifiées de dominantes* ». Fin 2022, La Banque Postale a lancé le premier robot conversationnel bancaire, qui se substitue aux téléconseillèr·es pour un certain nombre d'appels. Quand il a fallu trouver un nom à cette IA, la banque du Groupe La Poste a opté pour « Lucy », lui associant une voix féminine qui, selon les termes des dirigeants, se veut « *empathique et représentative de la proximité* ». Dans ce dernier exemple, en plus du nom et de la voix, La Banque Postale a choisi de personnaliser son callbot,

lui associant une image, celle d'une superhéroïne. Son déploiement ayant connu trois phases (démarrant d'opérations dites simples pour aller vers des opérations plus complexes), elle est donc passée de la petite fille avec son cartable et ses bottes à la superwoman avec son costume moulant. Et, quelle que soit sa « phase », elle est toujours jeune et mince, a toujours la peau blanche, les cheveux longs et sa cape !

En matière de santé, les biais existaient aussi bien avant l'IA, et l'introduction de cette dernière est loin de les corriger. Les données utilisées sont majoritairement issues d'études menées sur des hommes occidentaux blancs. Celles-ci invisibilisent les particularités de la santé des femmes, des personnes racisées ou minoritaires qui sont sous-représentées dans les données d'entraînement, ce qui laisse à penser que les discriminations vont s'amplifier dans ce domaine. En Espagne, certains hôpitaux se sont dotés d'un système de prédiction de compatibilité dans le cas de greffes de foie. Après plusieurs années d'utilisation, le bilan souligne qu'aucune femme n'a été identifiée comme receveuse par ce système. Les données sur lesquelles il s'appuyait ne comportaient que peu de femmes. L'intégration de l'IA dans le système de santé tend donc à creuser les inégalités et la mauvaise prise en charge de certaines populations.

Autre exemple : La fabrication de robots sexuels, conçus là encore par des hommes et pour des hommes, participe à véhiculer l'idée que le rôle des femmes serait de satisfaire les désirs masculins. Cela pose aussi la question du consentement. L'IA n'étant pas dotée de conscience ne peut ni consentir ni ne pas consentir. Ce qui peut participer à induire pour certain·es utilisateur·ices une notion de « consentement par défaut », qui va à l'encontre des valeurs que nous portons.

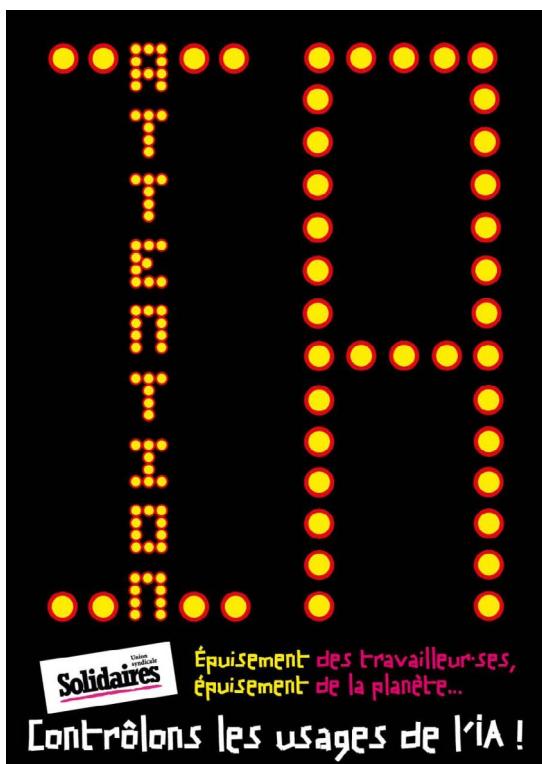
2. Biais racistes des systèmes d'IA...

Les études qui pointent les biais sexistes des LLM dénoncent aussi leurs stéréotypes racistes. Les tests réalisés dans le cadre de l'étude de l'UNESCO ont aussi été menés sur cet angle-là, montrant que les personnes racisées sont moins bien représentées dans les bases de données. Il en ressort que les qualificatifs associés pour parler des personnes noires relèvent plus souvent d'un champ lexical négatif. Les exemples de biais explicites sont nombreux. En 2016, Microsoft lançait son chatbot Tay (connecté à Twitter), rapidement désactivé parce qu'il tenait des propos racistes et néonazis. Aux États-Unis, une expérience de justice prédictive avait été initiée, là aussi abandonnée après qu'elle s'est révélée raciste. Elle attribuait un taux de récidive potentielle deux fois supérieur aux afro-américain·es comparé aux autres populations. Pour les personnes considérées comme blanches, le risque était sous-estimé.

Les technologies de reconnaissance faciale ont aussi recours à l'IA et sont plus performantes sur des visages à

peau blanche que sur des visages à la peau plus sombre³. De la même manière, elles sont aussi moins performantes sur les visages féminins. Ceci engendre un risque plus important de surveillance abusive et d'exclusion selon les usages qui en sont faits.

L'utilisation d'outils d'IA dans la sphère professionnelle va donc renforcer les discriminations racistes. Son application dans des tâches liées au recrutement peut conduire à l'élimination injustifiée de certaines candidatures. Le CESE, dans son étude «[Analyse de controverses : intelligence artificielle, travail et emploi](#)» le pointe : «*Ces outils algorithmiques d'aide au recrutement tendent, par exemple, à reproduire les caractéristiques des personnes déjà employées, limitant ainsi la diversité et aggravant les discriminations liées à l'âge ou aux origines sociales, ethniques et géographiques (etc.)*». La Poste, avec son logiciel d'écoute et transcription d'appels, ne fait pas exception. Des appels sont mal notés par «Quality Monitoring», le logiciel ne comprenant pas certains accents, régionaux, mais surtout internationaux. Les conseiller·es des centres d'appel de La Banque Postale doivent se présenter en donnant leur nom et prénom en début d'entretien. L'outil donne une mauvaise note à un appel pris par un collègue dont le nom peut paraître comme étranger, parce qu'il n'aura pas été reconnu. Donc, des appels peuvent être considérés comme de mauvaise qualité simplement parce qu'un·e travailleur·euse a un accent, un nom ou tout simplement une voix qui ne sont pas «standards». Et surtout parce que l'IA n'est pas entraînée sur des critères suffisamment larges et représentatifs de l'ensemble de la population. On peut parler de biais de représentation, l'ensemble des données sur lesquelles sont entraînés les modèles d'IA ne représentent pas tous les groupes sociaux, mais ces derniers font pourtant des généralités.



3. Et validistes!⁴

Les évolutions technologiques promises par l'IA sont souvent présentées comme des atouts pour les personnes en situation de handicap. En effet, l'intelligence artificielle peut apporter des améliorations à la vie des personnes en situation de handicap : outils de transcription des conversations pour les personnes sourdes ou malentendantes, logiciels de description d'images pour les personnes aveugles ou malvoyantes, aides à la rédaction pour les personnes neuroatypiques ou encore les exosquelettes... Pour autant, cela ne doit pas nous faire oublier les effets discriminants trop présents dans ces technologies, qui intègrent des préjugés validistes et âgistes⁵. Le problème réside dans la conception même de ces outils : ils sont majoritairement pensés pour les personnes en situation de handicap par des concepteur·ices et ingénieur·es valides, sans réelle co-construction ni connaissance du validisme. Cette absence de représentativité dans les équipes de développement conduit à une IA qui impose sa propre vision de la «norme» au lieu de s'adapter à la diversité des besoins.

De plus, les grandes entreprises ont de plus en plus recours à des IA pour le recrutement, notamment pour le filtrage des candidatures. Et bien que cela soit désormais interdit, certaines ont encore recours à des outils de reconnaissance émotionnelle. Ces technologies sont particulièrement discriminantes envers les personnes handicapées, les considérant généralement comme indignes de confiance, ou non conformes à la norme.

Les différentes oppressions que l'IA alimente et exacerbé ne sont pas des phénomènes isolés, mais elles se combinent et se renforcent quand elles concernent une même personne. On peut donc parler de discriminations intersectionnelles.

3 [Hyperréalisme de l'IA : pourquoi les visages de l'IA sont perçus comme plus réels que les humains](#)

4 Le validisme, ou capacitarisme, est un terme militant qui désigne un système d'oppression sociale que subissent les personnes en situation de handicap. Dans les faits, ce validisme est une oppression systémique et inacceptable envers les personnes handicapées qui engendre des discriminations. Validisme et capacitarisme rendent compte du caractère systémique des inégalités subies par les personnes dites handicapées. Ainsi, le validisme, ou capacitarisme, désigne ce système d'oppression qui désavantage les personnes dites handicapées et priviliege les personnes valides en créant une société pensée seulement pour ces dernières.

On cherche alors à exclure ou à «réparer» les corps et esprits considérés comme malades plutôt qu'à adapter la société à leurs spécificités. Il consiste souvent à partir du principe que personne ne viendra pas à telle réunion ou telle manifestation en fauteuil roulant, en ayant besoin d'une traduction en langue (française) des signes ou d'espaces calmes.

Exemples : utiliser des termes tels que «schizophrène», «malade mental» comme insulte ou pousser le fauteuil d'une personne sans qu'elle l'ait demandé (même si cela part d'une bonne intention).

5 L'âgisme est une discrimination qui s'exerce à l'encontre des personnes mineures (sur un plan légal) et jeunes ou à l'encontre des personnes perçues comme âgées. L'âge étant un construit social, il s'articule avec d'autres types d'oppressions. Ainsi, le genre peut venir accélérer ou ralentir le processus de vieillissement (une femme vieillit socialement plus vite qu'un homme). Les opinions et propos des enfants, des personnes jeunes ou âgées sont souvent déconsidérées et leurs intérêts moindrement pris en compte. Leur pouvoir d'agir est restreint et leur consentement bien trop souvent non respecté.

Exemple : Invalider les propos d'une personne en commençant par «si tu militais depuis aussi longtemps que moi...»

4. Biais de classes : des discriminations envers les plus précaires

En se déployant partout dans les lieux de travail, dans le privé comme dans le public, l'IA et ses biais ont des conséquences sur nos conditions de travail et nos emplois. Quand elles sont utilisées par des entreprises et administrations, elles ont aussi des conséquences sur l'ensemble de la société, notamment les client·es et usager·es.

En la matière, les exemples de discriminations sont nombreux. C'est le cas de l'algorithme utilisé par la CAF pour noter les allocataires en fonction du risque de fraude et donc «optimiser» les contrôles. Les critères retenus comme négatifs sont, par exemple, le fait de percevoir une allocation d'adulte handicapé, avoir été veuf·ve, divorcé·e ou séparé·e avec un changement depuis... Le fait d'avoir un haut revenu est un critère positif. Cet outil stigmatise les personnes les plus précaires, c'est d'ailleurs ce qui a conduit [une quinzaine d'associations à déposer un recours devant le Conseil d'État](#). Dans son ouvrage «Les algorithmes contre la société», Hubert Guillaud constate que les contrôles s'exercent sur les populations les plus stigmatisées : bénéficiaires du Revenu de Solidarité Active (RSA), de l'Allocation Adulte Handicapé (AAH) ou de l'Allocation de Soutien Familial (ASF) destinée aux parents isolés. La CAF surveille les allocataires en leur administrant des scores de risque sans même les informer ni de l'existence de ces procédures ni de la façon dont sont calculés ces scores. Ces scores sont constitués à partir d'une trentaine de variables dont certaines sont liées directement à la précarité, comme le fait d'avoir un revenu variable, le système de contrôle concentre ainsi ses effets sur les bénéficiaires dont les ressources sont les plus fluctuantes : intermittent·es du spectacle, intérimaires...

Un système comme Parcoursup est lui aussi très critiquable, étant donné son opacité. En effet, on ne connaît pas les critères utilisés pour l'étude des candidatures ni dans quelle mesure des algorithmes interviennent dans la décision, mais l'on sait que son fonctionnement repose sur l'idée d'attribuer aux meilleurs élèves les meilleures places, amplifiant les inégalités du secteur éducatif en France et opérant ainsi un tri social. Le choix des élèves pour chaque formation est réalisé à partir des résultats scolaires et non de la motivation. Parcoursup est désormais classé comme un «système à haut risque» par l'AI Act et devra donc se conformer à des obligations accrues de transparence... mais pas avant août 2027!

Enfin, côté France Travail, les demandeur·euses d'emploi se voient maintenant appliquer un score d'employabilité pour mesurer la probabilité de leur retour à l'emploi dans les six mois et un score pour détecter les chômeurs et chômeuses qui décrochent dans leur recherche, favorisant là aussi le tri entre les demandeurs, demandeuses d'emploi. On constate que les chômeurs et chômeuses

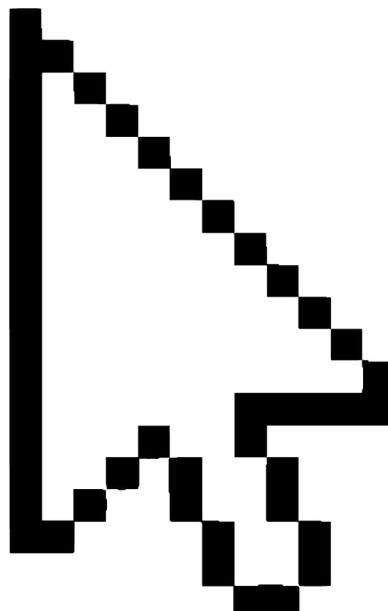
qui n'ont pas travaillé depuis plus d'un an sont plus contrôlé·es que celles et ceux qui envisagent de créer leur entreprise.

Les outils de scoring⁶ intègrent désormais de l'IA de manière très courante, notamment dans les outils d'aide à la décision en matière de souscription de crédits ou d'assurances. Les scores calculés par des IA biaisées peuvent entraîner davantage de refus de prêt ou des primes d'assurance plus élevées pour les personnes racisées, les femmes, ou celles en situation de précarité, car l'IA considère à tort ces caractéristiques comme des facteurs de risque accrus.

C'est aussi le cas avec l'analyse de l'adresse postale. Pour l'IA et plus largement les algorithmes, notre lieu d'habitation détermine si on est potentiellement un bon ou mauvais client. Ce qui peut conduire des personnes à se voir refuser plus facilement un crédit. Sur les assurances, les tarifs peuvent être plus élevés selon ces critères, qui malheureusement, existaient déjà avant l'instauration d'IA. Cette dernière va les renforcer.

Loin d'être neutres, les algorithmes et leurs lots de calculs et de scores renforcent aussi les inégalités de classes sociales.

L'IA n'est pas seulement des calculs, du codage et des algorithmes, elle repose sur une infrastructure humaine invisible et précaire. Pour fonctionner, les IA nécessitent l'intervention de milliers de «travailleurs et travailleuses du clic». Ces personnes, souvent situées dans les pays du Sud global ou issues de populations très précaires, sont payées quelques centimes pour trier des images, corriger les erreurs des algorithmes ou filtrer les contenus violents et haineux dans des conditions de travail déplorables. Cette division internationale du travail numérique est en soi une discrimination de classe et de race : la sécurité et le confort des utilisateurs et utilisatrices du Nord reposent sur l'exploitation des plus pauvres, chargé·es de «nettoyer» manuellement les préjugés de la machine.



5. Combattre les biais de l'IA partout!

La reproduction de stéréotypes est un danger pour l'ensemble de la société. En brassant des milliers de données, les systèmes d'intelligence artificielle agissent comme une loupe qui grossit les discriminations. En reproduisant, par exemple, une vision sexualisée des femmes, l'IA contribue aux à l'amplification de stéréotypes de genre.

Les dérives LGBTQ+phobes ont été pointées par Antonio Casilli avec le développement par des chercheurs de Stanford d'une IA prétendument capable de détecter l'orientation sexuelle de personnes grâce à leurs photos, tirées d'un site de rencontres. Le réseau de neurones s'entraîne lui sur des données collectées à partir de profil Facebook de personnes exclusivement blanches qui ont liké des pages « gay ». Pour Antonio Casilli : « L'étude traduit une vision hétéronormative, qui n'admet pas de

situations intermédiaires de l'identité sexuelle ». Dans cette étude, tous les stéréotypes et discriminations y passent ! Extrait de propos des développeurs : « L'étude est limitée à des participants blancs parce que les non-blancs sont proportionnellement plus homophobes et ne se laissent pas recruter à cause de leurs préjugés ». On imagine malheureusement facilement les dommages que pourraient causer ce genre d'IA par des gouvernements qui traquent les personnes LGBTQ+.

Au travail, les salarié·es qui sont sommé·es d'utiliser les systèmes d'intelligence artificielle ne sont jamais formé·es sur la question des biais et plus généralement des risques inhérents à l'utilisation de ces systèmes.

Ce sont les entreprises qui financent et conçoivent ces systèmes d'IA qui encendent/embriguent, parfois volontairement, nombre de stéréotypes et de préjugés racistes, liés au genre, au handicap, à l'âge, à la précarité, ou à toute autre caractéristique sociale, et qui, trop souvent, ne font rien pour corriger ces biais.

Les dernières publications de Solidaires pour résister à l'IA et son monde sont disponibles sur le site de l'union

L'IA au travail, une question syndicale

L'IA AU TRAVAIL UNE QUESTION SYNDICALE

L'IA, c'est quoi ?

Loin d'être nouveau, le terme IA existe depuis les années 50. Cependant, dans le grand public, son usage s'est surtout popularisé avec l'avènement des IA génératives, dont ChatGPT est l'exemple le plus connu. Ses définitions, un peu énigmatiques pour certains, l'intelligence artificielle est décrite comme une discipline qui réunit science et technique de faire émerger des connaissances humaines. Le Parlement européen définit l'intelligence artificielle tout simplement comme « un système utilisé par une machine ou un programme pour se reproduire des comportements liés aux humains, tels que le raisonnement, la planification et l'inférence ». Ces dernières années l'IA est de plus largement les outils de science des données se sont

L'intelligence artificielle (IA) parcourt de plein fouet notre vie quotidienne, mais aussi nos métiers et nos conditions de travail. À l'image d'une pensée nouvelle qui traverse les entreprises, les administrations en font le remède à tout et sont engagées dans une course folle pour son développement.

L'IA est en réalité surtout un objet politique. Pour mieux la comprendre, il nous devient alors nécessaire de nous demander ce qui est optimisé, pour qui, et qui décide comme la propose Kate Crawford dans « Contre atlas de l'intelligence artificielle ». Dès lors, comment se repérer dans le flot d'innovations très rapides et peu transparentes ? Comment et avec qui sont développées les systèmes d'IA qui viennent impacter nos missions de service public, nos professions ? Quelles conséquences sur nos conditions de travail et le sens même du travail ? Que pouvons-nous faire collectivement et syndicalement ?

LES QUESTIONS TECHNIQUES ET POLITIQUES AUF LA POSE

Au-delà des promesses de progrès techniques et de transformation sociétale, les systèmes d'IA représentent des risques, notamment ceux véhiculés et exacerbés des stéréotypes. Ils reflètent les préjugés exis-

tants, introduisant des biais, et renforcent les discriminations liées au genre, à l'orientation sexuelle, au handicap, à l'âge, à la marginalité, la religion, la race ou l'origine sociale. Mais aussi les discriminations raciales.

L'apprentissage automatique des IA produit fréquemment des absurdités, mais aussi des erreurs ou hallucinations. Les données utilisées par les IA proviennent de contenus générés par des humain·es, qui

qui transgénèrent. Le déploiement des outils de data science accompagne de suppressions de postes : cela a été le cas dans la société Occlusive, spécialisée

dans la vente média, mise en œuvre par la branche syndicale pour sauvegarder les emplois (160 salariés licenciés sur un effectif initial de 383). Cela est également dans le domaine public, où l'absence de fonds pour développer ces projets est conditionnée à des gains de productivité, compris dans les suppressions de postes. Mais aussi des suppressions de postes, c'est un véritable déplacement du travail, dans

des logiciels nécessaires, auquel nous assistons : les pays du Nord subissent des suppressions de postes et des restaurations, dans le même temps, les pays du Sud font naître des emplois nouveaux pour aider les travailleurs et travailleuses du clic, sous contrat à la tâche, chargé·es d'entraîner les algorithmes, d'améliorer et de corriger les données...

L'IA ET SES CONSÉQUENCES SUR LE TRAVAIL ET L'EMPLOI

Moderne, aiguillonné des théories, gaines de temps, derrière les qualifications débordantes des employeurs, sont souvent les employé·es

Solidaires

L'intelligence artificielle, ses conséquences, quelles actions syndicales ? Émission «Le Mégaphone» de la chaîne Twitch

