

Titre : Promouvoir l'adoption de l'IA dans les milieux d'emploi par l'entremise de l'explicabilité et de la confiance : une étude empirique

Rubrique : Article de recherche

Auteur(s)

1 : Viviane Masciotra, étudiante

2 : Jean-Sébastien Boudrias, professeur

Citation : Masciotra, V. et Boudrias, J.-S. (2024). Promouvoir l'adoption de l'IA dans les milieux d'emploi par l'entremise de l'explicabilité et de la confiance : une étude empirique. *Ad Machina*, 8(1), 84-113, <https://doi.org/10.1522/radm.no8.1840>

Affiliation des auteurs

1 : Université de Montréal

Courriel : viviane.masciotra@umontreal.ca

2 : Université de Montréal

Courriel : jean-sebastien.boudrias@umontreal.ca

Remerciements

Déclaration des conflits d'intérêts

- Aucun conflit d'intérêts à déclarer
 Conflit d'intérêts à déclarer (veuillez détailler)

Détails :

Résumé (250 mots)

L'intelligence artificielle (IA) est associée à plusieurs bénéfices pour les travailleurs et les organisations. Toutefois, ses capacités inédites sont propices à engendrer chez les humains de la crainte pour la pérennité de leur emploi, et de la réticence à utiliser l'IA. Dans la présente étude, nous explorons le rôle de la confiance envers l'utilisation de l'IA chez les travailleurs, ainsi que la capacité de l'explicabilité de l'algorithme à promouvoir la confiance. À cet effet, un devis expérimental à répartition aléatoire a été utilisé. Les résultats révèlent que la confiance favorise l'intention d'utiliser l'IA, mais que l'explicabilité ne contribue pas au développement de la confiance. De plus, l'explicabilité a eu un effet inattendu délétère sur l'intention d'utiliser l'IA.

Abstract

Artificial intelligence (AI) is associated with numerous benefits for workers and organizations. However, its novel capabilities are likely to generate fears for the sustainability of their jobs and reluctance to use AI among humans. In this study, the role of trust is studied in the use of AI among workers, as well as the ability of the explainability of the algorithm to promote trust. To achieve this, a randomized experimental design was used. The results reveal that trust promotes the intention to use AI, but that explainability does not contribute to the development of trust. In addition, explainability had an unexpectedly deleterious effect on the intention to use AI.

Mots clés

Intelligence artificielle, confiance, explicabilité, travail, intention d'utilisation

Droits d'auteur

Ce document est en libre accès, ce qui signifie que le lectorat a accès gratuitement à son contenu. Toutefois, cette œuvre est mise à disposition selon les termes de la licence [Creative Commons Attribution \(CC BY NC\)](https://creativecommons.org/licenses/by-nc/4.0/).



Promouvoir l'adoption de l'IA dans les milieux d'emploi par l'entremise de l'explicabilité et de la confiance : une étude empirique

Viviane Masciotra
Jean-Sébastien Boudrias

Introduction

Le terme intelligence artificielle (IA) décrit un système informatique capable de simuler certaines fonctions cognitives de l'intelligence humaine telles que l'apprentissage, le raisonnement et la planification (Pereira et al., 2021). Ainsi, l'IA peut émettre des prédictions, des recommandations et des décisions sur la base de données d'entrée, lesquelles peuvent provenir autant de l'humain que de la machine elle-même (Lane et al., 2023). En effet, ses capacités inédites d'apprentissage profond (*deep learning*) lui permettent d'analyser de vastes bases de données, d'identifier des patrons spécifiques au sein de celle-ci, de prendre des décisions et d'apprendre de ses expériences de manière autonome (Chowdury et al., 2023). Il va sans dire que l'IA a le potentiel de transformer les façons de faire habituelles de l'humain, y compris la manière dont celui-ci réalise ses tâches professionnelles. Par exemple, des algorithmes d'IA sont maintenant capables de paramétrer en temps réel de multiples données comportementales et physiologiques de patients par l'entremise de leur téléphone intelligent. L'analyse de ces données par l'IA offre aux praticiens une vision exhaustive du contexte entourant la condition de santé de chaque individu ainsi que des prédictions détaillées quant à son évolution (Rajpurkar et al., 2022). À l'instar de la médecine personnalisée, des applications de l'IA dans le domaine légal permettent aux juges et avocats d'adapter les sanctions et de diriger les criminels vers des programmes de réhabilitation appropriés aux besoins et caractéristiques propres à chaque individu afin de diminuer les probabilités de récidive (Custers, 2022).

Nonobstant son milieu d'application, la venue de l'IA est associée à plusieurs bénéfices, tant pour les travailleurs que pour les organisations. Du côté des travailleurs, l'IA prendrait en charge certaines tâches plus routinières, permettant ainsi à l'individu de consacrer davantage de temps et d'énergie à des tâches plus stimulantes, créatives, stratégiques et, somme toute, plus motivantes (Agrawal et al., 2019; Hmoud et Laszlo, 2019; Ore et Sposato, 2021; Pettersen, 2019; Siau et Wang, 2018; Xiang et al., 2020). Par exemple, dans le domaine de la gestion des ressources humaines, et plus précisément en contexte de sélection du personnel, l'IA prendrait en charge la rédaction de description des postes, l'élaboration de contenu de tests et l'analyse vidéo de candidats à des fins de présélection (Black et van Esch, 2020; Lee et al., 2023). Ceci permettrait aux professionnels des ressources humaines de dédier plus de temps à l'aspect relationnel et consultatif de leur rôle (Ramachandran et al., 2022).

Du côté des entreprises, l'IA permettrait non seulement de prendre en charge certaines tâches couramment effectuées par des humains, mais aussi de les réaliser de manière exponentiellement plus efficace (Giermindl et al., 2022). Par exemple, l'utilisation de technologies d'IA en sélection de personnel ferait en sorte de diminuer les délais de recrutement de 63 %, procurant aux entreprises un avantage compétitif considérable en répondant plus rapidement aux candidats les plus attrayants (Black et van Esch, 2020). De plus, en fonctionnant 24 heures sur 24 et en réalisant de nombreuses tâches simultanément, l'IA donnerait toute latitude aux organisations d'offrir un service en continu à leurs clients, tout en diminuant les erreurs liées à la fatigue ou à l'inattention pouvant être commises par l'humain (Kim et Heo, 2022; Wright et Schultz, 2018).

Bien que les bénéfices associés à l'IA soient attrayants, certains obstacles peuvent empêcher les travailleurs et les organisations d'en profiter. Un de ces obstacles réside dans la réticence à utiliser l'IA (Burton et al., 2019; Mahmud et al., 2022; Venkatesh, 2022). À cet effet, une étude menée auprès de 10 multinationales révèle que pour des investissements annuels de plusieurs millions de dollars, près de 80 % des initiatives d'implantation de l'IA au sein de ces entreprises étaient stagnantes au moment de l'enquête en raison de réticences d'utilisation de la technologie par les travailleurs (Dimensional Research, 2019). Une autre enquête, celle-ci menée auprès de 20 000 employés à travers le monde, dénote une préférence marquée pour les technologies de l'information traditionnelles ainsi que l'existence de réticences envers l'utilisation de l'IA (Russel et Susskind, 2021). En Amérique du Nord, plus de 60 % des répondantes et des répondants à cette enquête feraient plus confiance aux entreprises dont les processus internes sont réalisés à l'aide de technologies de l'information traditionnelles qu'aux entreprises qui utilisent l'IA.

À l'origine de cette réticence réside le sentiment d'incertitude et de prise de risque ressenti par les travailleurs appelés à collaborer avec l'IA, lequel serait associé à la nature opaque de la technologie (Hughes et al., 2019). En effet, une des caractéristiques de l'IA les plus marquantes, et qui contribue à la résistance des travailleurs à l'utiliser, est ce que l'on appelle le phénomène de « boîte noire ». Plus précisément, on caractérise de boîte noire le fonctionnement de l'IA parce que ses réponses sont produites par l'entremise de processus difficiles, voire impossibles, à comprendre même pour les experts (Barredo Arrieta et al., 2020). L'opacité engendrée par le manque d'explication adéquate des processus sous-jacents entrave ainsi l'acceptation de l'IA auprès des travailleurs et explique en partie les réticences exprimées (Gunning et al., 2019; Hasija et Esper, 2022; Mirbabaie et al., 2022).

Devant ce dilemme, il importe de mieux comprendre les facteurs pouvant mitiger le sentiment d'incertitude et de prise de risque associé à l'opacité de l'IA afin d'en promouvoir l'utilisation auprès des employés (Lichtenthaler, 2020). À cet effet, la littérature scientifique indique deux pistes de recherche prometteuses, quoiqu'encore peu explorées de manière empirique : la confiance des travailleurs envers l'IA et l'explicabilité de l'algorithme (Glickson et Woolley, 2020; Gulati et al., 2019; Siau et Wang, 2018; Toreini et al., 2019). Ainsi, l'objectif de la présente étude est de démontrer de manière empirique le lien entre la confiance des travailleurs et l'intention d'utiliser celle-ci, ainsi que la capacité de l'explicabilité de manière à favoriser la confiance envers l'IA.

1. Le rôle de la confiance

La recherche portant sur la confiance envers l'IA en étant à ses débuts, la plupart des écrits à ce sujet sont ancrés dans les prémisses plus largement étudiées de la confiance envers les technologies de l'information.

1.1 La confiance envers les technologies de l'information

Moray et Inagaki (1999) décrivent la confiance envers les technologies de l'information (la technologie) comme la croyance qu'une technologie accomplisse comme prévu, et de manière fiable, les tâches pour lesquelles elle a été conçue. Selon Madsen et Gregor (2000), la confiance envers la technologie correspond au degré pour lequel celle-ci incite son utilisateur à agir en fonction de ses recommandations, de ses actions et de ses décisions.

La confiance envers la technologie est depuis longtemps reconnue comme étant un moyen efficace d'en promouvoir son utilisation auprès des travailleurs. Gefen et al. (2003) ont joué un rôle de premier plan dans ce domaine, en bonifiant le modèle des prédicteurs d'utilisation des nouvelles technologies *Technology Acceptance Model* (TAM) de Davis (1989) par l'ajout de la confiance cognitive conceptualisée par McKnight (1998, 2002). Depuis, la recherche soutient largement le rôle prédictif de la confiance envers la technologie dans l'intention de l'utiliser (Gefen et al., 2003; Benbasat et Wang, 2005).



Diverses pistes ont été explorées afin d'expliquer ce lien. Tout d'abord, un niveau de confiance initial s'avère nécessaire afin de passer outre la perception de risque et d'incertitude liée à la présentation d'une nouvelle technologie (McKnight et al., 2002; Koufaris et Hampton-Sosa, 2004). En effet, la perception qu'une technologie agira dans le meilleur intérêt de son utilisateur aide à réduire l'anxiété associée aux relations technologies-humains (Hoff et Bashir, 2015; Söllner et al., 2016) et contribue à créer une relation positive et empreinte de sens (McKnight et al., 2011; Mirbabaie et al., 2022). Il a aussi été avancé que la confiance initiale des travailleurs envers la technologie contribuerait à accroître la satisfaction au travail (Tarafdar et al., 2010; Zielonka, 2022). Cette perception initiale serait ensuite modulable au fil du temps, selon l'expérience de l'individu avec la technologie (Söllner et al., 2012).

De nos jours, la majorité des individus ont eu l'occasion d'interagir avec de multiples technologies de l'information et de développer envers elles un certain niveau de confiance (Hengstler et al., 2016). Ainsi, les intérêts de recherche se tournent désormais vers la confiance envers l'IA.

1.2 La confiance envers l'IA

Considérant son implantation récente au sein d'organisations variées, il n'est pas surprenant que le sujet de la confiance des travailleurs envers l'IA génère beaucoup d'écrits. De manière similaire à la confiance envers la technologie, la confiance envers l'IA se traduit par l'impression que son fonctionnement lui sera favorable et sans préjudice, c'est-à-dire qu'elle lui permettra d'atteindre un but désiré, et ce, en dépit du risque couru (Kaplan et al., 2021; Ryan, 2020). Pour les besoins de cette étude, nous avons adopté la définition de la confiance envers la technologie formulée par Madsen et Gregor (2000), qui a été adaptée au contexte de l'IA par Ashoori et Weisz (2019). Elle se rapporte au niveau de certitude ressenti par un individu et à sa disposition à suivre les recommandations d'un algorithme d'IA dans le but d'atteindre un objectif spécifique.

1.3 Le rôle de la confiance dans l'utilisation de l'IA

La confiance des travailleurs envers l'IA constitue une condition de succès critique à son utilisation (Yang et Wibowo, 2022). Ainsi, le modèle de Gefen et al. (2003), qui positionne la confiance envers une technologie de l'information comme prédicteur de l'intention de l'utiliser, serait applicable au contexte de l'IA. Un nombre grandissant d'études soutiennent empiriquement cette application, et démontrent que plus un individu fait confiance à l'IA, plus il a l'intention d'utiliser cette technologie et ce, autant dans un contexte d'utilisation personnelle (Choung et al., 2021; Waung et al., 2021) que professionnelle (Liu et al., 2022). Plus précisément, la confiance permettrait d'assouplir la perception d'incertitude et de risque associée à l'utilisation de l'IA et à son opacité, et faciliterait son adoption par les individus (Chaudhry et al., 2022).

Bien que la recherche tende à progresser à ce sujet, elle en est encore à ses débuts, particulièrement dans la démonstration empirique du lien entre la confiance et l'intention d'utiliser l'IA en contexte de travail. La présente étude comble ce manque par l'entremise d'une vérification empirique de ce lien dans le cadre de la réalisation d'une tâche professionnelle dans le domaine des ressources humaines. Ainsi, nous proposons que :

Hypothèse 1 : La confiance envers l'IA entretient un lien significatif et positif avec l'intention d'utiliser cette technologie

2. L'explicabilité

Considérant le rôle clé de la confiance envers l'IA lorsqu'il s'agit d'en favoriser l'utilisation, il importe de se pencher sur les facteurs qui servent de base à son développement. Un facteur se démarque particulièrement dans la littérature, soit l'explicabilité de l'algorithme d'IA (Ashoori et Weisz, 2019; Liu et al., 2022; Gunning et al., 2019; Shin, 2021).

Dans le domaine de l'IA, le terme explicabilité est utilisé de manière interchangeable avec transparence, intelligibilité, compréhensibilité et interprétabilité du système (Guidotti et al., 2018). Certains auteurs décrivent l'explicabilité comme étant une caractéristique de l'algorithme menant à la transparence de l'IA. Pour les besoins de la présente étude, le terme explicabilité est préconisé, faisant référence au degré auquel les critères utilisés par un algorithme d'IA afin d'en venir à un résultat donné sont traduits de manière compréhensible pour son utilisateur (Arrieta et al., 2020; Wadden, 2021).

Alors que les technologies de l'information traditionnelles fonctionnent à l'aide de logiciels préprogrammés par des humains, l'IA, et en particulier l'apprentissage profond, fonctionne de manière autonome. Ceci se fait au moyen d'algorithmes capables d'autoapprentissage et d'autocorrection, selon les expériences et la rétroaction reçue (Glickson et Woolley, 2020). Dans ce contexte, l'apport de l'humain est limité et se résume à définir les données d'entrée et le type de sortie attendue du système (Castelvecchi, 2016). Puisqu'il est déterminé par la machine et largement incompréhensible par l'humain, le fonctionnement de l'IA est souvent associé à celui d'une « boîte noire », qui reflète son opacité (Gunning et al., 2019). Ce phénomène est par ailleurs inhérent à l'avancement de l'IA, c'est-à-dire que plus les algorithmes sont performants, sophistiqués et intelligents (en d'autres mots, autonomes), moins leur fonctionnement est compréhensible pour l'humain (Rossi, 2018).

2.1 Le rôle de l'explicabilité dans la confiance vers l'IA

L'opacité de l'IA est problématique dans la mesure où elle entrave le travail d'évaluation que doit réaliser l'humain afin de déterminer s'il peut, ou non, faire confiance à la machine. On retrouve un lot de recherches empiriques focalisées sur l'aspect technique de l'explicabilité, qui s'inscrit au niveau du développement des algorithmes d'IA (Cabrera-Sánchez et al., 2021). Dans le domaine de la psychologie du travail et des organisations, c'est plutôt l'« explicabilité située » qui est étudiée, c'est-à-dire une explication qui situe l'IA dans son contexte d'usage et qui illustre comment l'IA fonctionne et en arrive à ses résultats de manière compréhensible pour les travailleurs non-experts dans le domaine (Borel, 2024).

En théorie, l'explicabilité permettrait d'éclaircir la boîte noire de l'IA afin d'aider son utilisateur à mieux évaluer le risque couru en lui déléguant la réalisation d'une tâche donnée (von Eschenbach, 2021). Plus précisément, il a été démontré que la présence d'une explication portant sur le « comment » une IA en vient à prendre ses décisions rend son fonctionnement plus intelligible et moins risqué (Shin, 2021). De plus, il a été avancé que l'explication inciterait l'utilisateur de l'IA à évaluer positivement la compétence (Miller, 2019), ainsi que la fiabilité (Chowdury et al., 2023; Ferrario et Loi, 2022) de l'algorithme, deux sous-dimensions préconisées dans la conceptualisation de la confiance envers l'IA (McKnight et al., 2011).

2.2 L'explicabilité en contexte organisationnel (titre proposé)

Il est important de souligner qu'en contexte organisationnel, et plus spécifiquement auprès d'une population inexercée au domaine de l'IA, certaines explications s'avèrent plus optimales que d'autres (Tsiakas et Murray-Rust, 2022). Notamment, il a été théorisé qu'une explication trop transparente ou complexe risque de mener à une perception négative de l'IA (Confalonieri et al., 2020; Langer et König, 2021). En effet, la présentation d'éléments peu familiers, voire inconnus, concernant une nouvelle



technologie peut engendrer incertitude et crainte auprès de ses utilisateurs (Westphal et al., 2023). Ces dernières années, de plus en plus de chercheurs ont étudié de manière empirique le type d'explication s'avérant optimale afin de promouvoir la confiance chez différents utilisateurs. Bien que la réponse la plus sensible à cette question soit de personnaliser l'explication à chaque individu (Shulner-Tal et al., 2022), une méthode plus pragmatique de clarification du fonctionnement de l'IA à une population générale est l'explication simple permettant de transmettre des informations factuelles sur la manière dont une technologie a été conçue et entraînée (Ashoori et Weisz, 2019).

À l'heure actuelle, le rôle de l'explicabilité dans la promotion de la confiance envers l'IA n'a été que très peu étudié à la fois de manière empirique et dans un contexte organisationnel. La présente étude comble ce manque par l'entremise d'une manipulation expérimentale de l'explicabilité de l'IA dans le cadre de la réalisation d'une tâche par des professionnels en ressources humaines. Ainsi, nous proposons que :

Hypothèse 2 : Les travailleurs à qui on présente une IA dont les mécanismes sont clarifiés par une explication factuelle manifestent un niveau de confiance envers l'IA significativement plus élevé que les travailleurs à qui on présente une IA dont le fonctionnement est opaque.

3. Méthodologie

Afin de vérifier les hypothèses de recherche, un devis expérimental à répartition aléatoire (IA expliquée versus non expliquée) a été utilisé. Une méthodologie de vignettes a servi à mettre les participants recrutés en situation d'utilisation de l'IA dans leur domaine professionnel. Par la suite, ceux-ci ont répondu à un questionnaire élaboré par l'équipe de recherche portant sur la confiance envers l'IA et sur leur intention d'utiliser l'IA en contexte de travail.

3.1 Échantillon

Collecte de données

Une taille d'échantillon de 120 participants était visée aux fins de la présente étude afin d'effectuer une collecte de données réalisable dans une période circonscrite, tout en assurant une puissance statistique adéquate pour détecter les effets étudiés (Lakens, 2022). Le recrutement des participants a été réalisé en partenariat avec une firme-conseil en évaluation et en développement des ressources humaines. Un de ses représentants a transmis par courriel à 824 employés une invitation à participer à l'étude. Les employés intéressés ont été invités à contacter l'équipe de recherche, et ont ensuite reçu un hyperlien leur permettant d'accéder au questionnaire élaboré par l'équipe de recherche sur la plateforme informatique Qualtrics. Afin d'accéder au questionnaire, les participants devaient consentir aux modalités détaillées dans un formulaire de consentement, celui-ci ayant reçu l'approbation éthique de l'Université de Montréal. Les employés ne désirant pas consentir aux modalités décrites dans ce formulaire ont été invités à quitter la plateforme.

Confidentialité

Les données recueillies ont été anonymisées par la plateforme Qualtrics, celle-ci ayant remplacé l'identité des répondants par un identifiant alphanumérique. Les données recueillies par l'équipe de recherche ont ainsi été reçues de manière entièrement anonyme.

Participants

Les critères de sélection étaient d'avoir plus de 18 ans, une bonne compréhension de l'anglais et d'être employé depuis au moins six mois dans l'organisation afin d'assurer une compréhension minimale de son rôle et de ses tâches liées à l'évaluation des ressources humaines. Le recrutement au sein de la firme-conseil

a permis de circonscrire les deux premiers critères de sélection, car l'entreprise n'embauche pas en deçà de 18 ans et la maîtrise de l'anglais est une condition d'embauche. En ce qui a trait à l'ancienneté, une question a été posée aux participants au début du questionnaire et ceux-ci n'étaient en mesure d'accéder aux items que s'ils répondaient avoir une ancienneté supérieure ou égale à six mois.

Les 137 employés ayant démontré leur intérêt à participer à l'étude ont été assignés au hasard au groupe expérimental ou au groupe contrôle et un hyperlien vers la version appropriée du questionnaire leur a été envoyée. Au total, 126 travailleurs ont répondu au questionnaire à l'intérieur du délai communiqué de trois semaines et de deux rappels envoyés par courriel. Trois participants ont été exclus à défaut d'avoir rempli au moins 50 % du questionnaire. Deux autres ont été retirés parce qu'ils ont lu le scénario en moins de 30 secondes, ce qui laisse présager qu'ils n'ont pas eu le temps d'incorporer adéquatement l'information présentée. Le recrutement a pris fin lorsque les données de 120 participants ont été recueillies et de ceux-ci, 60 étaient assignés au groupe contrôle et 60 au groupe expérimental. Une analyse du chi carré indique qu'il n'existe pas de différence significative entre le groupe contrôle et expérimental en ce qui concerne leurs âge, genre et ancienneté. Les données démographiques concernant les répondants sont présentées au tableau 1.

Tableau 1
Caractéristiques démographiques des répondants et des répondantes

Caractéristiques	Groupe expérimental		Groupe contrôle		Total		χ^2	<i>p</i>
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%		
Genre							2,07	0,36
Femme	32	53,30	31	51,70	63	52,50		
Homme	28	46,70	29	48,30	57	47,50		
Âge							2,74	0,43
Entre 18 et 25 ans	24	40,00	28	46,70	52	43,40		
Entre 26 et 35 ans	28	46,70	29	48,30	57	47,50		
Entre 36 et 45 ans	6	10,00	2	3,30	8	6,70		
Entre 46 et 55 ans	2	3,30	1	1,70	3	2,50		
Ancienneté							2,10	0,55
Entre 6 mois et 1 an	9	15,00	4	6,70	13	10,80		
Entre 1 et 5 ans	43	71,70	45	75,00	88	73,30		
Entre 6 et 10 ans	6	10,00	8	13,30	14	11,70		
Plus de 10 ans	2	3,30	3	5,00	5	4,20		

Note. *N* = 120 (*n* = 60 pour chacun des groupes)

3.2 Procédure

La méthodologie de vignettes expérimentales utilisée implique de présenter un scénario hypothétique, quoique réaliste, aux participants et de recueillir leurs réponses en fonction de cette mise en situation. Cette méthodologie permet la manipulation de variables expérimentales ainsi qu'une vérification empirique adéquate dans l'étude de thèmes de recherche émergents, offrant une base pour le développement et le raffinement de théories plus approfondies (Atzmüller et Steiner, 2010; Lapatin et al., 2012; Robinson et Clore, 2001). Ceci est particulièrement indiqué dans l'étude de l'arrivée de l'IA dans les milieux de travail (Höddinghaus et al., 2020), cette dernière étant grandement attendue, mais à ce jour encore peu répandue, au sein des organisations.



Deux instruments de mesure identiques ont donc été développés, chacun précédé d'une vignette décrivant une situation de travail où le répondant était amené à utiliser les prédictions émises par un algorithme d'IA afin de rendre une décision d'embauche. Ce contexte a été sélectionné parce qu'il décrit l'utilisation d'un algorithme d'IA que la firme partenaire projette d'utiliser dans un futur rapproché, c'est-à-dire qui correspond à une situation réaliste qui maximise les chances que les participants puissent l'envisager. L'élément permettant de distinguer le groupe expérimental du groupe contrôle est l'ajout à la vignette expérimentale d'une explication simple et factuelle éclairant le lecteur quant au fonctionnement de l'algorithme d'IA, c'est-à-dire du « comment » celui-ci en arrive à son output. Le format de l'explication suit les recommandations existantes dans la littérature scientifique (Ashoori et Weisz, 2019). Les vignettes ayant été présentées aux participants ainsi que l'explication expérimentale sont disponibles à l'annexe A.

3.3 Mesures

Comme pour la plupart des instruments visant à mesurer la confiance envers l'IA (Vereschak et al., 2021), le présent outil a été développé en adaptant au contexte de l'IA des items validés en contexte d'utilisation de technologies de l'information. Tous les items sont disponibles à l'annexe B.

La confiance envers l'IA

Inspirée de la recherche portant sur la confiance envers la technologie, la conceptualisation de la confiance envers l'IA se décline en deux axes, soit la confiance affective et la confiance cognitive (Vereschak et al., 2021). La confiance affective est ancrée dans les émotions ressenties et se base sur des attitudes subjectives, tandis que la confiance cognitive est enracinée dans la pensée rationnelle et se base sur des faits objectifs et des observations passées (Glickson et Woolley, 2020). Tandis que la mesure de la confiance cognitive est adaptée à l'utilisation de l'IA de manière générale, il n'en est pas de même pour la confiance affective. En effet, la littérature scientifique spécifie que cette dernière serait davantage adaptée aux contextes où l'IA prend une forme anthropomorphique, c'est-à-dire visible et tangible tel un robot évoquant l'apparence d'un humain (Glickson et Woolley, 2020; Shi et al., 2021). Toutefois, elle serait moins pertinente dans un contexte où la technologie ne présente pas de ressemblance à l'humain dans sa forme. Dans ce type de scénario, ce serait principalement la confiance cognitive qui jouerait un rôle dans l'émission de confiance (McKnight et al., 2011). Puisque l'algorithme d'IA dépeint dans la présente étude ne présente aucune ressemblance à l'humain, la décision a été prise de circonscrire la recherche à la confiance cognitive.

À cet effet, c'est la conceptualisation de la confiance cognitive envers la technologie de McKnight et al., (2011) qui a été retenue du fait de sa prépondérance dans la littérature scientifique. Considérant la récence des applications pratiques d'IA, l'application de cette conceptualisation au domaine de l'IA détient actuellement un certain support empirique (Choung et al., 2022), quoique celui-ci demeure largement théorique (Vereschak et al., 2021). La confiance envers l'IA comprend les sous-dimensions de la fiabilité (*reliability*), de la fonctionnalité (*functionality*) et du support (*helpfulness*). La fiabilité réfère à la croyance que la technologie fonctionne de manière invariablement correcte en regard des tâches qui lui sont demandées. La fonctionnalité réfère à la croyance que la technologie possède les fonctions, ou caractéristiques, nécessaires afin de réaliser les tâches demandées. Le support réfère à la croyance que la technologie peut fournir une aide adéquate et prompte à ses utilisateurs, notamment au moyen d'une fonction d'aide à même l'interface digital. Puisque l'algorithme d'IA dépeint dans le cadre de la présente étude ne comprend pas de fonction de support, cette sous-dimension a été remplacée par une notion détenant aussi un large fondement empirique dans la littérature portant sur la confiance envers la technologie, soit la compétence (*competence*), qui réfère à la croyance que la technologie possède les capacités, ou l'intelligence, nécessaires afin de réaliser avec succès les tâches demandées (Lu et al., 2019).

Les sous-dimensions fiabilité et fonctionnalité de la confiance cognitive de McKnight et al. (2011) ont été opérationnalisées en instrument de mesure à diverses reprises (voir Gulati et al., 2019; Lankton et al., 2015). Dans le cadre de la présente étude, c'est l'opérationnalisation de Lankton et al. (2015) à trois items par dimension qui a été retenue pour ses qualités psychométriques ($\alpha \geq 0,95$ pour chaque sous-dimension). L'étude de Lankton et al. (2015) révèle la présence de corrélations non négligeables entre ces dimensions. En ce qui a trait à la sous-dimension compétence (Lu et al., 2019), c'est l'opérationnalisation de Shi et al. (2021) à quatre items qui a été retenue ($\alpha = 0,85$ pour la sous-dimension).

L'intention d'utilisation

Afin d'évaluer l'intention des participants d'utiliser l'algorithme d'IA dépeint dans le scénario, trois items ont été adaptés de l'outil de Venkatesh et al. (2003) et du TAM. Ces items ($\alpha = 0,90$ à $0,92$) prédisent de manière significative l'utilisation réelle de la technologie en trois temps de mesure subséquents ($\beta = 0,58$ à $0,61$, $p < 0,001$; Venkatesh et al., [2003]).

Variables de contrôle

La propension à faire confiance à la technologie (propension à faire confiance), l'expérience d'utilisation d'une technologie (expérience d'utilisation), l'âge et le genre sont les variables de contrôle à l'étude. La littérature scientifique démontre que celles-ci entretiennent des liens significatifs avec la confiance envers la technologie ou l'utilisation de la technologie. Premièrement, la propension à faire confiance fait référence à une caractéristique dispositionnelle stable qui se traduit en tendance à assumer d'emblée la fiabilité et la fonctionnalité d'une technologie (Rotter, 1971). Le lien entre la propension à faire confiance et la confiance initiale envers une technologie a été démontré dans le domaine des technologies de l'information (McKnight et al., 2011) et de l'IA (Fan et al., 2020). Trois items ont servi à mesurer la propension à faire confiance ($\alpha = 0,91$, Lankton et al. [2015]). De son côté, l'expérience d'utilisation atténue la perception de risque que représente une nouvelle technologie, augmentant ainsi la confiance envers cette dernière et l'intention de l'utiliser (Gefen, 2000; Gulati, 1995). Démontrés en contexte d'utilisation de technologies de l'information (Hoff et Bashir, 2015), ces liens ont aussi été observés avec l'IA (Cabiddu et al., 2022; Gillath et al., 2021; Liao et Sundar, 2022). Deux items ont servi à mesurer l'expérience d'utilisation ($\alpha = 0,87$, Gefen et al. [2003]).

L'âge influencerait aussi l'intention d'utiliser l'IA, dans la mesure où les adultes plus âgés auraient plus tendance à rejeter l'utilisation de nouvelles technologies (Knowles et Hanson, 2018). Bien que la démonstration empirique de ces effets soit jusqu'à présent partagée (Gillath et al., 2021; Hmoud et Várallyai, 2020), la récence de la recherche portant sur le sujet amène à tout de même inclure cette variable dans la présente étude. De plus, certaines différences pour ce qui est du genre ont été observées à propos des processus cognitifs qui sous-tendent la décision d'utiliser une nouvelle technologie. Notamment, il a été démontré que les femmes perçoivent initialement la technologie comme étant plus facile d'utilisation que les hommes (Zhang et al., 2022). Il a aussi été démontré que les femmes dépendent plus des normes sociales que les hommes lorsqu'il s'agit de prendre la décision d'utiliser une nouvelle technologie (Venkatesh et al., 2000). Bien qu'il n'y ait pas de démonstration explicite d'une différence entre les genres dans la confiance ou l'intention d'utiliser une technologie, la décision a été prise de vérifier la présence d'un effet considérant la nature exploratoire de l'étude.

Tous les items mentionnés précédemment, sauf ceux portant sur les caractéristiques démographiques, ont été évalués par une échelle Likert à 5 ancres allant de 1 (pas du tout d'accord) à 5 (tout à fait d'accord). Finalement, l'ensemble des analyses ont été réalisées par l'entremise du logiciel IBM SPSS Statistics 28.



4. Résultats

4.1 Validité de construit

Bien que l'application de la conceptualisation de la confiance envers les technologies de l'information au contexte de l'IA soit supportée sur le plan théorique, elle n'a jusqu'à présent été que très peu validée empiriquement. Ainsi, une analyse factorielle exploratoire (AFE) a été réalisée afin d'étudier les patrons de variance commune au sein des items développés dans la présente étude, permettant de générer des idées préliminaires quant à la manière dont se structure le construit de la confiance envers l'IA.

La méthode d'extraction de type maximum de vraisemblance a été utilisée aux fins de la présente AFE. Cette méthode est préconisée parce qu'elle permet d'obtenir une solution probable en appliquant le postulat de normalité multivariée à l'échelle de la population. Puisque ce postulat est plus difficilement respecté lorsque des données sont collectées dans la pratique, la méthode du maximum de vraisemblance permet de diminuer le poids des corrélations les plus susceptibles d'être affectées par l'erreur d'échantillonnage, et maximise la probabilité que les patrons de variance observés proviennent effectivement de la population (Tabachnick et Fidell, 2019). Afin de promouvoir l'interprétabilité de la structure factorielle présente dans les données, une rotation oblique (de type *direct oblimin*) a été appliquée. Celle-ci permet la présence de corrélations entre les facteurs (Fabrigar et al., 1999), une considération pertinente dans le cas de la présente étude. Puisque la rotation oblique permet, mais n'oblige pas, la corrélation entre les facteurs, elle peut répliquer les résultats d'une rotation orthogonale, ce qui la rend plus universellement applicable que cette dernière (Costello et Osborne, 2005).

Avant de procéder à l'AFE, une vérification a été réalisée dans le but de déterminer si les données issues de l'échantillon étaient distribuées de manière convenable pour procéder à une analyse factorielle. À cet effet, cinq indices ont été vérifiés, soit la qualité de représentation initiale, le KMO, le test de Bartlett, le déterminant de la matrice de corrélations et le pourcentage de résiduels. Tabachnick et Fidel (2019) suggèrent qu'une qualité de représentation initiale peut être considérée comme acceptable lorsque celle-ci est plus grande que 0,20, pourvu que le KMO soit supérieur à 0,60. Dans le cas de la présente étude, la qualité de représentation initiale moyenne (après extraction) était de 0,46 et le KMO de 0,88. Ceci indique que les items partagent un niveau acceptable de variance entre eux, et que les patrons de corrélation entre les données sont adaptés à l'analyse factorielle. Le test de Bartlett est significatif ($p < 0,001$), ce qui indique un niveau suffisant de redondance entre les items pour qu'une structure factorielle plus parcimonieuse soit adéquate, et le déterminant de la matrice de corrélations (0,011) est supérieur au seuil acceptable de 0,0001, ce qui pointe vers l'absence de multicollinéarité dans les données. Un bon ajustement entre les données observées au sein des participants à l'étude et les données attendues au sein de la population se traduit en résiduels inférieurs à 0,05, avec au maximum 50 % des résiduels supérieurs à ce seuil. Dans le cas présent, 9 résiduels sont supérieurs à 0,05, représentant 20 % des données. Le seuil de 50 % n'ayant pas été atteint, nous pouvons considérer comme acceptable la correspondance entre les données observées et attendues. En somme, le processus de vérification indique que les données recueillies au sein de l'échantillon sont convenables aux fins d'une AFE.

Le nombre de facteurs retenus lors d'une AFE dépend de différents critères, notamment des valeurs propres, du pourcentage de variance expliquée, des indices de saturation et de la pertinence avec le construit théorique à l'étude (Hair et al., 2006). Dans le cas de la présente analyse, deux valeurs propres sont supérieures à 1 (Kaiser, 1960), reflétant la présence d'un facteur principal qui explique 48,79 % de la variance totale du modèle, et d'un facteur secondaire qui explique 10,29 % de la variance. Les indices de saturation indiquent que la variance de 9 des 10 items du questionnaire de la confiance envers l'IA contribuent d'une magnitude suffisante ($> 0,40$; Stevens, [2009]) au premier facteur pour considérer qu'ils en font partie. Le dixième item (fonctionnalité 3) étant le seul à saturer sur un deuxième facteur, la décision

a été prise de le retirer de l'instrument. Une nouvelle AFE a été réalisée avec les 9 items restants, dont les résultats reflètent une structure à un facteur expliquant 51,47 % de la variance (tableau 2).

Tableau 2
Variance commune expliquée par chaque facteur extrait (modèle à 9 items)

Facteur	Valeur propre	% de la variance	% cumulatif
1	4,63	51,48	51,48
2	0,83	9,25	60,72
3	0,72	8,00	68,72
4	0,69	7,64	76,36
5	0,65	7,18	83,54
6	0,53	5,89	89,43
7	0,50	5,54	94,97
8	0,27	2,97	97,94
9	0,19	2,06	100,00

Note. Les valeurs propres supérieures à 1 ont été retenues (Kaiser, 1960).

Une analyse des indices de saturation indique que tous les items contribuent à la variance du facteur dans une magnitude d'au moins 0,50 (tableau 3). Ces indices étant satisfaisants ($> 0,40$; Stevens [2009]), la décision a été prise de procéder à la vérification des hypothèses avec la version de l'instrument de mesure à 9 items.

Tableau 3
Indices de saturation et qualité de représentation (modèle à 9 items)

	Saturations	Qualité de représentation (après rotation)
1. The AI technology presented in the scenario would perform its role well. (Compétence 2)	0,87	0,46
2. The AI technology presented in the scenario is competent and effective. (Compétence 4)	0,85	0,27
3. The AI technology presented in the scenario is capable and proficient in what it has to do. (Compétence 3)	0,81	0,32
4. The AI technology presented in the scenario is a reliable piece of software. (Fiabilité 1)	0,68	0,30
5. The AI technology presented in the scenario is knowledgeable about the prediction it makes. (Compétence 1)	0,60	0,29
6. The AI technology presented in the scenario is dependable. (Fiabilité 3)	0,57	0,35
7. The AI technology presented in the scenario has useful functionalities. (Fonctionnalité 1)	0,55	0,76
8. The AI technology presented in the scenario has the ability to do what I want it to do. (Fonctionnalité 2)	0,54	0,65
9. The AI technology presented in the scenario would not fail. (Fiabilité 2)	0,52	0,72



4.2 Fiabilité

La cohérence interne de l'instrument a été vérifiée par l'entremise de l'alpha de Cronbach, qui permet d'évaluer le niveau de convergence entre les items (Streiner, 2003). Le calcul des alphas de Cronbach pour chacune des variables à l'étude (tableau 4) démontre une consistance interne acceptable ($\alpha > 0,70$; Taber [2017]) pour la mesure de la confiance envers l'IA, de l'intention d'utiliser l'IA et de la propension à faire confiance à la technologie. L'expérience préalable avec l'IA a été évaluée par l'entremise de deux items. À cet effet, le coefficient de corrélation Spearman-Brown s'avère être une mesure moins biaisée de fiabilité que l'alpha de Cronbach (Eisinga et al., 2013). Dans le cas de l'expérience préalable avec l'IA, le coefficient de corrélation Spearman-Brown est de 0,50, ce qui se situe sous le seuil acceptable de 0,70 (de Vet et al., 2017). La décision a été prise de conserver cette variable dans les analyses puisqu'il s'agit d'une variable de contrôle, et non d'une variable centrale à l'étude, et que son effet ne sera pas examiné, mais plutôt maintenu constant dans les analyses afin d'isoler la variance propre à la relation entre les variables d'intérêt (Schmitt, 1996).

Tableau 4
Fiabilité des échelles de mesure

Dimension	Nb Items	α
Confiance envers l'IA	9	0,88
Intention d'utiliser l'IA	3	0,86
Propension à faire confiance à la technologie	3	0,76
Expérience préalable avec l'IA	2	0,50 ^a

Note. a = coefficient de corrélation Spearman-Brown

4.3 Vérification de la manipulation expérimentale

Une vérification de la manipulation expérimentale a été ajoutée au questionnaire afin de vérifier si les deux scénarios utilisés se distinguaient de manière significative au niveau de l'explicabilité, c'est-à-dire que l'explication ajoutée à la condition expérimentale a bien permis aux participants de mieux comprendre le fonctionnement de l'IA dépeinte dans le scénario. À cet effet, deux items ont été développés, un premier concernant l'explicabilité de l'IA et un deuxième concernant le temps de passation du questionnaire, un aspect qui n'a pas été manipulé entre les deux groupes.

Un test t à échantillons indépendants démontre une différence significative entre le groupe expérimental et le groupe contrôle au niveau de l'explicabilité (tableau 5). Plus précisément, les participants du groupe expérimental rapportent mieux comprendre que les participants du groupe contrôle le mécanisme par lequel l'IA dépeinte dans le scénario prend ses décisions. La taille de l'effet, évaluée par le d de Cohen, indique la présence d'un grand effet. Comme attendu, aucune différence significative n'a été observée en ce qui a trait au temps de passation du questionnaire. Ce résultat démontre que l'ajout de l'explication au scénario du groupe expérimental a effectivement permis de clarifier le « comment » l'IA en vient à prendre ses décisions pour les participants.

Tableau 5
Vérification de la manipulation expérimentale

Items	Groupe expérimental		Groupe contrôle		<i>t</i> (118)	<i>d</i> de Cohen
	<i>M</i>	<i>ÉT</i>	<i>M</i>	<i>ÉT</i>		
1. I understand the mechanism through which the AI technology presented in the scenario makes its decisions.	4,15	0,84	3,27	1,21	4,66***	0,85
2. This survey can be completed in a reasonable amount of time	4,57	0,77	4,58	0,65	-0,13	-0,02

Note. *** $p < 0,001$

4.4 Vérification de l'hypothèse 1

Afin de vérifier la première hypothèse de recherche, une régression hiérarchique a été réalisée. Celle-ci permet de contrôler la variance de covariables afin d'isoler l'effet de la variable indépendante. Dans le cadre de la présente étude, les covariables « propension à faire confiance » et « expérience d'utilisation » entretiennent un lien significatif avec la confiance et l'intention d'utiliser l'IA, contrairement à l'âge et au genre (tableau 6). La décision a donc été prise d'exclure ces dernières des analyses.

Tableau 6
Statistiques descriptives et matrice des corrélations entre les variables

Variables	<i>n</i>	<i>M</i>	<i>ÉT</i>	1	2	3	4	5	6
1. Âge	120	1,61	0,73	-					
2. Genre	120	1,55	0,53	0,04	-				
3. Propension à faire confiance	120	3,66	0,88	-0,01	0,05	-			
4. Expérience préalable	120	2,15	0,94	0,04	0,17	0,62**	-		
5. Confiance envers l'IA	120	3,34	0,66	-0,05	0,14	0,55**	0,33**	-	
6. Intention d'utilisation	120	3,35	1,00	0,02	-0,04	0,43**	0,28**	0,73**	-

Note. ** $p < 0,01$

Les résultats de l'analyse de régression hiérarchique (tableau 7) indiquent que la confiance envers l'IA entretient un lien significatif et positif avec l'intention d'utiliser cette technologie et ce, lorsque la propension à faire confiance et l'expérience préalable sont contrôlées. Ce résultat indique que l'hypothèse 1 est soutenue.



Tableau 7
Lien entre la confiance envers l'IA et l'intention d'utilisation

Variable	R ²	ΔR ²	β	t
Étape 1	0,21	-		
Propension à faire confiance			0,38	4,53***
Expérience préalable			0,18	2,06*
Étape 2	0,53	0,32		
Propension à faire confiance			0,04	0,55
Expérience préalable			0,04	0,54
Confiance envers l'IA			0,69	8,81***

Note. * $p < 0,05$, *** $p < 0,001$

4.5 Vérification de l'hypothèse 2

Afin de vérifier la deuxième hypothèse, une analyse de la covariance (ANCOVA) a été réalisée. Celle-ci permet de vérifier s'il existe des différences significatives entre les moyennes de différents groupes tout en contrôlant pour l'effet de covariables. Avant de procéder à l'ANCOVA, une vérification du postulat de l'homogénéité de la régression a été réalisée. L'objectif de celle-ci est d'évaluer si la relation entre une variable indépendante et une variable de contrôle est comparable d'un niveau à l'autre de la variable indépendante, à défaut de quoi les résultats de l'analyse risquent d'être biaisés (Schneider et al., 2015). Pour les besoins de cette analyse, le test proposé par Tabachnick et Fidell (2019) nous révèle la présence d'interactions significatives entre les covariables retenues aux fins de l'étude et la variable indépendante (tableau 8). Ceci indique que l'effet de l'explicabilité sur la confiance varie selon les différents niveaux d'expérience préalable.

Tableau 8
Vérification du postulat d'homogénéité de la régression

Items	B		SS _M	F(2, 117)	η _p ²
	Groupe expérimental	Groupe contrôle			
Expérience préalable avec l'IA x Explicabilité	0,21***	0,27***	6,34	8,10***	0,12
Propension à faire confiance à la technologie x Explicabilité	0,40***	0,42***	15,78	25,42***	0,30

Note. *** $p < 0,001$

En présence d'effets d'interaction, il n'est pas recommandé d'interpréter l'effet principal d'une ANCOVA (Leppink, 2018). Plutôt, une investigation des effets simples est préconisée afin de postuler l'existence, ou non, d'une différence significative entre deux groupes pour différents niveaux des variables de contrôle. Dans le cas de la présente étude, les effets simples de la relation entre l'explicabilité et la confiance à différents niveaux de propension à faire confiance aux technologies et d'expérience préalable avec l'IA ont été analysés afin de déterminer si les groupes contrôle et expérimental diffèrent significativement en matière de confiance envers l'IA.

Pour ce faire, les deux covariables ont été recodées afin de permettre l'analyse de la relation explicabilité-confiance à différents points de mesure. L'expérience d'utilisation de la technologie a été séparée en deux groupes, de part et d'autre de la moyenne de chaque variable. La décision de séparer les groupes au niveau de la moyenne a été prise afin de permettre l'investigation d'effets à différents niveaux de chaque covariable tout en maintenant un nombre suffisant de participants dans chacune des conditions. Cette opération s'est traduite par la création d'un groupe expérience élevée (> 2,15) et d'un autre groupe expérience faible (< 2,15). La propension à faire confiance a aussi été séparée de la même manière, engendrant la propension à faire confiance élevée (> 3,66) et la propension à faire confiance faible (< 3,66).

La comparaison entre la confiance moyenne des groupes expérimental et contrôle (tableau 9) indique qu'il n'y a pas de différence significative entre eux, autant pour les niveaux élevés que faibles des covariables. Ce résultat indique que l'hypothèse 2 n'est pas soutenue.

Tableau 9
Différence entre les groupes contrôle et expérimental (explicabilité) en matière de confiance envers l'IA à différents niveaux des covariables

Items	Groupe expérimental			Groupe contrôle			<i>t</i> (44)	<i>p</i>
	<i>M</i>	<i>ÉT</i>	<i>N</i>	<i>M</i>	<i>ÉT</i>	<i>N</i>		
Expérience faible	3,22	0,52	27	3,25	0,56	38	-0,23	0,82
Expérience élevée	3,38	0,86	33	3,55	0,86	22	-0,83	0,41
Propension faible	3,08	0,57	23	2,95	0,75	23	0,70	0,49
Propension élevée	3,45	0,63	37	3,62	0,52	37	-1,31	0,19

4.6 Résultats supplémentaires

Certaines analyses supplémentaires ont été réalisées afin d'explorer davantage les liens entre les variables à l'étude étant donné les résultats partagés obtenus. Notamment, la différence entre les groupes contrôle et expérimental a été explorée pour ce qui est de l'intention d'utiliser l'IA. Tout d'abord, on observe qu'à niveaux confondus d'explicabilité, les participants ayant une propension plus élevée à faire confiance à la technologie ont une intention d'utiliser l'IA significativement plus élevée que les participants dont la propension à faire confiance est faible (tableau 10) (cette différence dans les intentions d'utilisation n'a pas été observée entre les participants ayant une expérience faible et élevée d'utilisation de l'IA).

Tableau 10
Différences entre les niveaux faible et élevé de propension à faire confiance à la technologie et d'expérience préalable avec l'IA

Items	Propension faible		Propension élevée		<i>t</i> (118)	<i>d</i> de Cohen
	<i>M</i>	<i>ÉT</i>	<i>M</i>	<i>ÉT</i>		
Intention d'utiliser l'IA	3,01	1,05	3,57	0,90	-3,11**	0,58
	Expérience faible		Expérience élevée			
	<i>M</i>	<i>ÉT</i>	<i>M</i>	<i>ÉT</i>		
	3,26	0,97	3,47	1,02	-1,16	0,21

Note. ***p* < 0,01



La présence d'interactions significatives entre l'explicabilité et l'expérience préalable ainsi que la propension à faire confiance (tableau 8) indique que l'effet de l'explicabilité varie selon les différents niveaux des covariables. Lorsqu'on se penche sur l'impact de l'explicabilité sur l'intention d'utiliser l'IA, ceci se traduit en une intention d'utiliser l'IA significativement plus élevée lorsque l'explicabilité est faible (groupe contrôle) que lorsque l'explicabilité est élevée (groupe expérimental) (tableau 11). En somme, si on ne prend pas en compte l'explicabilité, les participants dont la propension à faire confiance à la technologie est plus élevée ont aussi davantage l'intention d'utiliser l'IA. Toutefois, chez les participants dont la propension à faire confiance est élevée, ceux à qui une explication a été fournie ont moins l'intention d'utiliser l'IA que ceux en condition de boîte noire. Aucune différence significative dans l'intention d'utiliser l'IA n'a été observée entre les groupes contrôle et expérimental à différents niveaux de l'expérience préalable avec l'IA.

Tableau 11
Analyses supplémentaires – Différence entre les groupes contrôle et expérimental (explicabilité) pour ce qui est de l'intention d'utiliser l'IA à différents niveaux des covariables

Items	Groupe expérimental			Groupe contrôle			t(44)
	M	ÉT	N	M	ÉT	N	
Expérience faible	2,99	0,92	27	3,45	0,98	38	-1,91
Expérience élevée	3,29	1,01	33	3,73	0,99	22	-1,58
Propension faible	2,88	0,98	23	3,13	1,12	23	-0,79
Propension élevée	3,32	0,94	37	3,81	0,80	37	-2,39*

Note. * $p < 0,05$

5. Discussion et conclusion

Cette étude visait à explorer l'effet de l'explicabilité de l'IA sur la confiance envers celle-ci, ainsi que le lien entre la confiance et l'intention d'utiliser cette technologie en contexte de travail. Les résultats de nos analyses révèlent les trois constats suivants. Tout d'abord, plus la confiance des travailleurs sondés était élevée, plus ceux-ci ont révélé avoir l'intention d'utiliser l'IA dans l'exercice de leur emploi. Deuxièmement, la présence d'une explication éclaircissant le « comment » l'IA en vient à prendre ses décisions n'a pas eu l'impact escompté sur la confiance des participants, comme le suggère l'absence de différence significative entre le niveau de confiance des groupes expérimental et contrôle. Troisièmement, et étonnamment, l'ajout de l'explication a eu un effet délétère sur l'intention d'utiliser l'IA. À la lumière de ces constats, la présente étude offre les contributions théoriques et pratiques suivantes.

5.1 Implications théoriques

Support pour le lien entre la confiance envers l'IA et l'intention de l'utiliser

Nos résultats renforcent la littérature émergente portant sur le rôle joué par la confiance envers l'IA lorsqu'il s'agit d'en favoriser l'utilisation. Cette littérature a largement été établie à partir du modèle de Gefen et al. (2003), qui a bonifié la modélisation des prédicteurs d'utilisation de nouvelles technologies abordées par le TAM (Davis, 1989) en y ajoutant la conceptualisation de la confiance cognitive de McKnight (1998; 2002).

À ce jour, peu de chercheurs ont testé la pertinence du modèle de Gefen et al. (2003), incluant la relation entre la confiance et l'intention d'utilisation, dans le contexte novateur de l'IA. Des chercheurs s'étant penchés sur le sujet, la majorité valide la présence d'un lien positif entre la confiance et l'intention d'utiliser l'IA (Chaudhry et al., 2022; Choung et al., 2021; Liu et al., 2022; Waung et al., 2021), quoique ce constat ne

fait pas entièrement l'unanimité (Tran et al., 2021). Il est à noter qu'une grande partie de ces études s'inscrivent dans le cadre d'applications spécifiques de l'IA au domaine de la médecine. La présente recherche permet ainsi d'élargir les connaissances en proposant des observations basées sur un échantillon ainsi qu'une application de l'IA propre au domaine de la gestion des ressources humaines.

Nouvelles perspectives quant aux impacts de l'explicabilité

La démonstration empirique de l'impact de l'explicabilité d'une IA, c'est-à-dire la clarification de la boîte noire de l'IA sur la confiance envers celle-ci, en est à ses débuts. et les résultats de la présente étude vont à l'encontre du courant théorique prédominant à ce sujet (Ashoori et Weisz, 2019; Chowdury et al., 2023; Ferrario et Loi, 2022; Liu et al., 2022; Miller, 2019). À cet effet, nous suggérons que la dynamique entre ces variables s'avère complexe et mérite d'être approfondie davantage.

Premièrement, bien que notre explication ait été élaborée en suivant les recommandations d'auteurs préconisées dans la littérature (Ashoori et Weisz, 2019) et ait effectivement clarifié le fonctionnement de l'IA, elle n'a pas eu l'impact escompté sur la confiance des travailleurs. À cet effet, il convient de relever que ce domaine d'étude en est à ses débuts et qu'il est possible que d'autres types d'explications soient plus pertinents afin de promouvoir la confiance des travailleurs. Par exemple, Larasati et al. (2020) suggèrent qu'en raison de la justification et du rationnel qu'elle fournit, une explication de type « contrastée » (qui porte sur les effets de causalité à l'intérieur d'un système d'IA et qui est davantage lié au « pourquoi » l'IA en vient à une décision plutôt qu'une autre) ou de type « exhaustive » (qui décrit de manière détaillée les critères sous-tendant les décisions du système) aurait un impact plus marqué sur la confiance qu'une explication de type « générale » (qui décrit le fonctionnement du système de manière simple et sans entrer dans les détails, semblable au type d'explication utilisée dans la présente étude).

Deuxièmement, nos résultats suggèrent que certains facteurs concomitants pourraient avoir réduit l'influence de l'explication sur la confiance, notamment le rôle potentiel de l'expérience préalable du travailleur avec la tâche réalisée par l'IA décrite dans le scénario (une tâche d'évaluation des compétences). À cet effet, Wang et Yin (2021) ont démontré que lorsqu'une explication portant sur le « comment » un algorithme d'IA en vient à prendre ses décisions est fournie, l'individu ayant déjà réalisé lui-même ce type de tâche devient enclin à comparer implicitement sa manière de faire à celle de l'IA. Ainsi, tout écart perçu par l'individu risque d'engendrer chez celui-ci un doute quant à la compétence de l'algorithme. Considérant la démonstration empirique d'un lien positif entre la compétence perçue d'une technologie d'IA et la confiance envers celle-ci (Shi et al., 2021), il est possible qu'un doute quant à la compétence de l'IA vienne affaiblir l'effet bénéfique de l'explication sur la confiance. Cette hypothèse souligne l'intérêt d'explorer davantage le rôle de l'expérience préalable du travailleur avec la tâche réalisée par l'IA et, de manière plus large, de considérer les caractéristiques propres aux tâches réalisées par l'IA lorsqu'il s'agit de promouvoir la confiance envers celle-ci.

Enfin, la présente étude démontre un impact négatif inattendu de l'explication sur l'intention d'utiliser l'IA, plus particulièrement lorsque le niveau de propension à faire confiance est élevé. Une hypothèse visant à expliquer ce résultat est ancrée dans le modèle *Unified Theory of Acceptance and Use of Technology* (UTAUT). Initialement développé par Venkatesh et al. (2003) dans le contexte de l'utilisation de technologies de l'information, le modèle connaît actuellement un nouvel essor avec l'utilisation de plus en plus répandue de l'IA. Dans les dernières années, Venkatesh (2022) a proposé une version du modèle adapté à l'IA. Celui-ci fait état de quatre prédicteurs d'intention et de comportement réel d'utilisation de l'IA, dont l'un d'eux est l'effort attendu (*effort expectancy*). L'effort attendu est défini comme la perception de facilité d'utilisation d'une technologie, c'est-à-dire que plus le niveau d'effort attendu d'une technologie est élevé, plus celle-ci semble facile à utiliser et plus forte sera l'intention de l'utiliser (Lin et al., 2022). Dans son adaptation du modèle UTAUT à l'IA, Venkatesh (2022) suggère de tenir compte de quatre nouvelles variables, soit les



caractéristiques propres à l'individu, à la technologie, à l'environnement et à l'intervention. Au sein du modèle, ces variables peuvent jouer le rôle d'antécédents aux prédicteurs existants, de modérateurs entre les prédicteurs et l'intention d'utiliser l'IA, ou même de nouveaux prédicteurs d'intention d'utilisation. En lien avec la présente recherche, la propension à faire confiance ferait partie des caractéristiques propres à l'individu, et entretiendrait le rôle de nouveau prédicteur des intentions d'utiliser l'IA. Ceci expliquerait l'observation d'une intention d'utilisation plus élevée chez nos participants ayant davantage la propension à faire confiance que chez nos participants ayant moins cette inclination.

En parallèle, nous avons observé que lorsque la propension à faire confiance est élevée, la présence d'une explication est associée à un niveau plus faible d'intention d'utilisation qu'en condition de boîte noire. Ce résultat donne lieu à diverses hypothèses. Tout d'abord, il est possible que les gens ayant des niveaux élevés de propension à faire confiance aient une perception d'emblée positive de l'IA, sans nécessairement avoir exploré son fonctionnement et ses particularités au préalable. Ainsi, il est possible que l'explication éclaircissant la boîte noire ait complexifié leur perception initialement positive de l'IA. Ceci fait référence au phénomène de « lune de miel » associé à l'émergence de l'IA et de nouvelles technologies (Asan et al., 2020; Bedué et al. 2021; Carton et al., 2020), qui se traduit en perceptions démesurément optimistes de celles-ci avant d'en avoir fait l'expérience. En lien avec le modèle UTAUT, ceci signifierait que l'explicabilité jouerait le rôle d'antécédent au prédicteur d'intention d'utilisation « effort attendu ». Plus précisément, l'explicabilité ne ferait pas qu'éclairer le fonctionnement de l'IA, elle en rendrait le fonctionnement plus complexe à première vue, c'est-à-dire dans un contexte où le travailleur n'a pas encore fait l'expérience de la technologie. Cette complexité engendrerait une diminution de l'effort attendu et, en retour, une baisse dans l'intention d'utilisation. Cette relation entre l'explicabilité et l'effort attendu ne serait toutefois significative qu'en présence de propension élevée, conférant à cette dernière un rôle modérateur. En effet, les individus ayant une propension plus faible à faire confiance seraient déjà sceptiques envers l'IA et auraient déjà une intention plus faible de l'utiliser. Ainsi, la présence d'une explication n'aurait pas d'impact sur leurs perceptions et intentions d'utilisation déjà mitigées.

En somme, nos observations révèlent un côté délétère inédit à la clarification de la boîte noire de l'IA. Ce résultat apporte un bémol à la recherche actuelle, qui est principalement favorable à l'éclaircissement de l'IA, et incite les chercheurs à prendre en considération les caractéristiques propres aux individus telles que la propension à faire confiance à la technologie en combinaison avec les explications données sur le fonctionnement de l'IA pour bien comprendre ce qui favorisera son utilisation.

Mesure de la confiance envers l'IA

La conceptualisation de la confiance cognitive envers l'IA a jusqu'à présent été peu étudiée, notamment en raison de l'arrivée récente de ce type de technologie dans les milieux de travail. La présente étude contribue ainsi à raffiner les connaissances portant sur ce sujet émergent en proposant une conceptualisation nous ayant permis d'expliquer 51 % de la variance dans la confiance rapportée par notre échantillon. Ce seuil est considéré comme acceptable dans le domaine des sciences humaines (Hair et al., 2019; Pett et al. 2003), et est comparable à la proportion de variance expliquée de 50 % par le modèle phare de McKnight et al. (2011) de la confiance envers les technologies de l'information, dont la grande majorité des études portant sur la confiance envers l'IA sont inspirées. Ainsi, notre outil offre aux chercheurs et entreprises la possibilité d'évaluer la confiance cognitive envers l'IA dans un contexte où l'arrivée d'une technologie d'IA est anticipée, c'est-à-dire avant que les travailleurs aient l'occasion de l'utiliser.

En ce qui a trait à la structure dimensionnelle de la confiance cognitive, nos données convergent vers une conceptualisation unidimensionnelle de celle-ci, plutôt que vers la conceptualisation en trois dimensions de McKnight et al. (2002; 2011). Certaines pistes peuvent potentiellement expliquer ce résultat. Tout d'abord, il est possible que le construit de la confiance envers l'IA diffère conceptuellement de la confiance envers

les technologies de l'information, et que les dimensions de la fiabilité, de la fonctionnalité et de la compétence soient moins saillantes dans ce contexte. Par contre, sous un angle méthodologique, il est aussi possible que l'approche de vignettes hypothétiques utilisées dans la présente étude n'ait pas permis de susciter chez les participants l'expérience de stimuli et de facteurs contextuels aussi saillants que l'on retrouverait dans un contexte d'utilisation réelle de l'IA. Il est donc envisageable que ce soit le contexte expérimental qui n'ait pas été en mesure de faire ressortir les nuances entre les trois dimensions postulées.

Limites de l'étude et pistes de recherche

Cette étude possède certaines limites qui éclairent diverses pistes de recherche. Premièrement, considérant l'arrivée récente de l'IA dans les organisations, la décision a été prise d'utiliser une méthodologie de vignettes expérimentales ayant fait ses preuves dans l'exploration de phénomènes émergents (Atzmüller et Steiner, 2010). Toutefois, cette approche limite la généralisation de nos résultats à un contexte d'utilisation anticipée de l'IA. Ceci dit, considérant le caractère novateur de l'IA et de son arrivée progressive dans les organisations, ce contexte d'utilisation anticipée s'avère pertinent au stade d'implantation de l'IA dans lequel se situent plusieurs organisations. Une prochaine étape essentielle demeure tout de même la réplication de cette étude dans un contexte d'utilisation réelle de l'IA afin d'en tirer des conclusions généralisables et de générer un outil de mesure applicable à une population plus large, en plus de permettre l'exploration de comportements réels d'utilisation de la technologie.

Une autre limite concerne le devis transversal de l'étude. Puisque l'ensemble des données ont été recueillies dans un même temps de mesure, nous ne pouvons pas réaliser d'inférences causales sur le lien entre la confiance et l'intention d'utiliser l'IA. À cet effet, un devis longitudinal permettrait d'y remédier. Combiné à un contexte d'utilisation réelle de l'IA, un devis longitudinal permettrait d'élargir notre compréhension des effets de l'explicabilité et de la confiance non seulement sur l'intention d'utiliser la technologie, mais aussi sur son utilisation effective à court et à plus long terme. De plus, ce type de devis permettrait d'observer comment l'expérience d'utilisation de l'IA module la confiance envers celle-ci au fil du temps.

Il est important de noter que la consistance interne de la mesure d'expérience préalable avec l'IA s'est avérée être sous le seuil de l'acceptabilité. Ceci apporte un bémol à la fiabilité et à la validité de nos résultats. Ainsi, les résultats de l'étude devraient être interprétés avec une certaine réserve. La recherche future pourrait adresser cette limite en employant une mesure plus robuste de l'expérience préalable avec l'IA, qui comprendrait notamment plus de deux items (Emons et al., 2007).

Finalement, notre étude a eu recours à un échantillon de convenance composé de travailleurs en ressources humaines provenant d'une même entreprise. Cette méthode d'échantillonnage, bien qu'efficace en matière d'accessibilité et de coût, alerte sur le risque que des biais propres à l'entreprise sondée influencent les résultats et limitent leur généralisation. Ainsi, nous encourageons la recherche future à répliquer nos analyses avec un échantillon plus diversifié, c'est-à-dire comprenant des participants provenant d'autres organisations et domaines d'emploi.

Outre ses limites, la présente étude propose de nouvelles perspectives de recherche. Notamment, notre étude renforce le soutien grandissant pour le rôle clé de la confiance afin de promouvoir l'intention d'utiliser l'IA (Choudhury et al., 2022; Liu et al., 2022; Waung et al., 2021). À cet effet, nous suggérons d'explorer le rôle de la confiance envers l'IA dans le cadre du modèle UTAUT. En plus de l'effort attendu (*effort expectancy*) décrit précédemment, le modèle comprend trois autres prédicteurs de l'intention d'utiliser et de l'utilisation réelle de l'IA. Ceux-ci sont les attentes de performance (*performance expectancy*), soit le degré auquel le travailleur perçoit que l'IA sera en mesure d'améliorer sa performance en emploi, l'influence sociale (*social influence*), soit la perception que les autres veulent que le travailleur utilise l'IA, et les conditions facilitantes (*facilitating conditions*), soit la croyance qu'une structure organisationnelle et technique existe afin



de soutenir l'utilisation de l'IA. À la lumière de nos résultats et du soutien grandissant pour le rôle de la confiance, il serait indiqué de tester le rôle de cette variable au sein du modèle et de préciser sur le plan théorique comment elle s'agence et interagit avec les quatre prédicteurs initiaux proposés par Venkatesh (2003). De plus, les rôles hypothésés précédemment de la propension à faire confiance à la technologie en tant que prédicteur d'intention d'utilisation et de modérateur entre l'explicabilité et l'effort attendu, ainsi que de l'explicabilité en tant qu'antécédent de l'effort attendu, méritent d'être examinés dans le cadre d'études empiriques. Jusqu'à présent, l'ensemble de ces relations a fait l'objet de très peu d'études empiriques au sein du modèle UTAUT.

5.2 Implications pratiques

Nous soulignons l'importance pour les organisations de considérer les conditions de succès propres aux individus sujets à utiliser l'IA dans l'exercice de leurs fonctions lorsqu'il s'agit d'en promouvoir l'utilisation. Notamment, dans le contexte où une technologie d'IA serait attendue dans une entreprise, les gestionnaires gagneraient à promouvoir la confiance de leurs employés envers cette technologie afin de générer des intentions positives d'utilisation. Bien que l'explicabilité de l'IA ne s'avère pas être la panacée initialement avancée, on note d'autres types de démarches pouvant être mises en place afin de favoriser la confiance des travailleurs envers l'IA dans l'optique de les préparer en amont du changement. À cet effet, on observe un impact positif sur la confiance lorsque les travailleurs sont invités à cocréer le processus d'implantation de l'IA avec l'organisation, et lorsque ceux-ci reçoivent une formation leur permettant de mieux comprendre l'IA, comment celle-ci sera utilisée dans leurs rôles spécifiques ainsi que ses implications pour l'organisation (Lane et al., 2023). L'instrument de mesure développé dans la présente étude pourrait servir à dresser un portrait de la confiance des travailleurs avant et après le déploiement de ces initiatives et permettrait à l'entreprise d'en évaluer le succès en ce qui a trait à la promotion de la confiance auprès de ses employés dans un contexte de préutilisation. De plus, notre étude souligne l'importance de considérer les caractéristiques propres aux individus, telle la propension à faire confiance à la technologie. Ainsi, les interventions susceptibles d'avoir plus de succès seront celles qui prendront en compte ces caractéristiques et qui pourront offrir un certain niveau de personnalisation et d'adaptation en fonction des individus.

RÉFÉRENCES

- Agrawal, A., Gans, J. S. et Goldfarb, A. (2019). Exploring the impact of artificial intelligence: Prediction versus judgment. *Information Economics and Policy*, 47, 1-6. <https://doi.org/10.1016/j.infoecopol.2019.05.001>
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. et Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Asan, O., Bayrak, A. E. et Choudhury, A. (2020). Artificial intelligence and human trust in healthcare: focus on clinicians. *Journal of Medical Internet Research*, 22(6). <https://doi.org/10.2196/15154>
- Ashoori, M. et Weisz, J. D. (2019). In AI we trust? Factors that influence trustworthiness of AI-infused decision-making processes. *arXiv*, 1-10. <https://doi.org/10.48550/arXiv.1912.02675>
- Atzmüller, C. et Steiner, P. M. (2010). Experimental vignette studies in survey research. *Methodology European Journal of Research Methods for the Behavioral and Social Sciences*, 6(3), 128-138. <https://doi.org/10.1027/1614-2241/a000014>
- Bedué, P. et Fritzsche, A. (2022). Can we trust AI? An empirical investigation of trust requirements and guide to successful AI adoption. *Journal of Enterprise Information Management*, 35(2), 530-549. <https://doi.org/10.1108/JEIM-06-2020-0233>
- Benbasat, I. et Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, 6(3), 4. <https://doi.org/10.17705/1jais.00065>
- Black, J. S. et van Esch, P. (2020). AI-enabled recruiting: What is it and how should a manager use it?. *Business Horizons*, 63(2), 215-226. <https://doi.org/10.1016/j.bushor.2019.12.001>

- Boon, S. et Holmes, J. (1991). The dynamics of interpersonal trust: Resolving uncertainty in the face of risk. Dans R. Hinde et J. Gorebel (dir.), *Cooperation and prosocial behaviour* (190-211). Cambridge University Press.
- Braun, M., Bleher, H. et Hummel, P. (2021). A leap of faith: is there a formula for “Trustworthy” AI?. *Hastings Center Report*, 51(3), 17-22. <https://doi.org/10.1002/hast.1207>
- Braganza, A., Chen, W., Canhoto, A. et Sap, S. (2021). Productive employment and decent work: The impact of AI adoption on psychological contracts, job engagement and employee trust. *Journal of Business Research*, 131, 485-494. <https://doi.org/10.1016/j.jbusres.2020.08.018>
- Bruhn, J. et Anderer, M. (2019). Implementing artificial intelligence in organizations and the special role of trust. *Media Trust in a Digital World: Communication at Crossroads*, 191-205. https://doi.org/10.1007/978-3-030-30774-5_14
- Burton, J. W., Stein, M. K. et 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>
- Cabiddu, F., Moi, L., Patriotta, G. et Allen, D. G. (2022). Why do users trust algorithms? A review and conceptualization of initial trust and trust over time. *European Management Journal*, 40(5), 685-706. <https://doi.org/10.1016/j.emj.2022.06.001>
- Carton, S., Mei, Q. et Resnick, P. (2020). Feature-based explanations don't help people detect misclassifications of online toxicity. *Proceedings of the International AAAI Conference on Web and Social Media*, 14, 95-106. <https://doi.org/10.1609/icwsm.v14i1.7282>
- Castelvecchi, D. (2016). Can we open the black box of AI?. *Nature News*, 538(7623), 20-23. <https://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731>
- Choung, H., David, P. et Ross, A. (2022). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human-Computer Interaction*, 39(3), 1-13. <https://doi.org/10.1080/10447318.2022.2050543>
- Chaudhry, I. S., Paquibut, R. Y. et Chabchoub, H. (2022). Factors influencing employees trust in AI & its adoption at work: Evidence from United Arab Emirates. *Proceedings of the International Arab Conference on Information Technology*, 1-7. <https://doi.org/10.1109/acit57182.2022.9994226>
- Chowdhury, S., Dey, P., Joel-Edgar, S., Bhattacharya, S., Rodriguez-Espindola, O., Abadie, A. et Truong, L. (2023). Unlocking the value of artificial intelligence in human resource management through AI capability framework. *Human Resource Management Review*, 33(1). <https://doi.org/10.1016/j.hrmr.2022.100899>
- Confalonieri, R., Coba, L., Wagner, B et Besold, T. R. (2021). A historical perspective of explainable Artificial Intelligence. *WITRES Data Mining and Knowledge Discovery*, 11(1). <https://doi.org/10.1002/widm.1391>
- Costello, A. B. et Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, 10(7). <https://doi.org/10.7275/yj1-4868>
- Custers, B. (2022). AI in criminal law: An overview of AI applications in substantive and procedural criminal law. Dans B. Custers et E. Fosch-Villaronga (dir.), *Law and artificial intelligence. Information technology and law series* (vol. 35). T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-523-2_11
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 13(3), 319-340. <https://doi.org/10.2307/249008>
- de Vet, H. C., Mokkink, L. B., Mosmuller, D. G. et Terwee, C. B. (2017). Spearman–Brown prophecy formula and Cronbach's alpha: different faces of reliability and opportunities for new applications. *Journal of Clinical Epidemiology*, 85, 45-49. <https://doi.org/10.1016/j.jclinepi.2017.01.013>
- Dietz, G. et Hartog, D. N. (2006). Measuring trust inside organisations. *Personnel Review*, 35(5), 557-588. <https://doi.org/10.1108/00483480610682299>
- Dimensional Research. (2019, mai). *Artificial intelligence and machine learning projects are obstructed by data issues global survey of data scientists, global survey of data scientists, AI experts and stakeholders*. <https://t.ly/FGo54>
- Eisinga, R., Grotenhuis, M. T. et Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?. *International Journal of Public Health*, 58(4), 637-642. <https://doi.org/10.1007/s00038-012-0416-3>



- Emons, W. H. M., Sijtsma, K. et Meteijer, R. R. (2007). On the consistency of individual classification using short scales. *Psychological Methods*, 12(1), 105-120. <https://doi.org/10.1037/1082-989X.12.1.105>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. et Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299. <https://doi.org/10.1037/1082-989X.4.3.272>
- Fan, W., Liu, J., Zhu, S. et Pardalos, P. M. (2020). Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). *Annals of Operations Research*, 294, 567-592. <https://doi.org/10.1007/s10479-018-2818-y>
- Ferrario, A. et Loi, M. (2022). The robustness of counterfactual explanations over time. *IEEE Access*, 10, 82736-82750. <https://doi.org/10.1109/ACCESS.2022.3196917>
- Gefen, D. (2000). E-commerce: the role of familiarity and trust. *Omega*, 28(6), 725-737. [https://doi.org/10.1016/S0305-0483\(00\)00021-9](https://doi.org/10.1016/S0305-0483(00)00021-9)
- Gefen, D., Karahanna, E. et Straub D. W. (2003). Trust and TAM in online shopping: an integrated model. *MIS Quarterly*, 27(1), 51-90. <https://doi.org/10.2307/30036519>
- Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U. et Redzepi, A. (2022). The dark sides of people analytics: Reviewing the perils for organisations and employees. *European Journal of Information Systems*, 31(3), 410-435. <https://doi.org/10.1080/0960085X.2021.1927213>
- Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B. et Spaulding, R. (2021). Attachment and trust in artificial intelligence. *Computers in Human Behavior*, 115. <https://doi.org/10.1016/j.chb.2020.106607>
- Glikson, E. et 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>
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F. et Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), 1-42. <https://doi.org/10.1145/3236009>
- Gulati, R. (1995). Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances. *Academy of Management Journal*, 38(1), 85-112. <https://doi.org/10.5465/256729>
- Gulati, S., Sousa, S. et Lamas, D. (2019). Towards an empirically developed scale for measuring trust. *Proceedings of the 31st European Conference on Cognitive Ergonomics*, 45-48. <https://doi.org/10.1145/3335082.3335116>
- Gunning, D. et Aha, D. W. (2019). DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine*, 40(2), 44-58. <https://doi.org/10.1609/aimag.v40i2.2850>
- Hair, J. F., Black, W., Babin, B., Andeson, R. et Tatham, R. (2006). *Multivariate data analysis* (6^e éd.). Pearson-Prentice Hall.
- Hair, J. F., Page, M. et Brunsveld, N. (2019). *Essentials of business research methods* (4^e éd.). Routedledge-Taylor & Francis.
- Hasija, A. et Esper, T. L. (2022). In artificial intelligence (AI) we trust: A qualitative investigation of AI technology acceptance. *Journal of Business Logistics*, 43(3), 388-412. <https://doi.org/10.1111/jbl.12301>
- Hengstler, M., Enkel, E. et Duelli, S. (2016). Applied artificial intelligence and trust: The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*, 105, 105-120. <https://doi.org/10.1016/j.techfore.2015.12.014>
- Hmoud, B. et Laszlo, V. (2019). Will artificial intelligence take over human resources recruitment and selection. *Network Intelligence Studies*, 7(13), 21-30. <https://dea.lib.unideb.hu/server/api/core/bitstreams/cff8e1b9-7db6-47e9-9642-6fe8e19f565d/content>
- Hmoud, B. I. et Várallyai, L. (2020). Artificial intelligence in human resources information systems: Investigating its trust and adoption determinants. *International Journal of Engineering and Management Sciences*, 5(1), 749-765. <https://doi.org/10.21791/IJEMS.2020.1.65>
- Höddinghaus, M., Sondern, D. et Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms?. *Computers in Human Behavior*, 116. <https://doi.org/10.1016/j.chb.2020.106635>
- Hoff, K. A. et 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>

- Hughes, C., Robert, L., Frady, K. et Arroyos, A. (2019). Artificial intelligence, employee engagement, fairness, and job outcomes: Managing technology and middle and low skilled employees. Dans *The changing context of managing people* (p. 61-68). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-78973-077-720191005>
- Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, 20(1). <https://doi.org/10.1177/001316446002000116>
- Kaplan, A. D., Kessler, T. T., Brill, J. C. et Hancock, P. A. (2023). Trust in artificial intelligence: Meta-analytic findings. *Human Factors*, 65(2), 337-359. <https://doi.org/10.1177/00187208211013988>
- Kim, J. Y. et Heo, W. (2022). Artificial intelligence video interviewing for employment: perspectives from applicants, companies, developer and academicians. *Information Technology & People*, 35(3), 861-878. <https://doi.org/10.1108/ITP-04-2019-0173>
- Knowles, B. et Hanson, V. L. (2018). The wisdom of older technology (non) users. *Communications of the ACM*, 61(3), 72-77. https://eprints.lancs.ac.uk/id/eprint/88296/1/CACM_Pure.pdf
- Koufaris, M. et Hampton-Sosa, W. (2004). The development of initial trust in an online company by new customers. *Information & Management*, 41(3), 377-397. <https://doi.org/10.1016/j.im.2003.08.004>
- Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, 8(1). <https://doi.org/10.1525/collabra.33267>
- Lane, M., Williams, M. et Broecke, S. (2023). *The impact of AI on the workplace: Main findings from the OECD AI surveys of employers and workers* (publication no 288). <https://doi.org/10.1787/ea0a0fe1-en>
- Lankton, N. K., McKnight, D. H. et Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems*, 16(10). <https://doi.org/10.17705/1jais.00411>
- Langer, M. et König, C. J. (2023). Introducing a multi-stakeholder perspective on opacity, transparency and strategies to reduce opacity in algorithm-based human resource management. *Human Resource Management Review*, 33(1). <https://doi.org/10.1016/j.hrmr.2021.100881>
- Lapatin, S., Gonçalves, M., Nillni, A., Chavez, L., Quinn, R. L., Green, A. et Alegria, M. (2012). Lessons from the use of vignettes in the study of mental health service disparities. *Health Services Research*, 47(3.2), 1345-1362. <https://doi.org/10.1111/j.1475-6773.2011.01360.x>
- Larasati, R., De Liddo, A. et Motta, E. (2020, mars). *The effect of explanation styles on user's trust*. Dans Workshop on explainable smart systems for algorithmic transparency in emerging technologies. <https://oro.open.ac.uk/70421/1/70421.pdf>
- Lee, P., Fyffe, S., Son, M., Jia, Z. et Yao, Z. (2023). A paradigm shift from “human writing” to “machine generation” in personality test development: An application of state-of-the-art natural language processing. *Journal of Business and Psychology*, 38(1), 163-190. <https://doi.org/10.1007/s10869-022-09864-6>
- Leppink, J. (2018). Analysis of covariance (ANCOVA) vs. moderated regression (MODREG): Why the interaction matters. *Health Professions Education*, 4(3), 225-232. <https://doi.org/10.1016/j.hpe.2018.04.001>
- Liao, Q. V. et Sundar, S. S. (2022). Designing for responsible trust in AI systems: a communication perspective. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1257-1268. <https://doi.org/10.1145/3531146.3533182>
- Lichtenthaler, U. (2020). Five maturity levels of managing AI: from isolated ignorance to integrated intelligence. *Journal of Innovation Management*, 8(1), 39-50. https://doi.org/10.24840/2183-0606_008.001_0005
- Lin, H. C., Ho, C. F. et Yang, H. (2022). Understanding adoption of artificial intelligence-enabled language e-learning system: An empirical study of UTAUT model. *International Journal of Mobile Learning and Organisation*, 16(1), 74-94. <https://doi.org/10.1504/IJML.O.2022.119966>
- Liu, X., He, X., Wang, M. et Shen, H. (2022). What influences patients' continuance intention to use AI-powered service robots at hospitals? The role of individual characteristics. *Technology in Society*, 70. <https://doi.org/10.1016/j.techsoc.2022.101996>
- Lu, L., Cai, R. et Gursoy, D. (2019). Developing and validating a service robot integration willingness scale. *International Journal of Hospitality Management*, 80, 36-51. <https://doi.org/10.1016/j.ijhm.2019.01.005>
- Madsen, M. et Gregor, S. (2000). Measuring human-computer trust. *Proceedings of the 11th Australasian conference on information systems*, 6-8. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b8eda9593fcb63b7ced1866853d9622737533a2>



- Mahmud, H., Islam, A. N., Ahmed, S. I. et Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, 175. <https://doi.org/10.1016/j.techfore.2021.121390>
- Makarius, E. E., Mukherjee, D., Fox, J. D. et Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*, 120, 262-273. <https://doi.org/10.1016/j.jbusres.2020.07.045>
- Mayer, R. C., Davis, J. H. et 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>
- Mcknight, D. H., Carter, M., Thatcher, J. B. et Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems (TMIS)*, 2(2), 1-25. <https://doi.org/10.1145/1985347.1985353>
- McKnight, D. H., Choudhury, V. et Kacmar, C. (2002). The impact of initial consumer trust on intentions to transact with a web site: a trust building model. *The journal of strategic information systems*, 11(3-4), 297-323. [https://doi.org/10.1016/S0963-8687\(02\)00020-3](https://doi.org/10.1016/S0963-8687(02)00020-3)
- McKnight, D. H., Cummings, L. L. et Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review*, 23(3), 473-490. <https://doi.org/10.5465/amr.1998.926622>
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1-38. <https://doi.org/10.1016/j.artint.2018.07.007>
- Mirbabaie, M., Brünker, F., Möllmann, N. R. et Stieglitz, S. (2022). The rise of artificial intelligence—understanding the AI identity threat at the workplace. *Electronic Markets*, 32, 73-99. <https://doi.org/10.1007/s12525-021-00496-x>
- Misztal, B. A. (2001). Trust and cooperation: the democratic public sphere. *Journal of Sociology*, 37(4), 371-386. <https://doi.org/10.1177/144078301128756409>
- Moray, N. et Inagaki, T. (1999). Laboratory studies of trust between humans and machines in automated systems. *Transactions of the Institute of Measurement and Control*, 21(4-5), 203-211. <https://doi.org/10.1177/0142331299021004>
- Ore, O. et Sposato, M. (2022). Opportunities and risks of artificial intelligence in recruitment and selection. *International Journal of Organizational Analysis*, 30(6), 1771-1782. <https://doi.org/10.1108/IJOA-07-2020-2291>
- Pasmore, W., Winby, S., Mohrman, S. A. et Vanasse, R. (2019). Reflections: sociotechnical systems design and organization change. *Journal of Change Management*, 19(2), 67-85. <https://doi.org/10.1080/14697017.2018.1553761>
- Pereira, V., Hadjielias, E., Christofi, M. et Vrontis, D. (2023). A systematic literature review on the impact of artificial intelligence on workplace outcomes: A multi-process perspective. *Human Resource Management Review*, 33(1). <https://doi.org/10.1016/j.hrmr.2021.100857>
- Pett, M. A., Lackey, N. R. et Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Sage.
- Pettersen, L. (2019). Why artificial intelligence will not outsmart complex knowledge work. *Work, Employment and Society*, 33(6), 1058-1067. <https://doi.org/10.1177/0950017018817489>
- Rajpurkar, P., Chen, E., Banerjee, O. et Topol, E. J. (2022). AI in health and medicine. *Nature medicine*, 28(1), 31-38. <https://doi.org/10.1038/s41591-021-01614-0>
- Ramachandran, K. K., Mary, A. A. S., Hawladar, S., Asokk, D., Bhaskar, B. et Pitroda, J. R. (2022). Machine learning and role of artificial intelligence in optimizing work performance and employee behavior. *Materials Today: Proceedings*, 51(8), 2327-2331. <https://doi.org/10.1016/j.matpr.2021.11.544>
- Ransbotham, S., Kiron, D., Candelon, F., Khodabandeh, S. et Chu, M. (2022). *Achieving individual and organizational value with AI*. MIT Sloan Management Review. <https://sloanreview.mit.edu/projects/achieving-individual-and-organizational-value-with-ai/>
- Robinson, M. D. et Clore, G. L. (2001). Simulation, scenarios, and emotional appraisal: Testing the convergence of real and imagined reactions to emotional stimuli. *Personality and Social Psychology Bulletin*, 27(11), 1520-1532. <https://doi.org/10.1177/01461672012711012>
- Rossi, F. (2018). Building trust in artificial intelligence. *Journal of International Affairs*, 72(1), 127-134. <https://www.jstor.org/stable/26588348>

- Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35(4), 651–665. <https://doi.org/10.1111/j.1467-6494.1967.tb01454.x>
- Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, 26(5), 443–452. <https://doi.org/10.1037/h0031464>
- Russel, S. et Susskind, D. (2021, octobre). *Positive AI Economic Futures*. World Economic Forum. https://www3.weforum.org/docs/WEF_Positive_AI_Economic_Futures_2021.pdf
- Ryan, M. (2020). In AI we trust: ethics, artificial intelligence, and reliability. *Science and Engineering Ethics*, 26(5), 2749–2767. <https://doi.org/10.1007/s11948-020-00228-y>
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350–353. <https://doi.org/10.1037/1040-3590.8.4.350>
- Schneider, B. A., Avivi-Reich, M. et Mozuraitis, M. (2015). A cautionary note on the use of the Analysis of Covariance (ANCOVA) in classification designs with and without within-subject factors. *Frontiers in Psychology*, 6(474), 1–12. <https://doi.org/10.3389/fpsyg.2015.00474>
- Shi, S., Gong, Y. et Gursoy, D. (2021). Antecedents of trust and adoption intention toward artificially intelligent recommendation systems in travel planning: a heuristic–systematic model. *Journal of Travel Research*, 60(8), 1714–1734. <https://doi.org/10.1177/0047287520966395>
- Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. *International Journal of Human-Computer Studies*, 146. <https://doi.org/10.1016/j.ijhcs.2020.102551>
- Shulner-Tal, A., Kuflik, T. et Kliger, D. (2023). Enhancing fairness perception—Towards human-centred AI and personalized explanations understanding the factors influencing laypeople’s fairness perceptions of algorithmic decisions. *International Journal of Human–Computer Interaction*, 39(7), 1455–1482. <https://doi.org/10.1080/10447318.2022.2095705>
- Siau, K. et Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal*, 31(2), 47–53. https://www.researchgate.net/publication/324006061_Building_Trust_in_Artificial_Intelligence_Machine_Learning_and_Robotics
- Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A. et Leimeister, J. M. (2012). Understanding the formation of trust in IT artifacts. *Proceedings of the International Conference on Information Systems (ICIS 2012)*. https://doi.org/10.1007/978-3-319-05044-7_3
- Söllner, M., Hoffmann, A. et Leimeister, J. M. (2016). Why different trust relationships matter for information systems users. *European Journal of Information Systems*, 25(3), 274–287. <https://doi.org/10.1057/ejis.2015.17>
- Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5^e éd.). Routedledge-Taylor & Francis.
- Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18
- Tabachnick, B. G. et Fidell, L. S. (2019). *Using multivariate statistics* (7^e éd.). Pearson Education.
- Taber, K. S. (2018). The use of Cronbach’s alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(1), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Tarafdar, M., Tu, Q. et Ragu-Nathan, T. S. (2010). Impact of technostress on end-user satisfaction and performance. *Journal of Management Information Systems*, 27(3), 303–334. <https://doi.org/10.2753/MIS0742-1222270311>
- Tran, A. Q., Nguyen, L. H., Nguyen, H. S. A., Nguyen, C. T., Vu, L. G., Zhang, M., Vu, T. M. T., Nguyen, S. H., Tran, B. X., Latkin, C. A., Ho, R. C. M. et Ho, C. S. (2021). Determinants of intention to use artificial intelligence-based diagnosis support system among prospective physicians. *Frontiers in Public Health*, 9. <https://doi.org/10.3389/fpubh.2021.755644>
- Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G. et Van Moorsel, A. (2020). The relationship between trust in AI and trustworthy machine learning technologies. *Proceedings of the 2020 conference on fairness, accountability, and transparency* (272–283). <https://doi.org/10.1145/3351095.3372834>
- Trist, E. L. et Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting: an examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. *Human Relations*, 4(1), 3–38. <https://doi.org/10.1177/001872675100400101>



- Tsiakas, K. et Murray-Rust, D. (2022). Using human-in-the-loop and explainable AI to envisage new future work practices. *Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments*. <https://doi.org/10.1145/3529190.3534779>
- Venkatesh, V. (2022). Adoption and use of AI tools: a research agenda grounded in UTAUT. *Annals of Operations Research*, 308, 1-12. <https://doi.org/10.1007/s10479-020-03918-9>
- Venkatesh, V., Morris, M. G. et Ackerman, P. L. (2000). A longitudinal field investigation of gender differences in individual technology adoption decision-making processes. *Organizational Behavior and Human Decision Processes*, 83(1), 33-60. <https://doi.org/10.1006/obhd.2000.2896>
- Venkatesh, V., Morris, M. G., Davis, G. B. et Davis, F. D. (2003). User acceptance of information technology: toward a unified view. *MIS Quarterly*, 27(3), 425-478. <https://doi.org/10.2307/30036540>
- Vereschak, O., Bailly, G. et Caramiaux, B. (2021). How to evaluate trust in AI-assisted decision making? A survey of empirical methodologies. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1-39. <https://doi.org/10.1145/3476068>
- von Eschenbach, W. J. (2021). Transparency and the black box problem: why we do not trust AI. *Philosophy & Technology*, 34(4), 1607-1622. <https://doi.org/10.1007/s13347-021-00477-0>
- Wadden, J. J. (2022). Defining the undefinable: the black box problem in healthcare artificial intelligence. *Journal of Medical Ethics*, 48(10), 764-768. <https://doi.org/10.1136/medethics-2021-107529>
- Wang, X. et Yin, M. (2021). Are explanations helpful? a comparative study of the effects of explanations in ai-assisted decision-making. *Proceeding of the 26th International Conference on Intelligent User Interfaces*. <https://doi.org/10.1145/3397481.3450650>
- Waung, M., McAuslan, P. et Lakshmanan, S. (2021). Trust and intention to use autonomous vehicles: manufacturer focus and passenger control. *Transportation Research Part F: Traffic Psychology and Behaviour*, 80, 328-340. <https://doi.org/10.1016/j.trf.2021.05.004>
- Westphal, M., Vössing, M., Satzger, G., Yom-Tov, G. B. et Rafeali, A. (2023). Decision control and explanations in human-AI collaboration: improving user perceptions and compliance. *Computers in Human Behavior*, 144. <https://doi.org/10.1016/j.chb.2023.107714>
- Wright, S. A. et Schultz, A. E. (2018). The rising tide of artificial intelligence and business automation: developing an ethical framework. *Business Horizons*, 61(6), 823-832. <https://doi.org/10.1016/j.bushor.2018.07.001>
- Xiang, Y., Zhao, L., Liu, Z., Wu, X., Chen, J., Long, E., Lin, D., Chen, C., Lin, Z. et Lin, H. (2020). Implementation of artificial intelligence in medicine: status analysis and development suggestions. *Artificial Intelligence in Medicine*, 102. <https://doi.org/10.1016/j.artmed.2019.101780>
- Yang, R. et Wibowo, S. (2022). User trust in artificial intelligence: A comprehensive conceptual framework. *Electronic Markets*, 32(4), 2053-2077. <https://doi.org/10.1007/s12525-022-00592-6>
- Yu, X., Xu, S. et Ashton, M. (2023). Antecedents and outcomes of artificial intelligence adoption and application in the workplace: the socio-technical system theory perspective. *Information Technology & People*, 36(1), 454-474. <https://doi.org/10.1108/ITP-04-2021-0254>
- Zhang, B., Zhu, Y., Deng, J., Zheng, W., Liu, Y., Wang, C. et Zeng, R. (2023). "I am here to assist your tourism": predicting continuance intention to use AI-based chatbots for tourism. Does gender really matter?. *International Journal of Human-Computer Interaction*, 39(9), 1887-1903. <https://doi.org/10.1080/10447318.2022.2124345>
- Zielonka, J. T. (2022). The impact of trust in technology on the appraisal of technostress creators in a work-related context. *Proceedings of the 55th Hawaii International Conference on System Sciences*. <http://hdl.handle.net/10125/80055>

ANNEXE A

Scénario Groupe Expérimental

Please read the following scenario carefully:

Imagine that you are a team leader in your organization and looking to hire new people to join your team. The HR department of the organization has just announced that recruitment decisions will now have to incorporate data gathered through an artificial intelligence (AI) supported assessment. During this assessment, participants are presented a series of situations depicting key events that happen in the day-to-day life of an employee in the role that you are looking to hire. After the presentation of each scenario, participants are asked to explain what they would do would they find themselves in that situation.

Traditionally (before the AI supported version of the assessment), participants were asked to select the best response for each question among multiple-choice answers. In the new, AI supported version of the assessment, participants are asked to answer each question in an open-ended, free-form text format. Their responses are then analyzed in real time by an artificial intelligence algorithm. At the end of the assessment, a report containing the participant's scores is automatically generated and directly sent to you by email.

In the new, AI supported version of the assessment, participants are asked to answer each question in an open-ended, free-form text format. This text is analyzed in real time by an artificial intelligence algorithm, which is trained to identify what a great and not great answer looks like through an extensive database of various optimal and suboptimal answers. Prepared by assessment experts, this database also indicates the number of points corresponding to each answer. The algorithm works to match each participant's response as closely as possible to one of the answers contained in the database, and attributes points to the participant accordingly. At the end of the assessment, a report containing the participant's scores is automatically generated and directly sent to you by email.

Keeping this scenario and the AI technology depicted in mind, please indicate the degree to which you agree or disagree with the following statements.



Scénario Groupe Contrôle

Please read the following scenario carefully:

Imagine that you are a team leader in your organization and looking to hire new people to join your team. The HR department of the organization has just announced that recruitment decisions will now have to incorporate data gathered through an artificial intelligence (AI) supported assessment. During this assessment, participants are presented a series of situations depicting key events that happen in the day-to-day life of an employee in the role that you are looking to hire. After the presentation of each scenario, participants are asked to explain what they would do would they find themselves in that situation.

Traditionally (before the AI supported version of the assessment), participants were asked to select the best response for each question among multiple-choice answers. In the new, AI supported version of the assessment, participants are asked to answer each question in an open-ended, free-form text format. Their responses are then analyzed in real time by an artificial intelligence algorithm. At the end of the assessment, a report containing the participant's scores is automatically generated and directly sent to you by email.

Keeping this scenario and the AI technology depicted in mind, please indicate the degree to which you agree or disagree with the following statements.

ANNEXE B

Mesure de la confiance cognitive

- **Fiabilité (Lankton et al., 2015)**

The AI technology presented in the scenario is a reliable piece of software (Fiabilité_1)

The AI technology presented in the scenario would not fail (Fiabilité_2)

The AI technology presented in the scenario is dependable (Fiabilité_3)

- **Fonctionnalité (Lankton et al., 2015)**

The AI technology presented in the scenario has features that would be useful to do my tasks (Fonctionnalité_3)

The AI technology presented in the scenario has the ability to do what I want it to do (Fonctionnalité_2)

The AI technology presented in the scenario has useful functionalities (Fonctionnalité_1)

- **Compétence (Shi et al., 2021)**

The AI technology presented in the scenario is knowledgeable about the prediction it makes (Compétence_1)

The AI technology presented in the scenario would perform its role well (Compétence_2)

The AI technology presented in the scenario is capable and proficient in what it has to do (Compétence_3)

The AI technology presented in the scenario is competent and effective (Compétence_4)

Mesure de l'intention d'utiliser l'IA (Venkatesh et al., 2003)

I would use the AI technology presented in the scenario in the future.

I plan on using the AI technology if it is available to me.

I would use the AI technology presented in the scenario in the near future.

Vérification de la manipulation expérimentale

I understand the mechanism through which the AI technology presented in the scenario makes its decisions.

This survey can be completed in a reasonable amount of time.

Mesure de la propension à faire confiance à la technologie (Lankton et al., 2015)

I tend to give a technology the benefit of the doubt when I first use it.

I usually trust a technology until it gives me a reason not to.

My approach is to trust new technologies until they prove that I shouldn't trust them.

Mesure de l'expérience d'utilisation de la technologie (Gefen et al., 2003)

I am familiar with the type of AI technology presented in the scenario.

I have used the AI technology presented in the scenario before.