

L'intelligence artificielle dans le recrutement des ressources humaines : Pour ou contre ?

Artificial Intelligence in the Recruitment of the Human Resource: For or Against?

Ali HUSSAIN*

Université D'El Oued, Algerie
Hussain-ali@univ-eloued.dz

Mohamed LESSOUED

Université D'El Oued, Algerie
Lessoued-mohamed@univ-eloued.dz

Reçu: 02/03/2024

Accepté: 13/05/2024

Publié: 30/06/2024

Résumé:

Notre article examine le potentiel de l'intelligence artificielle à réduire ou accroître la discrimination dans les processus de recrutement. Nous démontrons que l'IA peut favoriser des résultats équitables, mais mettons aussi en lumière le problème urgent du biais algorithmique. Nous soulignons l'importance de la perception des processus décisionnels par les travailleurs et ses implications. Nos conclusions indiquent que malgré le potentiel de décisions plus justes, les individus se méfient des décisions de l'IA, d'où le besoin de transparence. L'explicabilité des algorithmes et la transparence sur les données et les choix de conception sont essentielles pour établir la confiance entre humains et machines. L'évolution de l'IA vers plus ou moins de discrimination dépendra des décisions prises lors de sa conception et utilisation. Une plus grande transparence et explicabilité des technologies favorisera une confiance accrue et une participation élargie aux processus algorithmiques. Des politiques réglementaires doivent assurer l'équité et la non-discrimination dans les processus de recrutement intégrant de l'IA.

Mots clés: intelligence artificielle, recrutement, Autosélection, discrimination, algorithme

Code Jel : M510, M540, C63, J24, J23

*Auteur correspondant

Abstract:

Our article investigates the potential for artificial intelligence to decrease or increase discrimination in recruiting procedures. We illustrate that AI can improve egalitarian outcomes, but also highlight the important issue of algorithmic bias. We stress the importance of the perception of decision-making processes by workers and its ramifications. Our findings demonstrate that despite the promise for fairer outcomes, humans distrust AI decisions, hence the necessity for openness. The explainability of algorithms and transparency around training data and design choices are vital to develop confidence between humans and machines. The progress of AI towards more or less prejudice will depend on decisions made during its creation and use. Greater transparency and explainability of technology will foster improved confidence and broader engagement in algorithmic processes. Regulatory policies must assure equity and non-discrimination in recruiting processes using AI.

Keywords : artificial intelligence, recruitment, auto-selection, discrimination, algorithm.

JEL Classification Codes : M510, M540, C63, J24, J23

Introduction :

L'intelligence artificielle (IA), en tant que technologie prédictive puissante, peut être utilisée par les entreprises pour améliorer leur efficacité dans le cadre de différents processus, dont l'un est la pratique du recrutement dans la gestion des ressources humaines ((Rieke & Bogen, 2022); (Gonzalez et al., 2019); (Tambe et al., 2019)). Étant donné qu'il existe une multitude d'opinions et de préoccupations différentes concernant la mise en œuvre de l'IA en matière de recrutement, nous souhaitons nous concentrer sur un débat spécifique : **la question de savoir si elle a le potentiel d'augmenter ou de réduire la discrimination.** La discrimination à l'embauche est un phénomène étudié depuis longtemps, qui repose souvent sur des préjugés interpersonnels inconscients des recruteurs et des jugements subjectifs de personnes sur la base de différentes caractéristiques personnelles. Dans un contexte plus large, la discrimination contribue aux inégalités sur le marché du travail (Banks & Ford, 2008) en permettant aux personnes appartenant à des groupes surreprésentés d'accéder à des opportunités d'emploi attrayantes, tout en limitant l'accès des autres groupes.

La discrimination conduit également à des inadéquations de compétences et à une allocation inefficace des ressources (Adalet McGowan, 2015). La technologie de l'IA ne connaît pas de préférences naturelles pour l'embauche de candidats ayant une apparence physique particulière, un nom spécifique indiquant la nationalité ou un diplôme d'une université particulière, pour n'en citer que quelques-uns. L'un des principaux avantages potentiels de la technologie de l'IA par rapport aux humains est qu'elle peut, dans certaines conditions, conduire à des décisions apparemment objectives et fiables. Les algorithmes peuvent modéliser des prédictions sur la base de masses de données que les humains ne sont pas capables de collecter intégralement ((Agrawal et al., 2019); (van den Broek et al., 2021)). Les aspects qui restent discutables lorsque l'on envisage d'introduire l'IA dans le processus de recrutement ne sont pas seulement de savoir si ces hypothèses sont réellement fondées sur des preuves, mais aussi comment les personnes concernées perçoivent les processus de décision algorithmiques (Gonzalez et al., 2022).

Dans ce bref dossier, nous présentons les preuves scientifiques de l'interaction entre l'IA et les résultats des attitudes, comme la qualité et la diversité des attitudes. Dans la **sous-section 1** nous démontrons la capacité de l'IA à réduire la discrimination et à permettre des résultats d'embauche favorables, tout en mettant en lumière, dans la **sous-section**

2, un problème urgent dans la discussion sur l'IA dans le recrutement : Le biais de l'IA. Dans la **section 2**, nous soulignons l'importance de la perception des processus de décision algorithmiques par les travailleurs et démontrons les conséquences potentielles que cela pourrait avoir sur les résultats du marché du travail au niveau individuel et de l'entreprise. Enfin, dans la **section 3**, nous examinons les implications de nos conclusions pour les politiques de cadre réglementaire afin de garantir l'équité dans tous les processus de recrutement algorithmique.

Premièrement- l'IA peut-elle réduire la discrimination ?

La recherche a montré à maintes reprises que les processus de recrutement sont souvent associés à une discrimination fondée sur des caractéristiques personnelles qui n'ont en réalité rien à voir avec leur productivité ; ces caractéristiques potentielles sont par exemple la race ou la nationalité, qui sont déduites du nom ou de l'apparence des candidats ((Becker, 2010); (Bendick et al., 1994); (Bertrand & Mullainathan, 2004); (Kaas & Manger, 2012)). Plus récemment, des chercheurs ont proposé d'intégrer l'IA dans le processus de recrutement afin de surmonter la discrimination basée sur les préjugés inconscients des humains (van den Broek et al., 2021). Étant donné que les humains ont une capacité relativement limitée à traiter des informations provenant de différentes sources, l'IA peut fournir des connaissances plus objectives : les algorithmes de recrutement intelligents exploitent de vastes ensembles de données provenant de différentes sources de données, telles que les CV, les entretiens et les médias sociaux, de sorte qu'ils peuvent prédire quels candidats sont les mieux placés pour occuper le poste vacant (Agrawal et al., 2019), ce qui dépasse souvent les conclusions des experts techniques (van den Broek et al., 2021).

La pertinence de l'IA pour le processus de recrutement se reflète dans l'augmentation récente du nombre de fournisseurs de solutions d'IA pour les procédures de recrutement (Raghavan et al., 2020).

Les systèmes basés sur l'IA peuvent avoir différentes fonctions à différentes étapes du processus de recrutement, par exemple la suppression de formulations biaisées dans les offres d'emploi, le filtrage de CV, l'analyse d'entretiens vidéo ou l'évaluation de l'adéquation d'une personne à un poste ((Gonzalez et al., 2019); (Raghavan et al., 2020)). Les nombreuses possibilités d'application de l'IA aux premières étapes du processus de recrutement soulignent le fait que les algorithmes ont le potentiel d'influencer de manière substantielle les types de candidats qui se retrouvent dans le pipeline global (Rieke & Bogen, 2022).

Actuellement, le principal argument organisationnel en faveur de la mise en œuvre d'un processus de recrutement algorithmique est l'augmentation de l'efficacité et l'optimisation des métriques de recrutement, telles que : Temps d'embauche, coûts d'embauche, qualité d'embauche et rétention des employés ((Rieke & Bogen, 2022) ; (Gonzalez et al., 2019)). En outre, la capacité de l'IA à réduire les biais interpersonnels (van den Broek et al., 2021) contribue à la réalisation d'objectifs organisationnels en lien avec la promotion de la diversité et de l'inclusion (Rieke & Bogen, 2022). Des revues spécialisées telles que Forbes (Windley, 2021), Harvard Business Review (Chamorro-Premuzic & Akhtar, 2019) et Business Insider (Garfield, 2017) ont confirmé les avantages de l'utilisation de l'IA dans les recrutements, tout en préconisant une approche prudente lorsqu'on s'appuie trop sur les algorithmes, ce qui implique qu'il y a deux côtés à la médaille. Par conséquent, la recherche universitaire a commencé à examiner si l'IA réduit réellement ou élimine la discrimination à l'embauche. Dans les sections suivantes, les possibilités et les dangers identifiés de l'application de l'IA au recrutement sont développés plus en détail.

1. Qui affiche les meilleurs résultats à l'embauche IA ou Humains ?

Au sein du processus de recrutement, le screening des CV est un domaine qui a été identifié comme source de discrimination (Bertrand & Mullainathan, 2004). Deux études ((Cowgill & Stevenson, 2020); (Li et al., 2020)) ont développé des algorithmes d'IA pour le screening de CV et les ont comparés à la performance humaine. (Li et al., 2020) considèrent le recrutement comme un "problème de bandit", dans lequel les décideurs doivent mettre en balance l'exploration et l'exploitation. Dans ce contexte, les entreprises qui recrutent tentent de trouver un équilibre entre la sélection de candidats issus de groupes ayant déjà fait leurs preuves (Exploitation) et le sourcing de groupes sous-représentés (Exploration). Les algorithmes de recrutement actuels, basés sur des données de formation, prédisent des résultats positifs pour ces groupes à l'historique de réussite prouvé, mais ne tiennent pas compte des candidats issus de groupes qui pourraient avoir des antécédents non conventionnels, mais qui pourraient néanmoins exceller dans leur travail. L'intégration de cette exploration dans les algorithmes de recrutement peut constituer un pas en avant vers une approche plus inclusive (Li et al., 2020), ce qui correspond à l'idée de construire un vivier de candidats plus diversifié (Rieke & Bogen, 2022).

En conséquence, (Li et al., 2020) comparent les recruteurs humains à trois algorithmes : deux qui ont été formés par apprentissage supervisé et un qui comporte un degré d'incertitude plus élevé, c'est-à-dire un "bonus

d'exploration". Les résultats montrent que l'algorithme d'exploration donne des pourcentages beaucoup plus élevés de groupes sous-représentés dans le processus de sélection. Non seulement cet algorithme obtient des résultats significativement meilleurs que l'approche humaine, mais il dépasse également les deux algorithmes surveillés. Il est important de noter que les deux algorithmes d'apprentissage supervisé développés dans cette étude ont obtenu des résultats inférieurs à ceux de l'approche humaine lorsqu'il s'agit de sélectionner un groupe diversifié de candidats, ce qui souligne le potentiel des algorithmes dans l'établissement de préjugés supplémentaires, comme nous le verrons en détail ci-dessous. Les auteurs soulignent l'importance de la conception de l'algorithme pour sa performance : tous les algorithmes n'ont pas la même performance et c'est un facteur important à prendre en compte lorsqu'il existe une multitude de choix.

Dans un contexte de recrutement expérimental, (Cowgill, 2019) a constaté qu'un algorithme sélectionnait les parcours des candidats qui avaient obtenu de meilleurs résultats et qui étaient plus susceptibles de réussir les entretiens et de se voir proposer des emplois. L'algorithme a également augmenté le nombre de recrutements non conventionnels, y compris ceux des femmes, des minorités, des candidats non recommandés, des candidats issus d'universités non prestigieuses et de ceux qui n'avaient pas d'expérience professionnelle. En surface, il semble que l'algorithme ait été entraîné à l'aide de techniques similaires (c'est-à-dire LASSO et d'autres techniques d'apprentissage automatique standard) aux algorithmes supervisés de (Li et al., 2020), qui ont constaté que ces algorithmes pouvaient conduire à des préjugés supplémentaires.

Ces résultats contradictoires soulignent la nécessité de mener des recherches futures pour parvenir à un consensus sur la performance des algorithmes. L'intégration de différentes conceptions d'algorithmes, comme l'algorithme d'exploration de (Li et al., 2020), pourrait conduire à une analyse plus approfondie et à une meilleure compréhension de la manière dont la discrimination pourrait être réduite.

Comme le souligne (Cowgill & Stevenson, 2020), la composition de l'ensemble des données d'apprentissage pourrait également conduire à des résultats différents. Dans ses modèles, l'auteur montre comment un bruit statistique plus ou moins important peut influencer les résultats. Il est donc essentiel de reproduire les résultats de ces études à l'aide de données d'entraînement de composition différente.

Outre le screening de CV, l'IA pourrait également être utilisée pour l'analyse des tests d'aptitude professionnelle. Il s'agit de questionnaires

envoyés avant la sélection, afin que les recruteurs puissent en tenir compte dans leurs décisions.

(Hoffman et al., 2017) testent les résultats d'une telle IA et constatent des effets positifs sur la durée d'emploi des recrutements réussis par rapport à l'ancienne approche basée sur l'humain. Ils examinent également si la qualité des recrutements est plus élevée lorsque les recruteurs suivent les recommandations de l'algorithme ou lorsqu'ils rejettent activement ses recommandations. Ils montrent que les recruteurs qui rejettent les recommandations de l'algorithme prennent de moins bonnes décisions, que les travailleurs qu'ils ont sélectionnés ont des durées d'emploi plus courtes et qu'ils ne sont pas plus productifs. Leur étude contribue à démontrer que l'IA peut représenter une amélioration par rapport à un(e) recruteur(se) humain(e) : non seulement pour identifier des candidats plus performants, mais aussi pour constituer un pool de candidats plus diversifié. L'algorithme utilisé était propriétaire (Hoffman et al., 2017) et il y a peu de discussions sur sa conception ou sur la méthode d'apprentissage utilisée par l'algorithme. Pour une évaluation finale et objective de la technologie de l'IA, telle que discutée ci-dessus, il serait nécessaire de divulguer la technologie du modèle et les données utilisées pour l'apprentissage du modèle.

2. Toutes les IA ne sont pas impartiales !

Dans la section ci-dessus, nous avons décrit quelques cas où l'IA pourrait améliorer un décideur humain et réduire les résultats discriminatoires dans le processus de recrutement. Cependant, nous avons également montré que différents algorithmes peuvent donner des résultats différents. Comme le soulignent (Li et al., 2020) dans le cas des algorithmes d'apprentissage supervisé, il existe un risque qu'un algorithme introduise des biais supplémentaires.

Dans ce contexte, il existe de nombreuses histoires dans les médias qui soulignent le potentiel des algorithmes d'IA à renforcer les préjugés plutôt que de les réduire, ainsi que des études qui soulignent le biais de l'IA dans d'autres domaines, par exemple dans les outils de reconnaissance faciale (Leslie, 2020). C'est pourquoi cette section est consacrée à la mise en évidence des sources sous-jacentes des préjugés en matière d'IA, ainsi qu'à des exemples montrant comment un préjugé en matière d'IA se produit.

Les algorithmes d'IA actuels dépendent fortement de données d'apprentissage produites par l'homme et apprennent principalement en appliquant une combinaison de plusieurs approches statistiques aux données d'apprentissage. Ce processus d'apprentissage est généralement

appelé 'apprentissage automatique' ou 'apprentissage profond' (OCDE, 2019). Une récente étude bibliographique (Cecere et al., 2021) a identifié les causes fondamentales du biais de l'IA comme étant principalement dues à l'utilisation **(I)** de données et de modèles d'apprentissage non représentatifs, **(II)** de résultats mal étiquetés et **(III)** de programmeurs partiels ((Cecere et al., 2021) ; (Cowgill, 2019); (Cowgill & Stevenson, 2020); (Cowgill & Tucker, 2019)). Les algorithmes ne sont pas créés de manière intelligente ; ils doivent être entraînés sur des données avant d'être appliqués au monde réel. L'algorithme prend ces données et part du principe qu'elles sont représentatives du monde réel, et suppose donc que ses décisions s'appliquent à la fois aux données d'apprentissage et au monde réel. Si les données d'apprentissage sont biaisées ou non représentatives, l'algorithme pourrait naïvement reproduire les processus de sélection qui ont conduit à des données non représentatives, et ainsi établir davantage le biais ((Cowgill & Tucker, 2019); (Kleinberg et al., 2018)). De plus, comme les algorithmes d'IA ont tendance à être entraînés et à apprendre plus tard des comportements et des caractéristiques antérieurs des personnes, il est probable que la discrimination s'installe dans un cercle vicieux (Cecere et al., 2021). Dans le contexte de l'embauche, si les données de formation de l'IA sont basées sur des données d'emploi historiques, par exemple si elles contiennent un biais implicite qui favorise les Blancs par rapport aux employés afro-américains et latino-américains, l'algorithme peut reconnaître de tels modèles et ne pas tenir compte de certains groupes de candidats.

La prise de décision pourrait alors se faire en faveur du groupe surreprésenté qui, historiquement, est plus susceptible d'être convoqué à un entretien d'embauche ((Köchling & Wehner, 2020) ; (Tambe et al., 2019)). Le géant mondial de la technologie Amazon est un exemple de premier ordre dans ce scénario : En 2018, l'entreprise a découvert que son système de recrutement basé sur l'IA était basé sur des données historiques de performance qui étaient fortement dominées par les hommes et qui contenaient des scores de performance plus élevés pour les hommes blancs. En conséquence, l'algorithme formé sur cette sélection d'informations a donné des scores plus élevés pour les candidats masculins blancs, tandis que les femmes et les candidats dont les attributs étaient associés aux femmes ont été éliminés (Tambe et al., 2019).

Les biais algorithmiques peuvent également être dus à un mauvais étiquetage de certains résultats dans les données d'entraînement. Cela pourrait être dû aux goûts subjectifs des décideurs humains ou à un

mauvais raisonnement statistique (Cowgill & Tucker, 2019). Dans une étude récente, (Obermeyer et al., 2019) soulignent l'importance du choix d'étiquettes pertinentes pour l'apprentissage de l'algorithme en documentant les différences algorithmiques importantes qui se produisent dans le secteur de la santé.

Ils montrent que pour un même risque prédit par l'algorithme pour indiquer les besoins en matière de santé, les patients noirs en moins bonne santé ont reçu des évaluations de risque similaires à celles des patients blancs en meilleure santé, alors qu'ils présentent en fait des cas graves de diabète, d'hypertension, de cholestérol et d'anémie. La raison d'un tel préjugé racial est que la prédiction de l'algorithme de santé se base sur les coûts de santé plutôt que sur les maladies, et comme moins d'argent est dépensé pour les patients noirs qui ont le même niveau de besoin, l'algorithme conclut à tort que les patients noirs sont en aussi bonne santé que les patients blancs (Obermeyer et al., 2019).

Cet exemple illustre une conséquence potentielle qui pourrait survenir à la suite de résultats mal étiquetés ; les données de formation étiquettent les patients à faible coût comme étant en bonne santé, alors que dans certains cas, ces faibles coûts sont dus à un accès inégal aux soins de santé et au statut socio-économique ((Cecere et al., 2021) ; (Cowgill & Tucker, 2019); (Obermeyer et al., 2019)).

Tous les développeurs qui supervisent la programmation, l'étiquetage et la formation des algorithmes pourraient également avoir une influence sur l'apparition de préjugés algorithmiques, en particulier s'ils manquent de connaissances techniques sur l'éthique de l'IA ou s'ils ont eux-mêmes des préjugés significatifs. L'existence de tels préjugés peut conduire à des modèles d'entraînement non représentatifs, à des résultats mal étiquetés et/ou à un manque d'attention portée aux exemples d'entraînement disponibles et aux mesures de précision pendant le processus de développement (Cowgill & Tucker, 2019).

Dans une tentative d'explorer l'origine et la diffusion des principales sources de préjugés en matière d'IA, (Cowgill et al., 2020) ont cherché à savoir si les préjugés en matière d'IA proviennent de données de formation biaisées ou de programmeurs biaisés. Les résultats de leur étude sur le terrain montrent que les données d'entraînement biaisées sont le principal facteur de préjugés et que tous les développeurs sont hautement représentatifs et peuvent montrer des préjugés qui se répercutent sur les algorithmes qu'ils développent ((Cecere et al., 2021) ; (Cowgill et al., 2020)).

En bref, bien que chaque source de préjugés en matière d'IA soit spécifique à un cas, le mécanisme sous-jacent résulte du fait que les

préjugés humains sont transférés aux approches basées sur l'IA. Les algorithmes d'IA produisent des résultats discriminatoires et déformés lorsque les données et les modèles qu'ils ont appris à utiliser sont inexacts ou déformés ((Köchling & Wehner, 2020); (Mujtaba & Mahapatra, 2019)).

Si les données sont constituées d'un certain degré de préjugé culturel ou personnel, un algorithme d'IA peut adopter et reproduire le préjugé, ce qui donne des résultats injustes, comme par exemple une sous-représentation ou une surreprésentation de certains groupes de population. Le résultat discriminatoire qui en résulte soulève donc des questions sur la fiabilité des systèmes d'IA ((Lee, 2018) ; (OCDE, 2019)).

A la lumière des discussions ci-dessus, la compréhension des données d'apprentissage fournies à l'IA est très importante pour mesurer sa capacité à réduire ou à aggraver la discrimination. Lorsqu'une IA est entraînée avec des données discriminatoires, les preuves ci-dessus indiquent que cela renforce les préjugés. Cela suggère que les données existantes, en plus des conceptions algorithmiques choisies, sont difficiles à évaluer lors de la sélection d'un outil d'IA, car les différents niveaux de biais au sein des données d'entraînement nécessitent différentes approches algorithmiques.

Deuxièmement- Perception des décisions algorithmiques : fiable ou ambiguë ?

Alors que les preuves décrites ci-dessus indiquent que l'IA est un outil permettant de réduire la discrimination, l'application de l'IA au processus de recrutement suscite d'autres préoccupations au-delà de la discrimination. Bien que l'IA puisse réduire la discrimination et que cette réduction doive être testée avec précision, elle peut encore être inadaptée comme outil de recrutement.

L'une des principales préoccupations est de savoir si les personnes exposées à des décisions algorithmiques les considèrent comme plus équitables que les décisions humaines, qu'elles soient effectivement plus équitables ou non (Lee, 2018).

Cette perception humaine est importante pour deux raisons : premièrement, si les candidats perçoivent un algorithme, et donc le processus de recrutement, comme injuste, ils pourraient ne pas postuler à un emploi ou l'accepter. Une telle auto-sélection de la main-d'œuvre peut avoir un impact négatif sur l'organisation de recrutement, car les candidats peuvent avoir des gains de productivité plus importants qu'une alternative, ce qui entraîne une baisse globale des performances et de la

croissance. Les candidats alternatifs peuvent être si inefficaces qu'ils doivent être remplacés, ce qui augmente encore les coûts en raison d'une rotation plus élevée des employés et d'une attribution inefficace des compétences. Deuxièmement, si les candidats décident d'entrer dans une entreprise alors qu'ils ont des doutes sur le processus de recrutement, cela peut entraîner une perte de confiance entre l'employeur et l'employé. Ce manque de confiance peut avoir un impact sur la durée de la relation d'emploi et conduire à une rotation du personnel, ce qui est extrêmement coûteux pour les entreprises, estimé à plus de 16 milliards de dollars pour les entreprises américaines (Onovakpuri, 2017). Par conséquent, s'assurer que les candidats sont satisfaits du processus de recrutement et qu'ils ont confiance en l'équité de la décision d'embauche peut être considéré comme un avantage pour les entreprises.

Il est donc légitime de se demander si les algorithmes de recrutement et les mécanismes potentiels qui sous-tendent ces perceptions sont perçus comme positifs ou négatifs par les individus. (Bao et al., 2022) trouvent que les décisions d'embauche prises par l'IA semblent moins justes que des décisions humaines identiques. De même, (Acikgoz et al., 2020) montrent que les entretiens menés par une IA semblent moins équitables que ceux menés par des humains. De plus, (Gonzalez et al., 2019) constatent que les travailleurs font moins confiance aux entreprises qui utilisent des algorithmes pour prendre des décisions d'embauche.

(Gonzalez et al., 2022) esquissent des mécanismes potentiels pour le manque de confiance entre les décisions prises par des personnes et celles prises par des algorithmes. Les candidats sont souvent peu informés de la manière dont une IA est utilisée dans un processus décisionnel donné ou ne sont peut-être pas en mesure de comprendre ou de traiter ces informations, si elles sont disponibles.

Ils utiliseront alors des critères d'évaluation mentale simples pour juger de l'équité relative, en s'appuyant potentiellement sur des histoires médiatiques qui peuvent favoriser des résultats négatifs.

Les candidats peuvent également percevoir les décisions de l'IA comme moins équitables parce qu'ils ne sont pas familiers avec l'IA elle-même et le processus décisionnel dans lequel elle s'inscrit, alors qu'ils peuvent mieux comprendre un processus décisionnel basé sur l'homme en disposant de plus d'informations pour évaluer l'équité.

Ces mécanismes potentiels peuvent être atténués par une plus grande transparence dans l'utilisation de l'IA. Alors qu'une entreprise peut facilement expliquer où une IA est mise en œuvre dans un processus et si elle a le pouvoir de décision final, les candidats ont plus de mal à

comprendre les raisons qui poussent une IA à prendre ses décisions. Les IA modernes sont entraînées à l'aide de réseaux neuronaux et de techniques d'apprentissage en profondeur ; il s'agit de techniques opaques, car on ne sait pas exactement comment une IA parvient à prendre ses décisions. Les chercheurs s'efforcent de rendre les IA explicables (voir (Arrieta et al., 2020), pour une discussion) afin de mieux comprendre leurs processus de décision respectifs. Il est essentiel de continuer à développer cette approche si l'on veut que l'IA devienne "digne de confiance".

Troisièmement - implications politiques

Quels efforts réglementaires sont actuellement déployés pour garantir que l'IA soit centrée sur l'humain, digne de confiance et sûre ? Dans son projet de loi sur l'IA (COMMISSION, 2021), l'Union européenne considère l'utilisation de l'IA à des fins de recrutement comme un risque élevé.

Dans le cadre de cette classification, une technologie d'IA utilisée à des fins de recrutement n'est pas interdite, mais plutôt réglementée et doit remplir un certain nombre de conditions, telles que la mise en place d'un système de gestion des risques, des conditions de transparence et une surveillance humaine. Le système de gestion des risques exige l'identification et l'analyse des risques connus et prévisibles, l'estimation et l'évaluation de ces risques et des risques potentiels identifiés lors de la surveillance post-commercialisation. La possibilité que l'IA reproduise naïvement des préjugés humains est l'un des principaux risques évoqués dans ce dossier de synthèse et doit être pris en compte dans le cadre de cette législation. Une autre préoccupation soulevée dans ce dossier est l'exigence de transparence, qui exige que les utilisateurs soient en mesure d'interpréter la production du système et de comprendre son objectif. Cela permet de répondre à certaines des préoccupations concernant la perception de l'IA par les personnes et sa fiabilité relative. Si ces exigences de transparence sont remplies, elles peuvent contribuer à instaurer la confiance, mais cela n'est pas garanti et des travaux futurs sont nécessaires pour déterminer si ces exigences sont remplies et ce qui peut être fait d'autre pour instaurer la confiance.

Il y a des raisons de penser que la loi sur l'IA joue en faveur de certains des défis décrits ci-dessus. L'une des principales questions soulevées par cette législation est de savoir si les coûts ne vont pas étouffer l'innovation et l'application des algorithmes d'IA en raison d'une charge administrative trop lourde (Czarnocki, 2021). Inversement, certains craignent que des mesures telles que la loi sur l'IA, qui exigent

certaines formes d'examen, ne soient pas assez ambitieuses pour éradiquer les technologies nuisibles. (Sloane, 2022) souligne ce point en discutant d'une loi new-yorkaise qui exige l'examen des algorithmes d'embauche, arguant que ces examens n'évaluent pas suffisamment l'adéquation d'une technologie. Ce sont des débats qui se poursuivent ailleurs, mais qui doivent être pris en compte lors de l'évaluation de l'impact de la politique sur l'IA dans le recrutement.

Étant donné que la loi sur l'IA n'a pas encore été transposée dans la législation, il est trop tôt pour évaluer son impact de manière adéquate, mais il s'agit d'un développement intéressant et important qui aura un impact considérable sur l'évolution future de l'IA dans l'Union européenne.

Conclusion

Les discussions ci-dessus soulignent le fait que l'IA a un grand potentiel en tant qu'outil de réduction de la discrimination à l'embauche. Elles constituent des preuves significatives du fait que l'IA peut accroître la diversité des candidats retenus tout en améliorant leurs performances globales. Toutefois, des cas d'application bien documentés, tels que l'algorithme de recrutement discriminatoire d'Amazon mentionné ci-dessus, soulignent la nécessité d'une approche prudente. Comme nous l'avons décrit, le mode d'apprentissage actuel de l'IA comporte une forte composante d'interaction humaine.

Qu'il s'agisse de la gestion des données d'apprentissage appliquées par l'IA, de l'étiquetage de ces données ou de l'étiquetage des résultats, les biais peuvent être transmis de l'homme à l'IA. En conséquence, l'IA peut établir plus fortement des préjugés, ce qui perpétue ou même aggrave la discrimination. Les résultats des algorithmes supervisés de (Li et al., 2020) indiquent cette possibilité, l'algorithme d'exploration plus performant étant spécifiquement orienté vers la sélection d'un autre groupe de candidats. Cela souligne l'importance des décisions de conception pour la performance des algorithmes : En cas de mauvaise conception, une IA pourrait reproduire naïvement des préjugés inconsciemment entretenus par les humains et contrecarrer ainsi son objectif de procédures de recrutement objectives.

Comme le montre l'étude de la perception individuelle de l'IA en matière d'attitude, les personnes ont peu de confiance dans les décisions d'une IA, même si ces décisions sont aussi justes ou plus justes que les décisions humaines. Ce résultat souligne la nécessité de la transparence partout où l'IA est utilisée. Les gens ont tendance à se méfier de l'inconnu, ce qui nous amène à la conclusion que la transparence et

l'ouverture sont essentielles pour obtenir une plus grande acceptation et participation aux procédures de recrutement algorithmiques.

Alors que l'explicabilité de l'IA n'en est qu'à ses débuts, la transparence des utilisateurs et des fournisseurs d'outils de recrutement de l'IA en ce qui concerne les décisions de conception et les données de formation contribuerait également à instaurer la confiance entre les humains et les machines.

Les technologies n'évoluent pas de manière déterministe ; l'IA n'évoluera pas instinctivement vers plus ou moins de discrimination. Dans chaque cadre technologique, de nombreuses décisions doivent être prises concernant la conception et l'utilisation d'un artefact. Les technologies d'IA deviennent donc plus transparentes et plus explicables grâce à des décisions soigneusement prises, qui sont explicitement documentées en détail et mises à la disposition des personnes concernées.

Ce bref dossier a tenté de mettre en lumière le rôle de l'IA dans le recrutement et certaines des décisions cruciales qui peuvent être prises avec cette technologie. Lors de la conception ou de l'utilisation d'une IA, il est essentiel de prendre en compte les données d'apprentissage, la méthode d'apprentissage appliquée à l'IA et les critères de sélection utilisés.

RÉFÉRENCES

1. Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment*, 28(4), 399-416. <https://doi.org/10.1111/ijsa.12306>
2. Adalet McGowan, M. D. A. (2015). *Labour Market Mismatch and Labour Productivity* (1815-1973). (OECD Economics Department Working Papers, Issue. <http://dx.doi.org/10.1787/5js1pzx1r2kb-en>
3. Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction. *Journal of Economic Perspectives*, 33(2), 31-50. <https://doi.org/10.1257/jep.33.2.31>
4. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., & Benjamins, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities, and challenges toward responsible AI. *Information fusion*, 58, 82-115.
5. Banks, R. R., & Ford, R. T. (2008). (How) does unconscious bias matter: Law, politics, and racial inequality. *Emory LJ*, 58, 1053.
 Bao, L., Krause, N. M., Calice, M. N., Scheufele, D. A., Wirz, C. D., Brossard, D., Newman, T. P., & Xenos, M. A. (2022). Whose AI? How

- different publics think about AI and its social impacts. *Computers in Human Behavior*, 130, 107182. <https://doi.org/https://doi.org/10.1016/j.chb.2022.107182>
6. Becker, G. S. (2010). *The economics of discrimination*. University of Chicago Press.
 7. Bendick, M., Jackson, C. W., & Reinoso, V. A. (1994). Measuring Employment Discrimination through Controlled Experiments. *The Review of Black Political Economy*, 23(1), 25-48. <https://doi.org/10.1007/bf02895739>
 8. Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, 94(4), 991-1013. <https://doi.org/10.1257/0002828042002561>
 9. Cecere, G., Corrocher, N., & Jean, C. (2021). Fair or Unbiased Algorithmic Decision-Making? A Review of the Literature on Digital Economics. *A Review of the Literature on Digital Economics* (October 15, 2021).
 10. Chamorro-Premuzic, T., & Akhtar, R. (2019, 2019/05/17/). *Should Companies Use AI to Assess Job Candidates?* <https://hbr.org/2019/05/should-companies-use-ai-to-assess-job-candidates>
 11. COMMISSION, E. (2021). *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS*. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>
 12. Cowgill, B. (2019). *Bias and Productivity in Humans and Machines*. <http://dx.doi.org/10.17848/wp19-309>
 13. Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N., & Chaintreau, A. (2020, 2020). Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics.
 14. Cowgill, B., & Stevenson, M. T. (2020). Algorithmic Social Engineering. *AEA Papers and Proceedings*, 110, 96-100. <https://doi.org/10.1257/pandp.20201037>
 15. Cowgill, B., & Tucker, C. E. (2019). Economics, Fairness and Algorithmic Bias. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3361280>
 16. Czarnocki, J. (2021, 2021/09/03/). *Good intentions, unintended consequences? How the Proposal for the AI Act might kick innovation out of the EU - CITIP blog*. <https://www.law.kuleuven.be/citip/blog/good-intentions-unintended-consequences>
 17. Garfield, L. (2017). This AI startup is using games to help rid the hiring process of bias. *Business Insider*. <https://www.businessinsider.com/hiring-diversity-brain-games-artificial-intelligence-automation-2017-9>

18. Gonzalez, M., Capman, J., Oswald, F., Theys, E., & Tomczak, D. (2019). "Where's the I-O?" Artificial Intelligence and Machine Learning in Talent Management Systems. *Personnel Assessment and Decisions*, 5(3). <https://doi.org/10.25035/pad.2019.03.005>
19. Gonzalez, M. F., Liu, W., Shirase, L., Tomczak, D. L., Lobbe, C. E., Justenhoven, R., & Martin, N. R. (2022). Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes. *Computers in Human Behavior*, 130, 107179. <https://doi.org/https://doi.org/10.1016/j.chb.2022.107179>
20. Hoffman, M., Kahn, L. B., & Li, D. (2017). Discretion in Hiring*. *The Quarterly Journal of Economics*, 133(2), 765-800. <https://doi.org/10.1093/qje/qjx042>
21. Kaas, L., & Manger, C. (2012). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review*, 13(1), 1-20. <https://doi.org/doi:10.1111/j.1468-0475.2011.00538.x>
22. Kleinberg, J., Ludwig, J., Mullainathan, S., & Rambachan, A. (2018). Algorithmic Fairness. *AEA Papers and Proceedings*, 108, 22-27. <https://doi.org/10.1257/pandp.20181018>
23. Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13(3), 795-848. <https://doi.org/10.1007/s40685-020-00134-w>
24. Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 205395171875668. <https://doi.org/10.1177/2053951718756684>
25. Leslie, D. (2020). Understanding bias in facial recognition technologies. *Zenodo*. <https://doi.org/10.5281/zenodo.4050457>
26. Li, D., Raymond, L., & Bergman, P. (2020). *Hiring as Exploration*. <http://dx.doi.org/10.3386/w27736>
27. Mujtaba, D. F., & Mahapatra, N. R. (2019, 15-16 Nov. 2019). Ethical Considerations in AI-Based Recruitment. 2019 IEEE International Symposium on Technology and Society (ISTAS),
28. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453. <https://doi.org/10.1126/science.aax2342>
29. OCDE. (2019). *Artificial Intelligence in Society*. <https://doi.org/doi:https://doi.org/10.1787/eedfee77-en>
30. Onovakpuri, A. S. F. K. U. (2017). *A first-of-its-kind analysis of why people voluntarily left jobs in tech* (TECH LEAVERS STUDY, Issue. <https://www.kaporcenter.org/wp-content/uploads/2017/08/TechLeavers2017.pdf>
31. Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020, 2020). Mitigating bias in algorithmic hiring.

32. Rieke, A., & Bogen, M. (2022, 2022/09/22/). *Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias*. <https://www.upturn.org/work/help-wanted>
33. Sloane, M. (2022). To make AI fair, here's what we must learn to do. *Nature*, 605(7908), 9-9. <https://doi.org/10.1038/d41586-022-01202-3>
34. Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial Intelligence in Human Resources Management: Challenges and a Path Forward. *California Management Review*, 61(4), 15-42. <https://doi.org/10.1177/0008125619867910>
35. van den Broek, E., Sergeeva, A., & Huysman Vrije, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. *MIS Quarterly*, 45(3), 1557-1580. <https://doi.org/10.25300/misq/2021/16559>
36. Windley, D. (2021). Is AI The Answer To Recruiting Effectiveness? *Forbes*. <https://www.forbes.com/sites/forbeshumanresourcescouncil/2021/06/16/is-ai-the-answer-to-recruiting-effectiveness/?sh=2789b29f2d7c>