



HAL
open science

CROIRE OU NE PAS CROIRE LES ALGORITHMES... ? PERCEPTIONS ET COMPORTEMENT DES RECRUTEURS FACE AUX ALGORITHMES LORS DE LA PRE-SELECTION DE CV

Alain Lacroux, Christelle Martin Lacroux

► **To cite this version:**

Alain Lacroux, Christelle Martin Lacroux. CROIRE OU NE PAS CROIRE LES ALGORITHMES... ? PERCEPTIONS ET COMPORTEMENT DES RECRUTEURS FACE AUX ALGORITHMES LORS DE LA PRE-SELECTION DE CV. 33ème congrès de l'AGRH, AGRH (association francophone de gestion des ressources humaines), Oct 2022, Brest, France. hal-04095500

HAL Id: hal-04095500

<https://paris1.hal.science/hal-04095500v1>

Submitted on 11 May 2023

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

CROIRE OU NE PAS CROIRE LES ALGORITHMES... ? PERCEPTIONS ET COMPORTEMENT DES RECRUTEURS FACE AUX ALGORITHMES LORS DE LA PRESELECTION DE CV

Alain Lacroux¹ et Christelle Martin-Lacroux²

¹Univ. Paris 1 Panthéon-Sorbonne, PRISM, Paris, France

²Univ. Grenoble Alpes, Grenoble INP, CERAG, Grenoble, France

1. Introduction

Les systèmes algorithmiques intégrant l'Intelligence Artificielle (IA par la suite) font désormais partie de la panoplie des outils à la disposition des responsables RH : une enquête menée en 2018 auprès de 9000 recruteurs a par exemple révélé que 64 % d'entre eux " utilisent l'IA *a minima* " parfois " dans le cadre de leur activité de recrutement ; 79 % sont susceptibles de le faire dans les deux prochaines années, et 76 % pensent que l'IA aura un impact significatif sur le recrutement (LinkedIn, 2018).

L'IA est généralement définie comme l'utilisation de la technologie numérique pour créer des systèmes capables d'effectuer de manière autonome des tâches dont on pense généralement qu'elles nécessitent une intelligence humaine (Office for AI, 2019). Comme le soulignent Burton et ses collègues (2020), la décision assistée par algorithme concerne un vaste domaine, qui regroupe des thèmes tels que la prise de décision augmentée, les systèmes experts, et les aides au diagnostic. Dans la présente recherche, nous utilisons l'expression système d'aide à la décision algorithmique (ADSS par la suite pour *algorithmic decision-support system*) pour caractériser ces systèmes, dans lesquels "*les résultats peuvent être utilisés comme source d'information supplémentaire ou alternative pour les décideurs*" (Langer et al., 2020, p. 753). On peut trouver de nombreux exemples d'utilisation des ADSS au cours du processus de recrutement, qu'il s'agisse de l'étape de sourcing de candidats ou de présélection. Les solutions de sourcing intégrant des algorithmes permettent d'automatiser cette étape en identifiant les profils les plus en phase avec les postes proposés. On peut citer la solution « *EasySourcing* » proposée par la start-up Pipler¹ qui source, identifie et contacte les profils correspondant à un poste en mobilisant 50 sources différentes. L'IA peut être utilisée pendant les entretiens pour évaluer la richesse du vocabulaire du candidat, son débit et son ton de voix, ses expressions faciales afin d'évaluer son adéquation au poste. L'entreprise Cryfe² propose une solution appelée « *Authentic Behavior* » qui analyse la cohérence du langage corporel d'un candidat en situation d'entretien pour prédire l'authenticité de ce dernier. Les recruteurs peuvent également utiliser des *chatbots* pour interagir en temps réel avec les candidats en leur posant des questions de manière à évaluer leur adéquation au profil de poste. Toutes ces fonctionnalités sont regroupées par leurs développeurs sous le terme générique de recrutement prédictif.

Les chercheurs s'intéressent aujourd'hui aux raisons pour lesquelles les acteurs choisissent de recourir à ces outils d'aide à la décision (Burton et al., 2020). Plus spécifiquement, un champ de recherche traite de la question spécifique de la réticence des individus à utiliser les systèmes automatisés (Dietvorst et al., 2015) ou *a contrario* de leurs appréciation positive de cette

¹ <https://www.pipler.io/easysourcing/>

² https://try.cryfe.swiss/recrutement-intelligence-artificielle/?utm_source=google&utm_medium=cp&utm_campaign=search-fr&utm_term=intelligence

technologie (Logg et al., 2019). Dans le domaine du recrutement, on sait aujourd'hui encore peu de choses sur la façon dont les ADSS influencent les recruteurs. Le *gap* entre scientifiques et praticiens apparaît important lorsqu'il s'agit d'étudier comment et pourquoi les recruteurs perçoivent différents outils, et en particulier ceux intégrant l'IA (Chamorro-Premuzic et al., 2016; Gonzalez et al., 2019; Kuncel et al., 2013; Langer et al., 2019).

Les recherches disponibles montrent que les individus peuvent réagir de différentes manières aux systèmes automatisés, et que la confiance dans la qualité et la fiabilité des conseils apparaît comme un déterminant majeur de l'adoption effective des technologies d'automatisation (J. D. Lee & See, 2004). La confiance est définie comme "la volonté d'une partie (« *trustor* ») d'être vulnérable aux actions d'une autre partie (« *trustee* ») sur la base de l'attente que l'autre partie accomplira une action particulière importante pour le créancier, indépendamment de la capacité de surveiller ou de contrôler cette autre partie" (Mayer et al., 1995, p.712).

La confiance dans les systèmes automatisés joue un rôle central dans le type de réactions des utilisateurs (Parasuraman & Riley, 1997). Ces réactions peuvent s'appréhender selon la typologie suivante : l'utilisation (*use*), la mauvaise utilisation (*misuse*), la non-utilisation (*disuse*) et l'abus (*abuse*). L'utilisation (*use*) fait référence aux utilisateurs qui comptent sur l'automatisation pour effectuer des tâches qu'ils pourraient réaliser manuellement : dans de telles situations, un degré raisonnable de confiance dans un système d'aide à la décision automatisé est nécessaire pour s'engager dans son utilisation. La mauvaise utilisation (*misuse*) fait référence à des situations dans lesquelles les utilisateurs font trop confiance à aux systèmes automatisés, même si ceux-ci échouent ou se comportent de manière imprévisible. Cette confiance excessive, parfois appelée biais d'automation, signifie que les utilisateurs investissent plus de confiance que nécessaire dans les systèmes automatisés (Parasuraman & Riley, 1997). La non-utilisation (*disuse*) est définie comme le rejet par les individus des systèmes automatisés (Lee & See, 2004, p. 50) parce qu'ils ne leur font pas confiance. L'abus (*abuse*) est défini comme l'automatisation de fonctions qui ne tient pas compte des conséquences pour les humains ou les organisations (par exemple, la robotisation conduisant au remplacement des travailleurs humains par des machines). Selon Parasuraman & Riley (1997), l'utilisation, la non-utilisation, la mauvaise utilisation et l'abus de l'automatisation sont le résultat d'interactions complexes entre de nombreux facteurs, tels que les sentiments et les attitudes des utilisateurs (par exemple, la propension à faire confiance), le risque, la charge de travail ou encore des traits de caractère comme la confiance en soi. Parmi ces facteurs, la confiance joue un rôle crucial : une confiance excessive conduit à une mauvaise utilisation (*misuse*) et à un abus (*abuse*), tandis que l'absence de confiance conduit à une absence d'utilisation (*disuse*). Il est donc important de mieux comprendre les raisons qui poussent les managers à utiliser, à mal utiliser ou à ne pas utiliser certaines procédures ou certains outils de sélection.

Dans une récente revue de la littérature, Glikson & Woolley (2020) ont mis en évidence un niveau de confiance initial dans les ADSS assez faible. Selon eux, la méfiance des utilisateurs est si élevée que ces derniers pourraient refuser d'utiliser les outils intégrant de l'IA. Ces auteurs soulignent également la rareté des études de terrain qui évaluent la confiance dans les ADSS (« *embedded AI* » dans leur article) dans des contextes organisationnels. Comme mentionné plus haut, cette rareté est particulièrement préoccupante dans le domaine de la sélection du personnel où le marché des outils d'embauche prédictifs est en pleine croissance alors que la recherche académique n'en est qu'à ses débuts (Campion et al., 2016; Langer et al., 2019, 2021; Oberst et al., 2020; Tambe et al., 2019). Pourtant, les enjeux sont importants dans la mesure où la sélection du personnel implique des questionnements éthiques, juridiques, psychologiques et stratégiques dans les organisations.

Dans le domaine de la sélection du personnel, l'intervention humaine demeure importante : le processus de recrutement n'est pas entièrement automatisé, ce qui laisse une marge de manœuvre à l'utilisateur pour choisir d'utiliser ou non l'ADSS. Comme le mentionnent Langer et al. (2021, p. 3), cette marge de choix justifie la nécessité pour les gestionnaires d'évaluer la fiabilité relative des humains et des systèmes automatisés afin de décider à qui ils doivent se fier pour réaliser une certaine tâche. Une meilleure compréhension des mécanismes régissant cette confiance et de ses conséquences sur le processus de décision des recruteurs est une question importante.

Par conséquent, la présente recherche vise à explorer l'impact relatif des différentes sources de recommandations mises à disposition des recruteurs amenés à faire un choix de présélection de candidats (ADSS vs. humaine), ainsi que les caractéristiques individuelles (propension à la confiance, traits de personnalité) sur leur niveau de confiance et leur comportement. Ce faisant, nous souhaitons contribuer à répondre à l'appel de Glikson et Wolley (2020) en questionnant le rôle de la confiance dans l'IA au sein des organisations.

2. De l'aversion algorithmique au biais d'automatisation : le rôle central de la confiance

Les études sur l'acceptation et l'utilisation des nouvelles technologies dans les organisations se concentrent sur les réactions des utilisateurs. On peut citer le modèle d'acceptation de la technologie (TAM ; Davis et al., 1989) qui considère l'utilité perçue et la facilité d'utilisation comme les principaux déterminants des attitudes des utilisateurs et de leur intention comportementale d'utiliser et d'accepter le système (Davis, 1989). Récemment, de nouveaux prédicteurs de l'acceptation de la technologie sont apparus, comme la confiance. Dans la présente recherche, nous considérons que, tout comme dans les relations interpersonnelles, la confiance est un facteur humain bien établi qui joue un rôle central dans la détermination de la volonté des individus de se fier aux systèmes automatisés dans des situations caractérisées par l'incertitude et qui démontrent une forte relation avec l'utilisation de l'automatisation et les interactions homme-machine (Hoff & Bashir, 2015; Mayer et al., 1995).

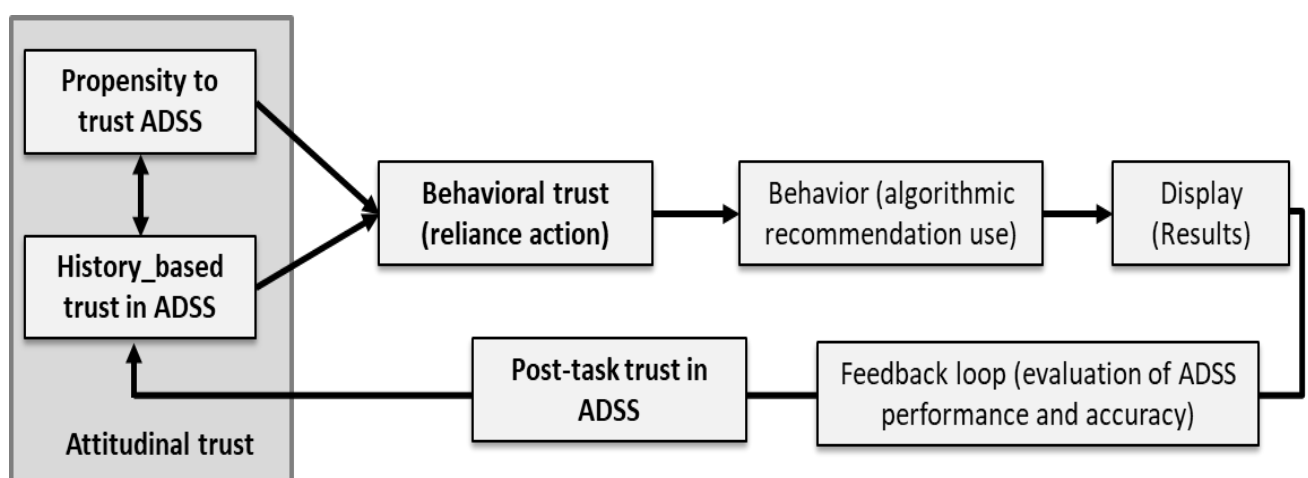
La confiance dans les systèmes automatisés a été étudiée dans le contexte interpersonnel, avec deux approches : la confiance comme une disposition stable ou la confiance comme une attitude dynamique, incluant une dimension comportementale. Concernant la confiance en tant que disposition stable, Merritt & Ilgen (2008) considèrent que les individus ont une propension générale à faire confiance ou à se méfier d'une machine, tout comme ils ont une propension générale à faire confiance ou à se méfier d'une autre personne. Selon eux, les facteurs qui influencent la confiance dispositionnelle varient peu au cours du temps, et ils influencent la prise de décision humaine lors des interactions avec un système autonome.

Une seconde approche considère que la confiance dans les systèmes automatisés est dynamique : elle peut varier en fonction des expériences passées avec le système (Muir & Moray, 1996). Sur la base de travaux antérieurs et de leurs propres résultats, Merritt et Ilgen (2008) proposent une distinction entre la propension à la confiance en tant que disposition stable (c'est-à-dire "*en général, je fais confiance à l'ADSS*") et la confiance basée sur l'histoire, fondée sur les interactions passées avec l'automatisation (c'est-à-dire "*d'après ma propre expérience, je fais confiance à cet ADSS*"). En se référant à la définition d'Azjen (1985) selon lequel l'attitude est le degré auquel une personne a une évaluation ou une appréciation favorable ou défavorable du comportement en question, ces deux dimensions de la confiance (une disposition stable et une confiance basée sur l'histoire) peuvent être considérées comme une confiance attitudinale.

Les modèles récents de la confiance dans les ADSS combinent la vision stable et dynamique de la confiance dans une perspective intégrative. On peut citer le modèle conceptuel de Lee & See (2004, p.68) de la dynamique régissant la confiance. Ce modèle prend ses racines dans le modèle conceptuel de la théorie de l'action raisonnée (Fishbein & Ajzen, 1975) ; il se base sur la distinction entre attitudes et comportements et est doté d'une boucle de rétroaction indiquant que le comportement du système a un effet de rétroaction sur la confiance. Dans ce modèle, la confiance attitudinale est un antécédent de l'action de confiance (*reliance action*), qui peut être considérée comme le résultat comportemental de la confiance (c'est-à-dire "*Je fais confiance aux conseils donnés par cet ADSS spécifique et je les utilise pour élaborer ma décision*"). L'action de confiance conduit à l'utilisation de l'automatisation : l'individu peut suivre la recommandation fournie par l'algorithmique (ADSS). Les recommandations faites par les systèmes sont ensuite évaluées par l'utilisateur dans le temps (ou instantanément, en les confrontant à sa propre évaluation). Cette confiance post-tâche qui découle de l'expérience de l'ADSS peut conduire un utilisateur à réviser son niveau de confiance initial : selon Meritt et Ilgens (2018), la confiance post-tâche contribue à construire une confiance basée sur l'historique.

Dans ce type de modèle, la confiance homme/machine est dynamique : le niveau de confiance initial peut augmenter ou diminuer au fil du temps, selon que les utilisateurs perçoivent les l'ADSS comme performant (**Figure 1**). Par conséquent, le niveau de confiance initial (confiance attitudinale) peut avoir un impact important sur l'utilisation de l'automatisation.

Figure 1 La dynamique de la confiance lors de l'interaction humain / ADSS



Un nombre important d'études sur les ADSS se concentre sur ce niveau initial de confiance et ses conséquences, avec une distinction entre la sur-confiance (*overtrust*) et la sous-confiance (*undertrust*), également définies comme aversion algorithmique et biais d'automatisation. La plupart des études sur les ADSS ont révélé de faibles niveaux de confiance initiale (aversion algorithmique) qui conduisant à une non-utilisation (*disuse*) de ces systèmes (Glikson & Woolley, 2020). L'aversion algorithmique se produit lorsque les utilisateurs ne font pas confiance aux systèmes automatisés, lorsqu'ils ne se fient pas aux décisions ou aux recommandations algorithmiques et préfèrent se fier à celles des humains (Dietvorst et al., 2015). Ce comportement est déjà bien documenté (Castelo et al., 2019; de Jong, 2020; Dietvorst, 2016; Dietvorst et al., 2018; Dietvorst & Bharti, 2020; Efendić et al., 2020; Filiz et al., 2021). Cette sous-confiance envers les systèmes automatisés conduit à leur non-utilisation

(*disuse*) (Lee & See, 2004). D'autre part, le biais d'automatisation se produit lorsque les utilisateurs font trop confiance aux systèmes automatisés, et lorsqu'ils se fient à tort aux algorithmes (Parasuraman & Riley, 1997). Le biais d'automatisation a été défini dans le contexte des systèmes de vol automatisés comme "des erreurs résultant de l'utilisation d'indices automatisés comme remplacement heuristique de la recherche et du traitement vigilants de l'information" (Mosier et al., 1998, p. 47). Cette confiance excessive dans les systèmes automatisés conduit à une mauvaise utilisation (*misuse*) de l'automatisation (Lee & See, 2004).

La plupart des recherches menées dans le domaine de l'aversion ou de la sur-confiance envers les algorithmes concernent les ADSS dédiés à des tâches relativement simples telles que le suivi de colis suspects (Merritt & Ilgen, 2008), la détection de cibles ((Dzindolet et al., 2003), la surveillance du fonctionnement d'un robot dans une usine de fabrication (J. Lee & Moray, 1992), ou la rédaction d'avis en ligne sur un service hôtelier (Jakesch et al., 2019).

Or, les tâches de sélection du personnel sont beaucoup plus complexes car elles requièrent des compétences hautement "humaines", le recruteur doit en effet faire face à des problèmes de complexité, de subjectivité et d'équité, et la décision qu'il prend est toujours le résultat d'un compromis entre plusieurs critères. La décision de sélection peut également avoir un impact sur la vie future du candidat, ce qui soulève des questions interpersonnelles, comme le montrent les recherches sur l'équité perçue dans la sélection du personnel (Konradt et al., 2017; Noble et al., 2021). La transposition des recherches antérieures sur l'automatisation au domaine de la sélection du personnel n'est donc pas simple et n'a pas encore été pleinement explorée : la question du biais d'automatisation ou de l'aversion algorithmique dans le contexte de la sélection du personnel mérite donc selon nous une plus grande attention.

3. Confiance et méfiance dans le contexte de la sélection du personnel

Des recherches antérieures ont établi que les caractéristiques des tâches réalisées par les ADSS jouent un rôle important. Lorsque les tâches impliquent des compétences humaines, les utilisateurs font davantage confiance à un expert humain qu'à un ADSS : c'est le cas lorsque les tâches impliquent une évaluation subjective du travail (M. K. Lee, 2018), un jugement éthique et moral (Bigman & Gray, 2018), ou plus généralement lorsque la tâche a un impact sur le destin de l'individu (Longoni et al., 2019). Toutes ces caractéristiques s'appliquent à la sélection du personnel, ce qui fait de la confiance dans les ADSS une question essentielle pour comprendre leur utilisation dans un contexte aussi complexe. Ces caractéristiques rendent également illusoire une délégation complète de cette tâche à des machines. C'est pourquoi l'automatisation complète des processus de sélection n'est pas vraiment à l'ordre du jour : les développeurs de solutions de recrutement prédictif préfèrent utiliser le terme "recrutement augmenté" plutôt que "recrutement automatisé" (Raisch & Krakowski, 2021). L'intervention humaine reste importante, et il semble plus pertinent d'étudier la confiance dans les ADSS lors de la sélection du personnel en conjonction avec l'expertise humaine, dans un contexte de coopération homme-machine.

Au cours du processus de recrutement, la présélection des candidats (via l'analyse des dossiers de candidatures) est une étape cruciale pour l'étude de l'impact des ADSS pour plusieurs raisons. Premièrement, cette étape est déjà impactée par les systèmes automatisés. La présélection des CV est une tâche fastidieuse pour laquelle il existe un large éventail de solutions, allant de l'extraction automatique et du classement des candidatures aux conseils pour choisir les candidats répondant le mieux aux critères du poste à pourvoir. Les ADSS constituent donc déjà une option viable en matière d'aide à la décision des recruteurs (Hickman et al., 2021). Deuxièmement, la présélection des CV est soumise à des heuristiques et les recherches ont

démontré un fort impact des stéréotypes et des préjugés (Derous & Ryan, 2019) : l'un des arguments les plus forts en faveur des ADSS (outre le gain de temps) est leur "objectivité" et leur capacité à réduire les biais humains. Troisièmement, la présélection des CV est une étape cruciale du processus de recrutement, où la plupart des candidats sont éliminés, et pour cette raison, elle peut être considérée comme risquée sur le plan éthique et juridique, surtout lorsque les recruteurs s'appuient fortement sur les ADSS.

Les études sur la confiance dans les ADSS au cours du processus de recrutement demeurent à ce jour peu nombreuses (Burton et al., 2020). La plupart d'entre elles ont révélé une tendance à la non utilisation (*disuse*) de l'automatisation et mis en évidence un comportement d'aversion. Les discussions entre un recruteur et un candidat sont perçues comme plus justes et plus flexibles qu'une recommandation algorithmique (Diab et al., 2011). Dans la lignée de ces résultats, Langer et al. (2021) ont conclu que les individus dans un contexte de sélection du personnel percevaient la recommandation humaine comme plus digne de confiance que celle d'un système automatisé. Selon les auteurs, ces résultats peuvent s'expliquer par la capacité perçue plus forte de l'humain à accomplir des tâches où des questions éthiques doivent être prises en compte. Des résultats convergents montrent que lorsque la tâche est perçue comme subjective ou exigeant des compétences humaines, elle est jugée comme plus précise et plus digne de confiance lorsqu'elle est effectuée par un humain plutôt que par un système d'aide à la décision (Castelo et al., 2019 ; M. K. Lee, 2018). Cette préférence pour les recommandations humaines existe d'ailleurs même si cette recommandation est fortement subjective (Oberst et al., 2020) ou dans la situation où les systèmes automatisés font des erreurs dans une tâche similaire à un recrutement (Dietvorst et al., 2015). Pour les décisions managériales, dans un contexte de coopération homme-machine, une étude montre par exemple que les managers préféreraient que les recommandations humaines aient une influence de 70% sur les décisions finales, par rapport aux recommandations algorithmiques (Haesevoets et al., 2021). Selon les recherches, il apparaît que les systèmes automatisés ne sont ni capables d'accomplir des tâches où des questions éthiques doivent être prises en compte (Langer et al., 2021), ni capables d'accomplir des tâches subjectives qui nécessitent des compétences hautement humaines (Castelo et al., 2019). Par conséquent, nous soutenons qu'au stade de la présélection des CV, les recruteurs préfèrent la recommandation humaine et sont plus influencés par cette recommandation. Par conséquent, nous posons les hypothèses suivantes :

H1-1 : Les recruteurs considèrent que la recommandation d'un expert humain est plus fiable que celle de l'ADSS.

H1-2 : Les recruteurs sont plus influencés dans leur tâche de présélection par une recommandation humaine que par une recommandation émanant d'un ADSS.

Dans leur revue, Glikson & Wooley (2020) ont montré l'effet durable des erreurs des systèmes automatisés sur la confiance des utilisateurs. Lorsque les ADSS commettent des erreurs, la confiance de l'utilisateur et sa dépendance à l'égard de l'ADSS diminuent de manière significative (Dzindolet et al., 2003). Dans la lignée de ces résultats, Dietvorst et ses collègues (2015) ont mis en évidence l'effet néfaste d'une erreur : les utilisateurs préfèrent se fier aux prévisions humaines plutôt qu'aux prévisions algorithmiques, même en cas d'erreurs humaines. Par conséquent, nous posons les hypothèses suivantes :

H2-1 : les recruteurs considèrent comme plus fiable une recommandation non pertinente provenant d'un expert humain par rapport à une recommandation algorithmique non pertinente.

H2-2 : les recruteurs sont plus influencés dans leur tâche de présélection par une recommandation humaine non pertinente que par une recommandation algorithmique non pertinente.

La recherche a également mis en évidence des associations entre plusieurs traits de personnalité et la confiance. En ce qui concerne les traits de personnalité du *Big Five*, le névrosisme est corrélé négativement à l'accord avec les systèmes automatisés (McBride et al., 2012; Szalma & Taylor, 2011). D'autre part, l'extraversion entraîne une plus grande propension à faire confiance aux machines que l'introversion (Merritt & Ilgen, 2008). Les traits de disposition, tels que la propension générale à faire confiance et la confiance en soi, se sont également avérés avoir un effet sur la confiance dans les systèmes automatisés. La propension à faire confiance aux ADSS s'est avérée être un antécédent du comportement de confiance (*reliance action*) dans plusieurs études impliquant une interaction entre les systèmes automatisés et les humains (Jessup et al., 2019; Singh et al., 1993), conformément au modèle dynamique de la confiance dans l'automatisation (J. D. Lee & See, 2004).

La confiance en soi est connue pour influencer la formation de la confiance. Lee & Moray (1994) ont ainsi démontré dans une étude impliquant des contrôleurs automatiques que la confiance dans l'automatisation et la confiance en soi dans la capacité de contrôler manuellement le processus sont négativement liées : l'utilisation de l'automatisation est favorisée lorsque la confiance dans l'automatisation dépasse la confiance en soi, et vice versa.

Tous ces résultats ont été tirés d'études menées sur des tâches impliquant une collaboration homme-machine. Les tâches décrites dans ces études sont bien moins complexes que celles effectuées par les recruteurs, mais considérant que les traits de personnalité et de disposition sont relativement indépendants du contexte de l'interaction homme-machine (car ils sont considérés comme des antécédents de ces comportements d'interaction), nous proposons une transposition des résultats précédents à l'utilisation de l'ADSS dans la tâche de sélection des CV. Nous posons donc les hypothèses suivantes :

H3-1 : L'extraversion est positivement liée à la propension à faire confiance à l'algorithme (a) et au comportement de confiance (b).

H3-2 : Le neuroticisme est négativement lié à la propension à faire confiance à l'algorithme (a) et au comportement de confiance (b).

H3-3 : La propension des recruteurs à faire confiance à l'ADSS est positivement associée au comportement de confiance.

H3-4 : La confiance en soi dans le recrutement est négativement liée à la propension à faire confiance à l'algorithme (a) et conduit à un comportement de confiance plus faible (b).

4. Méthodologie

4.1 Participants

Dans cette recherche, nous avons choisi d'analyser la confiance des participants et leurs comportements (c'est-à-dire la notation et le classement de dossiers de candidature) après avoir été exposés à des recommandations. Dans un souci de validité écologique de nos résultats, nous avons sélectionné un échantillon de professionnels impliqués dans le processus de recrutement au sein de l'organisation dans laquelle ils travaillent. Dans les recherches précédentes sur la confiance dans le contexte de la sélection, les participants impliqués étaient principalement des étudiants ou des adultes actifs (Diab et al., 2011 ; Dietvorst, 2016 ; Dietvorst et al., 2015 ;

Langer et al., 2021 ; M. K. Lee, 2018) et seules quelques études ont recouru à des échantillons de participants réellement impliqués dans des tâches de recrutement dans leur organisation (Oberst et al., 2020). Les participants ont été sélectionnés à partir d'un panel en ligne appartenant à une société spécialisée. Panelabs® donne accès à un panel qui contient 500 000 participants français et fournit un haut niveau de contrôle de qualité requis par les chercheurs. Les données de panel en ligne sont de plus en plus utilisées pour obtenir des échantillons dans les études, notamment en comportement organisationnel (Bowling & Lyons, 2015; Bruk-Lee et al., 2016; Djurdjevic et al., 2019; Krishnakumar et al., 2019; Lacroux & Martin-Lacroux, 2020). Nous avons obtenu 746 réponses exploitables.

L'âge des répondants s'échelonne de 20 à 69 ans avec une moyenne de 40,5 ans (ET = 9,97) ; 47,9 % sont des hommes et 52,1 % des femmes. Ils sont en emploi dans des entreprises comptant de 10 à plus de 1000 salariés (avec une classe médiane de 50 à 199). En ce qui concerne leur expérience professionnelle en matière de sélection du personnel, 50,8 % des répondants présentent une expérience significative de 3 ans ou plus, et 34,4 % une expérience de 1 à 3 ans.

4.2 Procédure

Les participants ont volontairement participé à l'expérimentation en ligne. Il leur a été demandé de lire une description de poste pour un responsable des ressources humaines. Ensuite, deux extraits de CV leur ont été proposés, de niveau de qualification différent (voir **4.6 Etude pilote**). L'ordre de présentation de ces deux CV a été contrebalancé pour éviter les effets d'ordre. Les participants ont reçu des recommandations synchrones en complément de la description de l'offre et des deux extraits de CV : ils ont été exposés de manière aléatoire à 5 conditions, dans un plan inter-sujet 2×2 , avec un groupe de contrôle dans lequel les participants n'étaient exposés à aucune recommandation. Les facteurs manipulés sont le type de recommandation (humaine ou algorithmique) et la pertinence de la recommandation (pertinente ou non pertinente).

Nous avons considéré comme non pertinente la recommandation (humaine ou algorithmique) pour laquelle la candidature la moins qualifiée était recommandée ; en revanche, la recommandation pertinente (humaine ou algorithmique) était celle qui recommandait le meilleur candidat, c'est à dire celui présentant le niveau de qualification le plus en phase avec les exigences du poste. Dans le cadre de cette tâche expérimentale, les participants ont été invités à remplir un questionnaire dans lequel des questions d'attention et de contrôle de la qualité des manipulations ont été insérées. Après avoir lu la description du poste, les extraits des deux CV et la recommandation sélectionnée, les participants devaient évaluer les deux candidats et remplir un questionnaire (voir **4.4 Mesures**).

4.3 Matériel expérimental

Pour l'offre d'emploi, nous avons sélectionné un poste de responsable des ressources humaines et élaboré une description de poste inspirée de postes similaires publiés sur des sites d'emploi.

Dans un deuxième temps, deux extraits de CV ont été élaborés avec le même modèle et des rubriques conformes aux CV publiés sur des sites d'emploi (nom, adresse personnelle et adresse électronique, formation, expériences professionnelles, compétences, maîtrise de langues étrangères). Ces deux candidatures ne se différenciaient qu'en termes de niveau d'expérience professionnelle, de niveau de diplôme, de niveau de maîtrise d'une langue étrangère et de niveau de compétences de gestion. L'un des CV correspondait exactement à la description du poste en termes de niveau d'études, d'expérience professionnelle, de maîtrise des langues étrangères et de compétences de gestion, tandis que le second était moins adapté (niveau d'études inférieur, expérience professionnelle inférieure, compétences de gestion et maîtrise des langues

étrangères inférieures). Pour des raisons de réalisme, les deux candidatures étaient toutefois conformes à l'offre d'emploi (en termes de formation et d'expériences professionnelle). Quatre experts en gestion des ressources humaines ont examiné l'offre d'emploi ainsi que les deux candidatures afin de s'assurer de leur réalisme et de la qualité du matériel expérimental.

Dans un troisième temps, nous avons créé cinq conditions concernant les recommandations accompagnant les CV (**Annexe 1** pour la présentation de deux conditions). Les recommandations de l'expert humain consistaient en des extraits de conversation téléphonique avec le responsable d'une agence de recrutement (l'expert humain) à laquelle l'entreprise recruteuse avait déjà fait appel à plusieurs reprises pour de précédents recrutements. Les participants étaient informés que le responsable de l'agence de recrutement avait reçu la description du poste ainsi que les deux CV des candidats. Les recommandations algorithmiques consistaient en une synthèse écrite issue d'une solution de recrutement prédictif développée par une startup fictive (*HR Predict*) spécialisée dans le développement de solutions algorithmiques avancées intégrant les derniers développements en matière d'intelligence artificielle (apprentissage automatique et adaptatif). Les participants ont été exposés aux recommandations de *HR predict* qui proposaient de classer les candidats en fonction de leur taux d'adéquation avec la description du poste.

4.4 Mesures

Nous avons d'abord recueilli différentes mesures de traits de caractère. La personnalité des participants a été évaluée à l'aide d'une version courte de 10 items de l'inventaire *Big Five* (Rammstedt & John, 2007). Un exemple d'item est " *Je me vois comme quelqu'un de détendu, qui gère bien le stress* " (alpha Cronbach moyen = 0,57). Pour mesurer la confiance des participants dans leur capacité à recruter, nous avons adapté l'échelle de compétence perçue de Williams et al., 1998; Williams & Deci, 1996. La propension à faire confiance aux systèmes automatisés a été évaluée chez les participants exposés aléatoirement aux conditions algorithmiques à l'aide de l'échelle de confiance entre les personnes et l'automatisation adaptée de Jian et al. (2000). Deux exemples d'items Likert à 7 points sont : " *Les ADSS sont trompeurs* (codage inverse) " ; " *Les ADSS sont fiables* " (Alpha de Cronbach = 0,84).

Les participants ont ensuite été invités à lire la description du poste, les extraits des deux CV et ont été assignés au hasard à une condition de recommandation. Il leur a demandé de choisir la meilleure candidature et de les évaluer toutes les deux en les notant sur une échelle de 10 points. Pour évaluer l'influence des recommandations sur le comportement des participants, nous avons utilisé deux types de mesures. Tout d'abord, nous avons calculé la différence de score entre le meilleur et le second CV (variable *diffscore*) : plus cette différence est importante, plus la recommandation a eu d'impact. Si un *diffscore* est négatif pour une condition particulière, cela indique que les participants préfèrent le second CV au meilleur. Ensuite, nous avons calculé et comparé les fréquences des CV qui ont été classés en première position.

La confiance dans l'expert humain ou dans les recommandations de l'ADSS a été mesurée à l'aide de deux items proposés par Dietvorst (2016) et notés sur une échelle de Likert en 9 points : " *Dans quelle mesure êtes-vous digne de confiance dans la recommandation fournie par [l'expert/la solution algorithmique] ?* " et " *Dans quelle mesure pensez-vous qu'il est probable que [la solution algorithmique/l'expert] prenne de mauvaises décisions de sélection ?* ". Compte tenu du fait que ces mesures ont été prises après la réalisation de la tâche, nous considérons qu'elles permettent d'évaluer la confiance post-tâche envers l'expert ou l'automatisation.

4.5 Variables de contrôle

Plusieurs recherches ont montré que les individus dotés d'une plus grande expertise dans un domaine sont moins susceptibles de se fier à l'automatisation que des individus novices (Dikmen & Burns, 2022; Sanchez et al., 2014). Ainsi, une expérimentation récente a permis de conclure que les individus ayant une expérience de la conduite de véhicules agricoles étaient plus réticents à se fier aux alarmes automatisées pendant une tâche d'évitement de collision que les individus ayant peu ou pas d'expérience agricole (Sanchez, Rogers, Fisk, & Rovira, 2011). Dans une expérimentation manipulant la connaissance dans un domaine, les participants avec un accès à la connaissance du domaine se sont moins fiés à l'assistant intégrant l'IA lorsque celui-ci était incorrect, en comparaison avec des participants sans accès à la connaissance (Dikmen & Burns, 2022). Parce que le niveau d'expertise dans une tâche spécifique a un impact sur la confiance et les croyances concernant l'automatisation de cette tâche, nous avons également mesuré l'expertise professionnelle des participants en matière de recrutement avec deux questions " *Depuis combien d'années êtes-vous impliqué dans ces procédures de sélection/recrutement ?*" et " *En moyenne, au cours de chacune de ces procédures de sélection, combien de candidatures examinez-vous ?*" Un score composite représentant l'expertise en matière de recrutement a ainsi été calculé comme $\log(\text{expérience} \times \text{nombre de candidatures examinées})$. En lien avec leur expertise, nous avons également mesuré la confiance du participant dans sa propre décision en utilisant un item proposé par Dietvorst (2016), mesuré sur une échelle de Likert à 9 points (" *Quelle confiance accordez-vous à votre propre décision ?*").

4.6 Etude pilote et contrôle de la qualité des manipulations

Avant l'étude principale, nous avons réalisé une étude pilote sur un premier échantillon de 164 recruteurs fourni par la société de sondage en ligne. Notre objectif était de vérifier que les deux CV étaient correctement distingués et que les manipulations des recommandations étaient efficaces. Les résultats montrent que dans le groupe témoin (condition sans recommandation), les deux candidatures ont été significativement différenciées par les participants (différence moyenne = 0,26, T-test apparié : $t(163) = 3,32, p = .018$) et que le meilleur CV était plus souvent classé en premier que le moins pertinent (respectivement 62,5% contre 37,5%). Concernant l'impact des conditions expérimentales, une ANOVA de la différence entre les scores des deux CV a révélé que cet impact différait significativement selon les conditions expérimentales ($F(4,159) = 2.47, p = 0.04, \eta^2 = 0.06$). Ces résultats indiquent que les participants ont bien pris en compte les recommandations lors de l'évaluation des deux candidatures.

5. Résultats

5.1 Hypothèses relatives à la confiance des recruteurs

En ce qui concerne les hypothèses sur la confiance, H1-1 stipule que les recruteurs considèrent comme plus digne de confiance une recommandation provenant d'un expert humain par rapport à une recommandation algorithmique. Les résultats montrent que les moyennes de la confiance perçue par type de recommandation diffèrent ($m = 6,54, sd = 1,32$ pour la recommandation algorithmique ; $m = 6,95, sd = 1,08$ pour la recommandation d'un expert humain). La différence est statistiquement significative, avec un effet modéré : différence moyenne = 0,406 [0,19, 0,62] ; test t de Welch : $t(457,6) = -3,64, p < .001, d$ de Cohen = -0,34 [-0,52, -0,15]). Par conséquent, **l'hypothèse H1-1 est supportée**. L'hypothèse 2-1 stipule que les recruteurs considèrent comme plus digne de confiance une recommandation non pertinente provenant d'un expert humain par rapport à une recommandation algorithmique. Les moyennes de la confiance perçue par type de recommandation diffèrent ($m = 6,51$ et $sd = 1,24$ pour la recommandation algorithmique ; $m = 6,93$ et $sd = 1,1$ pour la recommandation d'un expert humain). La différence est statistiquement significative, avec un effet modéré : différence moyenne = 0,42 [0,16, 0,69] ;

test t de Welch : $t(301,3) = -3,16$, $p = 0,002$, d de Cohen = $-0,36$ $[-0,13, -0,58]$). **L'hypothèse H2-1 est donc supportée.**

5.5 Hypothèses relatives au comportement des recruteurs

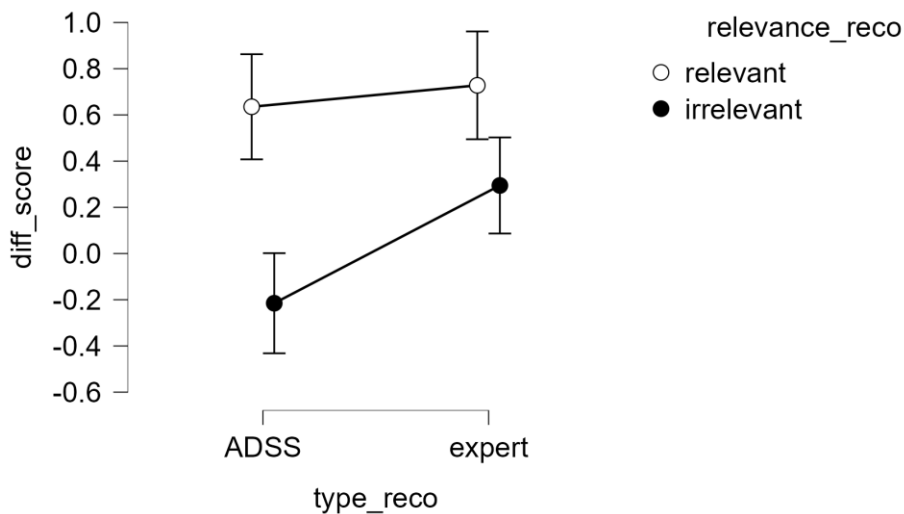
En ce qui concerne les hypothèses relatives au comportement des recruteurs, l'hypothèse H1-2 pose que les recruteurs sont plus influencés par la recommandation humaine que par la recommandation algorithmique (c'est-à-dire que, par rapport au groupe de contrôle, la recommandation humaine augmente davantage la différence de score entre le meilleur et le second CV que la recommandation algorithmique). Les moyennes des diffscores par type de recommandation diffèrent ($m = 0,55$ et $sd = 1,36$ pour le groupe de contrôle ; $m = 0,20$ $sd = 1,45$ pour la recommandation algorithmique ; $m = 0,51$ $sd = 1,37$ pour la recommandation humaine). La différence est statistiquement significative, avec un faible effet : différence moyenne = $0,406$ $[0,19, 0,62]$; ANOVA (Welch) : $F(2,400.3) = 4.96$, $p = 0,007$, $\eta^2 = 0.013$). L'analyse de contrastes simples (**Tableau 1**) révèle qu'il n'y a pas de différence statistiquement significative entre la condition de recommandation d'expert et le groupe de contrôle, mais nous avons trouvé une différence significative entre la recommandation algorithmique et le groupe de contrôle. Par conséquent, **l'hypothèse H1-2 n'est pas supportée**. Nos résultats suggèrent que les recruteurs sont plus influencés par la recommandation algorithmique que par la recommandation d'un expert et que cette influence ne va pas dans le sens attendu (la différence de score entre le meilleur et le second CV a été réduite). L'analyse des décisions de classement confirme cette tendance : le pourcentage de meilleurs CV classés en première position lorsque le participant a reçu une recommandation humaine (66,5%) est plus élevé que dans le groupe de contrôle (64,2%), mais la différence n'est pas significative selon un test Z de proportions ($Z = 0,40$, $p = 0,68$). Inversement, le pourcentage de meilleurs CV classés en première position lorsque le participant a reçu une recommandation algorithmique (55,8 %) est inférieur à celui du groupe de contrôle (64,2 %), la différence est plus importante que dans le cas précédent, mais non significative ($Z = 1,77$, $p = 0,07$).

Tableau 1 : Contrastes simples – type de recommandation

Comparaison	Estimation	95% CI		SE	df	t	p
		Inf	Sup				
Recom algo - control	-0.355	-0.629	-0.081	0.140	743	-2.539	0.011
Recom expert - control	-0.039	-0.315	0.237	0.141	743	-0.278	0.781

Nous avons proposé H2-2 qui pose que les recruteurs sont plus influencés par une recommandation humaine non pertinente que par une recommandation algorithmique non pertinente. Pour tester cette hypothèse, nous avons effectué une ANOVA factorielle en tenant compte du type et de la pertinence des recommandations, qui a révélé que les deux facteurs ont un effet direct sur les diff_scores (pour le type : $F(1,595) = 7.2$, $p = 0.007$, $\eta^2 = 0.011$; pour la pertinence : $F(1,463) = 32,73$, $p < .001$, $\eta^2 = 0,05$) avec une interaction non significative ($F(1,595) = 3,45$, $p = .06$, $\eta^2 = .005$). La pertinence de la recommandation a un effet plus fort que le type de recommandation comme le montre la **Figure 2**.

Figure 2 Influence du type de recommandation et de sa pertinence sur diffscore



L'analyse de contrastes simples (**Tableau 2**) indique une différence significative entre les recommandations algorithmiques et humaines, mais dans une direction différente de celle attendue : les participants ont été plus influencés par une recommandation algorithmique non pertinente que par une recommandation humaine non pertinente : en présence d'une recommandation algorithmique non pertinente, les recruteurs ont favorisé le deuxième CV au détriment du meilleur (la moyenne de diff_score est négative). Par conséquent, **l'hypothèse H2-2 n'est pas supportée**.

L'analyse des décisions de classement confirme cette tendance : le pourcentage de meilleurs CV classés en première position est plus élevé lorsque le participant a reçu une recommandation humaine non pertinente (59,6%) par rapport à une recommandation algorithmique non pertinente (43,6%), la différence est significative selon un test Z pour les proportions ($Z = 2,77$, $p = .005$). Cela signifie que les participants ont été plus influencés par la recommandation algorithmique non pertinente, ce qui les a amenés à favoriser le second CV.

Tableau 2 : Statistiques descriptives et contrastes (VD = diff_score)

Type de recommandation	Pertinence	Mean	SD	N				
algorithmique	Non pertinente	-0.183	1.382	126				
	Pertinente	0.690	1.392	116				
Humaine	Non pertinente	0.345	1.259	113				
	Pertinente	0.777	1.463	112				
Contraste	Estimation	95% CI LL	95% CI UL	SE	df	t	p	
Non pertinente algorithmique- non pertinente humaine	-0.510	-0.810	-0.210	0.153	302	-3.340	<.001	

5.3 Hypothèses relatives aux traits dispositionnels du recruteur

Concernant l'influence de la personnalité et des traits de disposition sur la confiance et le comportement, une matrice de corrélation sur un sous-échantillon de participants exposés à des recommandations algorithmiques (N=242) a été traitée pour tester les hypothèses sur les associations entre les variables (**Tableau 3**).

Les traits de personnalité ont une influence limitée sur les attitudes envers les recommandations algorithmiques et le comportement ultérieur sur notre échantillon. Pour affiner l'analyse en prenant en compte l'impact de la pertinence des recommandations identifié dans H2.2, nous avons calculé les corrélations entre les traits de personnalité et la confiance sur des sous-échantillons, et avons distingué les conditions pertinentes et non pertinentes (**Tableau 4**). Lorsque les recruteurs sont exposés à des recommandations algorithmiques non pertinentes, l'extraversion influence leur comportement de confiance. Cela pourrait s'expliquer par la dimension grégaire de l'extraversion qui augmente la tendance à faire confiance aux gens (Freitag & Bauer, 2016) : cette tendance pourrait se transposer au contexte de la confiance dans les systèmes automatisés, suivant les travaux de Mayer et al. (1995). D'autre part, nos résultats suggèrent que lorsque les recruteurs présentent un score élevé sur le trait névrosisme et sont exposés à une recommandation algorithmique non pertinente, ils sont moins influencés dans leur comportement de confiance. Ceci n'est pas surprenant dans la mesure où les individus présentant ce trait de caractère sont connus pour être plus prudents et pour prendre leurs décisions avec plus de précaution.

Table 3 : Corrélations par paires entre les traits de personnalité, la propension à faire confiance et la confiance dans les recommandations algorithmiques

		Pearson	r
Propension confiance algorithmes	- Extraversion	0.116	*
Propension confiance algorithmes	- Névosisme	-0.132	*
Propension confiance algorithmes	- Confiance dans sa capacité à recruter	0.141	*
Confiance dans la recom. ADSS	- Propension confiance algorithmes	0.608	***
Confiance dans la recom. ADSS	- Extraversion	0.156	**
Confiance dans la recom. ADSS	- Névosisme	-0.094	
Confiance dans la recom. ADSS	- Confiance capacité à recruter	0.224	***

* $p < .05$, ** $p < .01$, *** $p < .001$

Tableau 4. Corrélations par paires entre les traits de personnalité et la confiance dans les recommandations algorithmiques (conditions pertinentes vs. non pertinentes)

Condition cohérence des recommandations	Pearson r
Confiance dans la recom. ADSS - Extraversion	0.132
Confiance dans la recom. ADSS - Nevrocisme	0.010
Condition incohérence des recommandations	Pearson r
Confiance dans la reco. ADSS - Extraversion	0.201*
Confiance dans la recom. ADSS - Nevrocisme	-0.201*

* $p < .05$

L'ensemble des tests de nos hypothèses est présenté ci-dessous (**Tableau 5**) :

Tableau 5 : Synthèse des tests d'hypothèses

Hypothèse	Test	Résultat
H1-1	T-test sur la confiance (post-tâche)	Supportée
H1-2	ANOVA sur Diffscores avec contraste	Non supportée
H2-1	T-test sur la confiance (post-tâche) pour la condition avec la recommandation algorithmique incohérente	Supportée
H2-2	ANOVA sur Diffscores sur les conditions recommandations incohérentes avec contraste	Non supportée
H3-1	Corrélation Extraversnion – Confiance (propension & post-tâche)	Supportée
H3-2	Corrélation Neuroticisme-Confiance (propension a) et post-tâche b))	(a) Supportée (b) Non supportée
H3-3	Corrélation Confiance Propension-post-tâche	Supportée
H3-4	Corrélation confiance en sa capacité de recruteur- confiance (propension & post-tâche)	Non supportée

6.2 Discussion

6.3.6.1 Discussion Générale

Dans un contexte où les recherches demeurent limitées lorsqu'il s'agit de comprendre les mécanismes de confiance dans les systèmes automatisés, en particulier dans des contextes où les décisions ont un impact sur le devenir d'individus (ici les candidats), nos résultats contribuent à une meilleure compréhension de ces processus de confiance.

Le résultat le plus notable est toutefois paradoxal. Dans un premier temps, nous avons montré que les recruteurs font plus confiance à la recommandation d'un expert humain qu'à celle d'un système d'aide à la décision automatisé. Dans le prolongement de l'étude de Langer et ses collègues (2021), nos résultats indiquent que les recruteurs ne considèrent pas que les systèmes automatisés soient plus performants en matière de présélection de candidats, comme en témoignent leur évaluation de la fiabilité post-tâche pour les ADSS (plus faible que celle des recommandations d'un expert humain).

Pourtant, nous avons montré dans un second temps que les recruteurs semblent davantage influencés par l'ADSS dans leur comportement, lorsqu'ils doivent classer les candidatures. Lorsque les recommandations fournies aux recruteurs ne sont pas pertinentes, ils sont également davantage influencés par la recommandation algorithmique. Ce résultat peut s'expliquer par le fait que les recruteurs peuvent être convaincus que l'ADSS est moins susceptible de faire des erreurs ou bien que les recruteurs considèrent que le fonctionnement de l'ADSS est moins biaisé que la recommandation émanant d'un individu. Cela pourrait générer une confiance excessive et des conséquences négatives potentielles. Ainsi, nos résultats mettent en évidence un possible biais d'automatisation : dans notre échantillon, les recruteurs font une confiance excessive aux recommandations algorithmiques non pertinentes. Cette confiance excessive dans les ADSS peut conduire à une mauvaise utilisation des systèmes automatisés (*misuse* au sens de Lee & See, 2004) et soulève des questions liées aux risques associés à l'utilisation de ADSS inefficaces ou imprécis : manque de qualité et de reproductibilité des décisions algorithmiques, opacité des algorithmes, inexplicabilité des décisions, biais et risques de discrimination (Besse et al., 2018).

Nos résultats ne correspondent pas complètement à la conclusion répandue selon laquelle les gens se méfient des algorithmes dans le domaine de la sélection de collaborateurs. Nos résultats mettent plutôt en évidence que les recruteurs ont certes une plus grande confiance envers les

experts humains (conformément à Diab et al., 2011 ; M. K. Lee, 2018 et Oberst et al., 2020), mais qu'ils ne se sont pas comportés en conséquence en présence de recommandations algorithmiques non pertinentes. Nos résultats mettent ainsi en évidence un décalage entre la confiance déclarée des recruteurs envers l'expert humain et leur comportement dans le cas d'une recommandation algorithmique non pertinente pour laquelle l'ADSS est plus suivie.

En considérant que le biais d'automation se définit comme la tendance à suivre une décision algorithmique si elle est incorrecte (Parasuraman & Manzey, 2010), nous avons souhaité explorer les prédicteurs de ce biais sur le sous-échantillon dans lequel les participants ont été exposés à une telle condition (N= 156). Nous avons créé une variable binaire définissant le biais d'automation comme la situation dans laquelle l'ADSS recommande le CV le moins qualifié et que les participants le classent en première position. Nous avons ensuite construit un modèle de régression logistique pour identifier les prédicteurs possibles, en utilisant une procédure par étapes (stepwise). Nous avons introduit dans le modèle les variables utilisées précédemment ainsi que différentes variables démographiques et de personnalité non utilisées précédemment (les traits de caractère que sont l'ouverture, l'agréabilité et la conscience mais également le sexe du recruteur, son expertise, la comparaison des participants entre leur propre évaluation et celle de l'ADSS). Le meilleur modèle sélectionné après la procédure *stepwise* ne retient au final que trois prédicteurs (**Tableau 6**).

Tableau 6 : Prédicteurs du biais d'automation (régression logistique, procédure stepwise)

Paramètre	Est	SE robust	CI 95% inf	CI 95% sup	Odds Ratio
Comparaison soit-ADSS	0.428	0.115	0.202	0.653	1.534
Conscience	-0.492	0.130	-0.746	-0.238	0.611
Expertise du recruteur	0.305	0.217	-0.120	0.731	1.357

CI95% obtenu par bootstrap accéléré corrigé du biais (1000 samples)

R² (Nagelkerke) = 0.176

Concernant le trait Conscience, nos résultats sont cohérents avec d'autres recherches suggérant que la confiance est négativement liée à la Conscience dans le cas des individus (Freitag & Bauer, 2016). Les personnes présentant ce trait de personnalité sont plus prudentes et mieux informées, prennent des décisions avec soin et ne font pas facilement confiance aux actions ou aux décisions des autres. De plus, elles se considèrent généralement comme très compétentes (McCrae & Costa, 2003). C'est peut-être la raison pour laquelle elles semblent moins influencées par les recommandations algorithmiques non pertinentes.

De manière surprenante, nos résultats montrent que l'expertise du recruteur est positivement liée au biais d'automation, ce qui ne correspond pas aux résultats antérieurs indiquant que les personnes avec une plus grande expertise du domaine sont moins susceptibles de se fier à l'automatisation que les opérateurs novices (Dikmen & Burns, 2022 ; Sanchez et al., 2014). Nos résultats indiquent en effet que les individus expérimentés dans le recrutement se fient davantage à l'assistant d'IA lorsque celui-ci est incorrect. Pour expliquer ce résultat paradoxal, nous pouvons avancer que le niveau d'expertise déclaré par les recruteurs concerne le processus de recrutement " traditionnel " (et non le processus de recrutement assisté par algorithme). L'expertise professionnelle des recruteurs n'est pas facilement transférable aux ADSS en raison de la grande difficulté à comprendre leur fonctionnement : la plupart des ADSS basés sur l'intelligence artificielle sont opaques pour leurs utilisateurs, pour des raisons techniques (utilisation d'algorithmes d'apprentissage supervisé complexes) et commerciales (secret industriel). Ce "problème de la boîte noire" est une question émergente dans la gestion des ressources humaines. Le fait que les utilisateurs soient confrontés à des systèmes automatisés

prévisibles mais opaques et inexplicables semble augmenter leur aversion algorithmique, selon une étude basée sur des entretiens (Ochmann et al., 2021). Ochmann et ses collègues ont conclu que lorsque les utilisateurs sont conscients du manque de transparence et d'explicabilité des recommandations basées sur l'IA, leur attitude à l'égard de ces dernières est négative. Dans notre étude, seuls 5% des recruteurs de notre échantillon déclarent utiliser l'ADSS régulièrement et 62,5% ne l'ont jamais utilisé. Nous pensons donc que les recruteurs n'étaient peut-être pas conscients du problème de la boîte noire, et que cette ignorance les a encouragés à suivre les recommandations, même incohérentes, de l'ADSS. Une fois de plus, ces conclusions sont en accord avec nos résultats sur les variables attitudinales (confiance) mais pas sur les variables comportementales (classement des candidats). Le manque de confiance n'est donc pas directement lié à la non-utilisation de l'ADSS.

6.4 Limites et voies de recherches futures

Notre travail comporte des limites qui constituent autant de pistes de recherches futures. Nous avons par exemple constaté de faibles effets des recommandations sur la confiance et le comportement. Nous pouvons expliquer ces résultats par notre recherche de validité écologique. Dans un souci de réalisme et de crédibilité, nous avons créé des dossiers de candidature assez proches en termes de qualification. Par conséquent, ces candidatures ne différaient pas tellement, ce qui explique les tailles d'effet limitées.

Ensuite, d'autres variables identifiées dans la littérature pourraient être prises en compte pour mieux comprendre les antécédents du comportement de confiance des recruteurs comme la charge de travail, ou les contraintes de temps (Lee et See, 2004). Nous pensons également qu'il serait intéressant d'étudier dans quelle mesure fournir aux recruteurs des doubles recommandations humaines et algorithmiques pourraient influencer leurs attitudes et comportement et les interactions entre ces deux types de recommandations.

Une question importante à aborder dans les travaux futurs sur les outils de recrutement prédictifs concerne le " problème de la boîte noire ". Dans le domaine de l'acceptation de la technologie, Kroenung & Eckhardt (2015) ont proposé de distinguer l'utilisation volontaire et obligatoire des systèmes automatisés. Dans notre étude, comme dans de nombreuses situations dans les processus de la vie réelle, les recruteurs ne " choisissent " pas vraiment d'utiliser ce type de systèmes. Ils se trouvent dans un contexte dans lequel le recours à ces outils est imposé. Compte tenu des coûts et des problèmes de mise en œuvre des solutions d'aide à la décision automatisées, l'utilisation ou non de ces systèmes dans une organisation est une décision stratégique prise par la direction générale à laquelle les recruteurs doivent se conformer. Dans l'étude d'Ochmann et al. (2021), un contexte obligatoire semble renforcer l'aversion pour les ADSS dans les discours des utilisateurs, par rapport à un contexte dans lequel ce recours est volontaire. Cependant, il est à noter qu'Ochmann et ses collègues n'ont pas mesuré les comportements réels mais les attitudes des individus, ce qui ouvre une perspective de recherche prometteuse sur le lien entre attitude et comportement en matière d'usage de systèmes d'aide à la décision automatisés, en lien avec nos propres résultats qui semblent démontrer un manque de cohérence entre attitudes et comportements dans la présélection de CV assistée par l'IA.

Au niveau managérial, le fait que nos résultats soulignent la tendance des recruteurs à être influencés par des recommandations non pertinentes fournies par un ADSS appelle à une extrême vigilance dans le choix de ces outils. Suivre aveuglément des recommandations fournies par des outils manquant de transparence et d'explicabilité peut générer des risques juridiques et de réputation en cas de discrimination basée sur les algorithmes pour les entreprises utilisant ces systèmes : des explications insuffisantes sur les décisions basées sur les

algorithmes peuvent nuire à la réputation d'une entreprise en cas de discrimination, même si des travaux récents suggèrent que la discrimination algorithmique provoque moins d'indignation morale que la discrimination humaine (Bigman et al., 2020). Dans ce contexte, Friedman & Nissenbaum (1996) ont distingué 3 types de biais liés à l'utilisation des ADSS : les biais préexistants, qui affectent les données sur lesquelles les outils se basent pour faire des prédictions ; les biais techniques ou algorithmiques causés par le fonctionnement du système ; les biais émergents causés par l'utilisateur du système (biais liés au comportement du recruteur placé devant un choix final). Selon Barocas & Selbst (2016), les données étant fréquemment imparfaites, les algorithmes reproduisent les préjugés issues des décisions antérieurement prises par les individus. Concernant les biais techniques causés par le fonctionnement du système, des résultats récents ont mis en évidence par exemple un biais dans l'analyse faciale automatisée (Buolamwini & Gebru, 2018). Les auteurs ont constaté que tous les systèmes automatisés obtenaient les meilleures performances pour les individus à la peau plus claire et pour les hommes en général. Les entreprises qui commercialisent les outils de recrutement prédictifs affirment à juste titre que la décision finale appartient toujours à un humain. Cependant, le choix des candidats présélectionnés peut être influencé par la manière dont les ADSS présentent la recommandation : la recommandation est généralement donnée sous forme de classements présentés sans marge d'erreur. Les travaux précurseurs de Tversky & Kahneman (1974) ont popularisé une série de biais décisionnels liés à la présentation des solutions, tels que le biais d'ancrage ou l'effet de cadrage, qui influencent fortement les décisions, surtout lorsque celles-ci comportent un risque (Kühberger, 1998). Par exemple, lorsque les résultats apparaissent sous la forme d'un classement ou d'une sélection, le risque de biais de présentation est élevé (Craswell et al., 2008) : le candidat classé en première position reçoit un avantage, même si la différence de score avec les autres candidats est faible. Par conséquent, les effets de l'interface ADSS sur le choix des recruteurs est également une question qui mérite d'être approfondie.

References

- Ajzen, I. (1985). *From intentions to actions : A theory of planned behavior*. Springer.
http://link.springer.com/chapter/10.1007/978-3-642-69746-3_2
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *Calif. L. Rev.*, 104, 671.
- Besse, P., Castets-Renard, C., Garivier, A., & Loubes, J.-M. (2018). *Can Everyday AI be Ethical? Machine Learning Algorithm Fairness* (SSRN Scholarly Paper ID 3391288). Social Science Research Network. <https://papers.ssrn.com/abstract=3391288>
- Bigman, Y., Gray, K., Waytz, A., Arnestad, M., & Wilson, D. (2020). *Algorithmic Discrimination Causes Less Moral Outrage than Human Discrimination*. PsyArXiv.
<https://doi.org/10.31234/osf.io/m3nrp>
- Buolamwini, J., & Gebru, T. (2018). Gender shades : Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 77-91.
- Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2), 220-239.
<https://doi.org/10.1002/bdm.2155>
- Campion, M. C., Campion, M. A., Campion, E. D., & Reider, M. H. (2016). Initial investigation into computer scoring of candidate essays for personnel selection. *Journal of Applied Psychology*, 101(7), 958-975. <https://doi.org/10.1037/apl0000108>

- Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, 56(5), 809-825. <https://doi.org/10.1177/0022243719851788>
- Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New Talent Signals : Shiny New Objects or a Brave New World? *Industrial and Organizational Psychology*, 9(3), 621-640. <https://doi.org/10.1017/iop.2016.6>
- Craswell, N., Zoeter, O., Taylor, M., & Ramsey, B. (2008). An experimental comparison of click position-bias models. *Proceedings of the 2008 international conference on web search and data mining*, 87-94.
- de Jong, S. (2020). *Algorithmic appreciation or aversion : Does the representation of an algorithm change the trust placed in it?* [MsC Thesis]. Leiden.
- Derous, E., & Ryan, A. M. (2019). When your resume is (not) turning you down : Modelling ethnic bias in resume screening. *Human Resource Management Journal*, 29(2), 113-130. <https://doi.org/10.1111/1748-8583.12217>
- Diab, D. L., Pui, S.-Y., Yankelevich, M., & Highhouse, S. (2011). Lay perceptions of selection decision aids in US and non-US samples. *International Journal of Selection and Assessment*, 19(2), 209-216.
- Dietvorst, B. J. (2016). *Algorithm aversion*. Publicly Accessible Penn Dissertations. Paper 1686. <http://repository.upenn.edu/edissertations/1686>
- Dietvorst, B. J., & Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error STUDY MATERIAL. *Psychological science*, 31(10), 1302-1314.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion : People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion : People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3), 1155-1170.
- Dikmen, M., & Burns, C. (2022). The Effects of Domain Knowledge on Trust in Explainable AI and Task Performance: A Case of Peer-to-Peer Lending. *International Journal of Human-Computer Studies*, 102792.
- Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58(6), 697-718. [https://doi.org/10.1016/S1071-5819\(03\)00038-7](https://doi.org/10.1016/S1071-5819(03)00038-7)
- Efendić, E., Van de Calseyde, P. P., & Evans, A. M. (2020). Slow response times undermine trust in algorithmic (but not human) predictions. *Organizational behavior and human decision processes*, 157, 103-114.
- Filiz, I., Judek, J. R., Lorenz, M., & Spiwojs, M. (2021). The Tragedy of Algorithm Aversion. *Wolfsburg Working Papers*, 21(02), 1-22.

- Freitag, M., & Bauer, Paul. C. (2016). Personality traits and the propensity to trust friends and strangers. *The Social Science Journal*, 53(4), 467-476. <https://doi.org/10.1016/j.soscij.2015.12.002>
- Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, 14(3), 330. <https://doi.org/10.1145/230538.230561>
- Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence : Review of Empirical Research. *Academy of Management Annals*, 14(2), 627-660. <https://doi.org/10.5465/annals.2018.0057>
- Gonzalez, M., F., Capman, J., F., Oswald, F. L., They, E., R., & Tomczak, D., L. (2019). “Where’s the I-O?” Artificial Intelligence and Machine Learning in Talent Management Systems. *Personnel Assessment and Decisions*, 5(3), 33-44.
- Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in Human Behavior*, 119, 106730. <https://doi.org/10.1016/j.chb.2021.106730>
- Hoff, K. A., & Bashir, M. (2015). Trust in Automation : Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors*, 57(3), 407-434. <https://doi.org/10.1177/0018720814547570>
- International Data Corporation. (2020). Worldwide Spending on Artificial Intelligence Is Expected to Double in Four Years, Reaching \$110 Billion in 2024, According to New IDC Spending Guide. Retrieved from <https://www.idc.com/getdoc.jsp?containerId=prUS46794720>
- Jessup, S. A., Schneider, T. R., Alarcon, G. M., Ryan, T. J., & Capiola, A. (2019). The measurement of the propensity to trust automation. *International Conference on Human-Computer Interaction*, 476-489.
- Kroenung, J., & Eckhardt, A. (2015). The attitude cube—A three-dimensional model of situational factors in IS adoption and their impact on the attitude–behavior relationship. *Information & Management*, 52(6), 611-627. <https://doi.org/10.1016/j.im.2015.05.002>
- Kühberger, A. (1998). The Influence of Framing on Risky Decisions : A Meta-analysis. *Organizational Behavior and Human Decision Processes*, 75(1), 23-55. <https://doi.org/10.1006/obhd.1998.2781>
- Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions : A meta-analysis. - PsycNET. *Journal of Applied Psychology*, 98(6), 1060-1072. <https://psycnet.apa.org/doi/10.1037/a0034156>
- Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews : Acceptance under the influence of stakes. *International Journal of Selection and Assessment*, 27(3), 217-234. <https://doi.org/10.1111/ijsa.12246>
- Langer, M., König, C. J., Sanchez, D. R.-P., & Samadi, S. (2019). Highly automated interviews : Applicant reactions and the organizational context. *Journal of Managerial Psychology*.
- Langer, M., König, C., Back, C., & Hemsing, V. (2021). *Trust in Artificial Intelligence : Comparing trust processes between human and automated trustees in light of unfair bias*. <https://doi.org/10.31234/osf.io/r9y3t>
- Lee, J. D., & See, K. A. (2004). Trust in Automation : Designing for Appropriate Reliance. *Human Factors*, 31.

- Lee, M. K. (2018). Understanding perception of algorithmic decisions : Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 1-16.
<https://doi.org/10.1177/2053951718756684>
- LinkedIn (2018). Global Recruiting Trends 2018. The 4 ideas changing how you hire. Retrieved from <https://business.linkedin.com/content/dam/me/business/en-us/talent-solutions/resources/pdfs/linkedin-global-recruiting-trends-2018-en-us2.pdf>
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation : People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90-103.
<https://doi.org/10.1016/j.obhdp.2018.12.005>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model Of Organizational Trust. *Academy of Management Review*, 20(3), 709-734. <https://doi.org/10.5465/amr.1995.9508080335>
- McBride, M., Carter, L., & Ntuen, C. (2012). The impact of personality on nurses' bias towards automated decision aid acceptance. *International Journal of Information Systems and Change Management*, 6(2), 132-146. <https://doi.org/10.1504/IJISCM.2012.051148>
- McCrae, R. R., & Costa, P. T. (2003). *Personality in Adulthood : A Five-factor Theory Perspective*. Guilford Press.
- Merritt, S. M., & Ilgen, D. R. (2008). Not All Trust Is Created Equal : Dispositional and History-Based Trust in Human-Automation Interactions. *Human Factors*, 50(2), 194-210.
<https://doi.org/10.1518/001872008X288574>
- Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias : Decision making and performance in high-tech cockpits. *The International journal of aviation psychology*, 8(1), 47-63.
- Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39(3), 429-460.
- Oberst, U., De Quintana, M., Del Cerro, S., & Chamarro, A. (2020). Recruiters prefer expert recommendations over digital hiring algorithm : A choice-based conjoint study in a pre-employment screening scenario. *Management Research Review*, 44(4).
- Ochmann, J., Zilker, S., & Laumer, S. (2021). The Evaluation of the Black Box Problem for AI-Based Recommendations : An Interview-Based Study. In F. Ahlemann, R. Schütte, & S. Stieglitz (Éds.), *Innovation Through Information Systems* (p. 232-246). Springer International Publishing.
https://doi.org/10.1007/978-3-030-86797-3_16
- Office for AI. (2019). Understanding artificial intelligence. Retrieved December 1, 2021, retrieved from <https://www.gov.uk/government/publications/understanding-artificial-intelligence>
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation : An attentional integration. *Human factors*, 52(3), 381-410.
- Parasuraman, R., & Riley, V. (1997). Humans and automation : Use, misuse, disuse, abuse. *Human factors*, 39(2), 230-253.
- Sanchez, J., Rogers, W. A., Fisk, A. D., & Rovira, E. (2014). Understanding reliance on automation : Effects of error type, error distribution, age and experience. *Theoretical issues in ergonomics science*, 15(2), 134-160. <https://doi.org/10.1080/1463922X.2011.611269>

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced "complacency" : Development of the complacency-potential rating scale. *The International Journal of Aviation Psychology*, 3(2), 111-122.

Szalma, J. L., & Taylor, G. S. (2011). Individual differences in response to automation : The five factor model of personality. *Journal of Experimental Psychology: Applied*, 17(2), 71-96.
<https://doi.org/10.1037/a0024170>

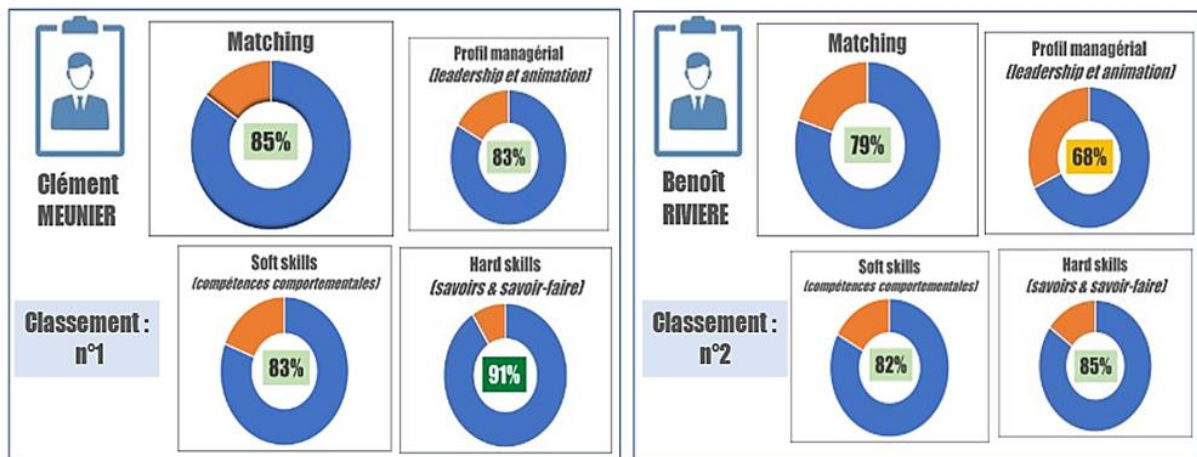
Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial Intelligence in Human Resources Management : Challenges and a Path Forward. *California Management Review*, 61(4), 15-42.
<https://doi.org/10.1177/0008125619867910>

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty : Heuristics and Biases. *Science*, 185(4157), 1124-1131. <https://doi.org/10.1126/science.185.4157.112>

Annexe 1 : Extrait des recommandations présentées aux recruteurs

Condition Recommandation algorithmique cohérente

L'entreprise ayant diffusé l'offre et ayant reçu les deux candidatures a acquis la solution de recrutement prédictif proposée par *HR Predict*, startup spécialisée dans le développement de packages intégrés d'aide au recrutement. Cette solution est basée sur l'utilisation d'algorithmes avancés intégrant les derniers développements en matière d'intelligence artificielle (apprentissage supervisé adaptatif, réseaux de neurones). En se basant sur les informations disponibles sur les candidats précédemment recrutés, ainsi que sur l'extraction de données du dossier de candidature, *HR Predict* propose de classer les candidats selon leur taux d'adéquation (*matching*) à l'offre : la solution propose un score d'adéquation globale et un classement, avec trois sous-critères : les compétences relationnelles, les compétences techniques et l'adéquation du profil managérial avec les besoins du poste.



Condition recommandation expert humain cohérente

Extraits de la conversation que vous avez eue avec M.D, dirigeant du cabinet de recrutement MPI, auquel l'entreprise a fait appel plusieurs fois pour des recrutements de cadres. Vous l'avez consulté par téléphone après lui avoir envoyé les deux profils complets des candidats ainsi que l'offre d'emploi publiée.

« J'ai regardé rapidement vos deux profils.... Bon...et bien à mon avis il n'y a pas photo. Clément Meunier a tout bon. Il a un super parcours et il colle au profil. En plus, il est bilingue...je ne vois pas vraiment de faiblesses dans cette candidature.

Benoît Rivière est bien aussi : je trouve qu'il remplit globalement les critères et il a un parcours intéressant... L'anglais n'est pas un problème : il pourra s'adapter.

Mais j'ai deux petits doutes : il a managé une toute petite équipe de 7 personnes et puis il est plutôt dans l'opérationnel... ça reste quand même un bon profil.