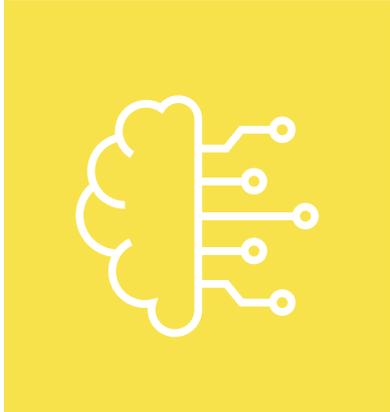


Une perspective d'équité en matière d'intelligence artificielle



Une perspective d'équité en matière d'intelligence artificielle

Carmina Ravanera et Sarah Kaplan

Institute for Gender and the Economy,
Rotman School of Management, Université de Toronto
Le 15 août 2021

Remerciements

Les auteurs tiennent à remercier les professeurs Joshua Gans, Avi Goldfarb et Gillian Hadfield pour leurs commentaires utiles sur les premières ébauches de ce projet.

Ce rapport est disponible en français et en anglais à l'adresse suivante :

www.gendereconomy.org/artificial-intelligence

Le projet «Une perspective d'équité en matière d'intelligence artificielle» est cofinancé par le Conseil de recherches en sciences humaines et une subvention dans le cadre du programme Compétences futures du gouvernement du Canada no 872 2020 0011.

"An Equity Lens on Artificial Intelligence" is co-funded by the Social Sciences and Humanities Research Council and the Government of Canada's Future Skills Program Grant #872-2020-0011.

Table des matières

- 1** **Résumé**

- 2** **Une perspective d'équité en
matière d'intelligence artificielle**
 - Introduction

 - Une arme à double tranchant :
IA et l'équité (ou l'iniquité)

- 4** **IA et répercussions inévitables**

- 5** **Défis liés à l'élimination des iniquités en IA**

- 8** **IA, automatisation et emploi**

- 9** **Répercussions sur la recherche, les politiques
et la pratique**

- 10** **Vers l'équité en IA**

- 12** **Notes**

Résumé

L'intelligence artificielle (IA) s'obtient des machines qui peuvent simuler certaines formes d'intelligence humaine, par exemple, relever des tendances, faire des prévisions et prendre des décisions. Aujourd'hui, l'IA est utilisée par des entreprises de nombreux secteurs à diverses fins, allant de l'embauche d'employés à l'évaluation des risques, en passant par la formulation de recommandations en matière d'investissement et de condamnations criminelles. Cependant, il est bien connu que les relations sociales et les contextes sont représentés et reproduits dans la technologie. L'IA ne fait pas exception : elle recèle le potentiel de renforcer les biais, la discrimination et les iniquités sous jacents. Bien que l'IA puisse être utilisée dans l'intérêt de groupes marginalisés, il est nécessaire que les entreprises et les gouvernements mettent l'accent sur l'équité dans l'IA pour atténuer les possibles préjudices. Nous fournissons ici une ressource aux chercheurs et aux praticiens pour qu'ils voient l'IA du point de vue de l'équité afin de résumer les recherches et les connaissances existantes sur le lien entre l'IA et l'équité (ou l'iniquité) et de proposer des considérations dont les dirigeants des secteurs public et privé doivent tenir compte dans la mise en œuvre de l'IA.

Concept clé - L'IA est une arme à double tranchant qui peut à la fois atténuer et renforcer les biais:

- Comme l'IA utilise des méthodes de prévision statistique qui peuvent être vérifiées, elle peut créer des résultats qui aident les groupes marginalisés dans des situations où les décisions humaines peuvent être assombries par des biais cognitifs.
- Malgré ce potentiel, puisque les inégalités et les iniquités sont souvent représentées dans les technologies, certains systèmes d'IA peuvent renforcer ou ont renforcé la marginalisation de certains groupes, notamment les femmes, les minorités de genre et les communautés racisées et à faible revenu. Les produits et services fondés sur l'IA peuvent utiliser des ensembles de données préjudiciables qui reproduisent ces biais et amplifient les stéréotypes et la marginalisation, parfois dans le but d'obtenir certains avantages, et élargir les asymétries de pouvoir.
- Le renforcement de l'iniquité et de l'inégalité est survenu à cause de biais intégrés ou d'omissions importantes dans les ensembles de données; de la complexité et de compromis liés à l'harmonisation de l'IA avec les valeurs sociales lorsque les profits sont également en jeu; d'un manque de transparence de la part de ceux qui créent et mettent en œuvre de l'IA; d'un manque de responsabilisation à l'égard du public ou d'autres utilisateurs de l'IA et d'une participation limitée des groupes marginalisés et diversifiés dans le secteur de la technologie.
- Il existe également des répercussions possibles variées de l'IA et de l'automatisation sur les emplois et le travail. Il est possible que des femmes, des groupes racisés et des groupes à faible revenu soient plus susceptibles de perdre leur poste en raison de l'automatisation dans un nombre croissant d'emplois de cols bleus, blancs et roses.

Ces résultats nous amènent à inviter les entreprises et les gouvernements à tenir compte des considérations suivantes :

- Les entreprises technologiques et les gouvernements peuvent se concentrer sur des initiatives visant à assurer une représentation équitable dans le développement de l'IA.
- Les créateurs, les chercheurs et les responsables de la mise en œuvre de systèmes d'IA peuvent prioriser l'harmonisation de l'IA avec des valeurs sociales comme l'équité, malgré les compromis possibles par rapport à l'efficacité et à la rentabilité.
- Les gouvernements peuvent créer des politiques sur l'IA qui accordent la priorité à la responsabilisation et à la transparence et exigent des entreprises qu'elles adhèrent à ces principes.
- Les gouvernements et les entreprises peuvent travailler de sorte à contribuer à la sécurité économique des travailleurs qui sont doublement touchés par les nouvelles technologies et la pandémie mondiale en portant une attention aux programmes de recyclage ou de perfectionnement des compétences.
- Les chercheurs universitaires peuvent approfondir leurs connaissances sur l'IA et l'iniquité, notamment en poursuivant des travaux interdisciplinaires sur les répercussions sociales, politiques et environnementales de l'IA et en concevant des solutions de rechange nouvelles et différentes qui priorisent l'atténuation des préjudices.

Une perspective d'équité en matière d'intelligence artificielle

Introduction

L'intelligence artificielle (IA) s'obtient des machines qui peuvent simuler certaines formes d'intelligence humaine. Certaines conceptualisations de l'IA font référence à des machines qui agissent de la même manière que les humains, tandis que d'autres mettent davantage l'accent sur l'apprentissage automatique qui permet de cerner des tendances, d'obtenir un résultat optimal pour un problème donné, de faire des prévisions et de prendre des décisions fondées sur des renseignements antérieurs.¹ L'IA utilise des algorithmes pour atteindre ces résultats, habituellement en « apprenant » à partir de grands ensembles de données, puis en s'ajustant et en s'améliorant à partir de nouvelles données. Bien que l'IA ne soit pas un concept nouveau ou récent, elle est aujourd'hui intégrée dans la vie des gens et ne fait que devenir de plus en plus omniprésente.

L'IA est utilisée par les entreprises de tous les secteurs à diverses fins, comme l'embauche d'employés, la réalisation d'interventions chirurgicales, le tutorat dans diverses matières scolaires, la prise de décisions pour la détermination de peines criminelles, l'automatisation de la conduite automobile et la prévision des lieux où se produiront des crimes. Elle sert à formuler des recommandations sur ce que les gens regardent à la télévision ou sur la musique qu'ils écoutent, à choisir les publicités à montrer aux utilisateurs sur les médias sociaux et à afficher les résultats sur les moteurs de recherche en ligne.² Il ne s'agit pas d'une liste exhaustive. En fait, il peut être difficile de trouver aujourd'hui un secteur ou un domaine où l'IA n'intervient pas d'une quelconque façon. L'IA est devenue omniprésente au point que certains chercheurs ont laissé entendre qu'elle était devenue un nouveau type d'infrastructure. Plutôt que d'être un produit physique ou une institution, comme des routes ou des systèmes éducatifs, elle est immatérielle et souvent invisible, mais elle est néanmoins une « modératrice des relations sociales, des pratiques et des actions », y compris la répartition du pouvoir.³ Les relations sociales et les valeurs sont représentées et reproduites dans la technologie et l'IA ne fait pas exception. Cela signifie également que les biais, la discrimination et les inégalités qui sont profondément enracinés dans la société peuvent aussi être profondément enracinés dans cette technologie.⁴

Bien que ces nouvelles technologies modifient le fonctionnement de l'économie, des entreprises et de la société, les communautés aux quatre coins du Canada continuent de composer avec des inégalités et des iniquités sociales, économiques et politiques qui peuvent amplifier les répercussions de l'IA ou être amplifiées par

celle-ci. En 2020 et 2021, les impacts économiques de la pandémie de COVID-19 ont été ressentis plus fortement par les groupes qui étaient déjà marginalisés, plus particulièrement les femmes, les communautés racisées et les personnes à faible revenu. Les chercheurs et les analystes de politiques ont suggéré que les politiques de relance doivent porter une attention particulière à ces groupes pour éviter une augmentation des inégalités.⁵

Pour comprendre les impacts de l'IA sur l'économie et la société au Canada, surtout dans le contexte du ralentissement économique causé par la pandémie de COVID-19, il faut comprendre ses impacts sur les groupes marginalisés. L'IA peut être utilisée de façon novatrice pour produire des résultats qui sont dans l'intérêt de diverses communautés. Toutefois, la recherche a également montré qu'il est nécessaire de mettre l'accent sur l'IA équitable pour les entreprises et les décideurs afin d'atténuer les préjudices.

De nombreuses disciplines et entreprises participent à ce discours. De nombreux points de vue proviennent de domaines aussi variés que le droit, l'informatique et la philosophie. Au moment de la publication, de nouvelles recherches auront probablement vu le jour. Il ne s'agit donc pas de couvrir toute la documentation sur ce sujet, mais de fournir une ressource qui met en évidence les débats, les questions et les enjeux, explore le « pourquoi » et le « comment » de l'incidence de l'IA sur l'équité et l'égalité et présente aux dirigeants, aux décideurs et aux étudiants des observations sur l'utilisation actuelle et ultérieure de l'IA. La marginalisation est fondée sur de nombreuses situations sociales et leurs intersections, comme le genre, la race, l'indigénéité, le statut socioéconomique, le statut d'immigration, la capacité et la religion. Toutefois, la présente synthèse porte principalement sur le genre et la race, car sa portée devait être limitée pour servir son objectif. Néanmoins, les auteurs reconnaissent que la technologie touche les personnes ayant différentes identités sociales de diverses façons et que des recherches plus approfondies sont nécessaires pour comprendre ces répercussions de façon plus exhaustive.⁶

Une arme à double tranchant : IA et l'équité (ou l'iniquité)

Potential de l'IA

L'IA recèle le potentiel d'améliorer les résultats pour les gens de tous les secteurs. Idéalement, elle élimine les répercussions possibles de l'erreur humaine en faisant des prévisions exactes et en aidant les humains à prendre des décisions. Par exemple, dans les milieux de travail, l'IA utilisée pour l'embauche pourrait trouver sans biais le meilleur candidat pour un poste. Dans le domaine des soins de santé, elle peut aider à diagnostiquer des maladies et à trouver des traitements. Dans les institutions financières, elle peut prédire la probabilité que des personnes soient en défaut de paiement hypothécaire. Dans les gouvernements, elle peut évaluer les cas de réfugiés afin d'obtenir des résultats plus justes pour les demandeurs.

Le pouvoir prévisionnel de l'IA est important compte tenu du fait que les prévisions et les décisions des humains sont assombries par des biais, notamment cognitifs. Bien souvent, les gens ne comprennent pas entièrement pourquoi ils font certaines prévisions et leur intuition peut être influencée par

leurs expériences ou opinions. Comme l'ont fait remarquer les chercheurs, les prévisions effectuées par l'IA à l'aide de techniques de prévision statistique ont tendance à surpasser celles réalisées par les humains possédant de l'expertise et de l'expérience.⁷ De plus, les prévisions et les prises de décision faites par des humains sont souvent nébuleuses – il est difficile de comprendre et d'approfondir les divers facteurs qui influencent les gens. La prise de décisions de l'humain est également difficile à vérifier. Ainsi, dans la mesure où les algorithmes peuvent être vérifiés et modifiés, l'IA pourrait être un outil d'atténuation de la discrimination, des biais et d'autres formes de marginalisation. C'est-à-dire qu'elle pourrait « devenir un contrepoids puissant à la discrimination humaine et une force positive pour le bien social sous de multiples formes ».⁸

Par exemple, des chercheurs ont montré qu'aux États-Unis, les juges commettent souvent des erreurs au sujet de la probabilité que les défendeurs libérés sous caution commettent un crime, en partie à cause de leurs biais et parce qu'ils se basent fortement sur une accusation actuelle plutôt que sur d'autres renseignements pertinents. Les juges ont tendance à traiter les défendeurs à haut risque comme présentant un faible risque si l'accusation actuelle est mineure et les défendeurs à faible risque comme présentant un risque élevé si l'accusation actuelle est grave. Cela peut entraîner des erreurs de deux types, soit des taux de détention inutilement élevés ou la libération de délinquants à risque élevé, ce qui augmente le taux de criminalité. Des chercheurs ont constaté qu'un algorithme créé pour prédire statistiquement les résultats est plus précis que l'évaluation de juges dans ces cas, ce qui pourrait entraîner moins de détention inutile et moins de crimes. De plus, comme les algorithmes peuvent être ajustés pour optimiser des résultats précis, ce type d'IA pourrait être mis en œuvre dans le but de réduire la détention inutile qui a des répercussions importantes dans les communautés racisées. Autrement dit, elle pourrait être programmée pour réduire les disparités raciales dans les taux de détention tout en maintenant le même taux de criminalité.⁹ Un autre exemple est un outil algorithmique utilisé par le service à l'enfance, à la jeunesse et à la famille du comté d'Allegheny en Pennsylvanie. Il vise à prévoir le risque de préjudice pour un enfant que les responsables de l'enquête sur le signalement pourraient être incapables de faire rapidement et avec autant d'exactitude, ce qui aiderait à mieux orienter les ressources vers les cas à risque élevé. Le New York Times a rapporté en 2018 qu'avec cet outil, les signalements à risque élevé sont examinés plus souvent et que le pourcentage de cas à faible risque examinés inutilement a diminué.¹⁰

Des chercheurs ont également suggéré qu'un système d'IA d'assistance qui a été utilisé dans les cas de réfugiés pourrait réduire les décisions nuisibles aux demandeurs d'asile. Encore une fois, les décisions concernant les cas de réfugiés sont prises par des humains dans des circonstances qui pourraient être nébuleuses, incertaines et préjudiciables, et ce, souvent en raison de données limitées concernant un demandeur. Cela peut entraîner le refus de l'asile dans un cas où il aurait dû être accepté. Les prévisions fondées sur l'IA peuvent offrir à la fois une prévision et une probabilité qui mesure le degré d'incertitude de la prévision. Ainsi, son utilisation dans

le traitement des demandes de réfugiés rendrait explicites les incertitudes dans les données qui éclairent la prise de décisions pour les demandeurs d'asile, par exemple en démontrant qu'il n'y a pas suffisamment de données pour conclure avec exactitude qu'un demandeur d'asile ne sera pas persécuté s'il retourne dans son pays d'origine. Si les structures légales changeaient pour dissiper le doute en faveur du demandeur plutôt que le contraire – ce qui nécessiterait une volonté politique et un changement – un tel système d'IA pourrait aider à s'assurer que les réfugiés ne se voient pas refuser les protections nécessaires en raison d'une erreur humaine.¹¹

L'idée générale ici est que les algorithmes et l'information menant aux prévisions peuvent être programmés au moyen de paramètres décisionnels précis et vérifiés par la suite. Ils peuvent être utiles dans des circonstances changeantes d'iniquité et de marginalisation dont des décideurs humains pourraient autrement ne pas tenir compte. Toutefois, les chercheurs et les défenseurs ont montré que lorsque l'IA est utilisée en pratique, elle n'agit bien souvent pas comme le contrepoids espéré à la discrimination et peut au contraire la renforcer. Les inégalités sociétales peuvent être et sont reproduites dans l'IA et les atténuer peut être difficile. Par exemple, l'outil d'évaluation des risques utilisé pour les services à l'enfance dans le comté d'Allegheny a été critiqué parce qu'il a eu une incidence disproportionnée sur les familles pauvres : l'algorithme utilise la pauvreté comme indicateur de risque élevé de négligence et de mauvais traitements, ce qui est en fait une hypothèse injuste.¹² De même, une étude de l'évaluation du risque par algorithme auprès de défendeurs accusés de crimes en Floride a révélé qu'elle notait incorrectement les défendeurs noirs comme de futurs criminels deux fois plus souvent que les défendeurs blancs, notamment parce que la race était étroitement corrélée à des facteurs jugés à risque élevé, comme le chômage.¹³ De plus, il est possible que les gens n'adhèrent pas aux évaluations fournies par les outils de prévision. Selon une étude d'un outil d'évaluation des risques dans le système de justice pénale du Kentucky, les juges sont plus enclins à mettre de côté les recommandations algorithmiques pour les défendeurs noirs que pour les défendeurs d'autres races, ce qui entraîne des conditions de cautionnement plus strictes pour les défendeurs noirs que les défendeurs blancs similaires.¹⁴ Ces exemples permettent de penser que les relations de pouvoir et les inégalités intégrées dans la société façonnent les données qui alimentent les algorithmes, les algorithmes eux-mêmes et la façon d'utiliser les algorithmes. Cela signifie que le potentiel transformateur de l'IA comporte des risques et des défis importants, et actuellement, des chercheurs et des défenseurs tentent de les contrer.

IA et répercussions inévitables

La technologie et les systèmes d'IA ne sont pas neutres ou objectifs. Ils s'inscrivent plutôt dans un contexte social et historique qui peut marginaliser certains groupes, dont les femmes, les communautés racisées et les communautés à faible revenu. Il existe donc de nombreux exemples de façons dont les systèmes d'IA peuvent reproduire les biais et la marginalisation.¹⁵ Cela peut se produire par des biais ou des lacunes dans les données utilisées pour former des algorithmes, ainsi que par la mise en œuvre de produits et de services alimentés par l'IA de façons qui renforcent des stéréotypes, la marginalisation et des relations de pouvoir mondiales.

Biais et lacunes dans les données

Comme les biais et les inégalités existent à tous les niveaux de la société, il s'ensuit que les données à partir desquelles une certaine IA est construite contiennent de tels biais et que l'IA peut ensuite les reproduire. L'attention portée à la reproduction de la discrimination fondée sur le genre par l'IA n'est pas nouvelle, mais elle demeure un défi constant. Dans les années 1970 et 1980, une école de médecine du Royaume Uni a utilisé un programme informatique pour sélectionner les candidats. Il a fini par rejeter les personnes de genre féminin et portant un nom non européen parce que l'algorithme était basé sur des données antérieures sur des demandes acceptées et que ces personnes y étaient peu représentées.¹⁶ De même, en 2015, Amazon a mis au point un système de recrutement appuyé par l'IA, aujourd'hui hors service, qui a éliminé certaines femmes du bassin de candidats, et ce, parce qu'il était basé sur les modèles d'embauche antérieurs dans lesquels les hommes dominaient.¹⁷

Les mêmes problèmes sont survenus pour des lacunes liées à la race dans les données. Dans le secteur de la santé, un système d'IA utilisé pour détecter les lésions cutanées cancéreuses a été conçu à l'aide d'une base de données contenant principalement des populations ayant la peau pâle, ce qui l'a rendu moins susceptible de fonctionner sur des populations ayant la peau plus foncée.¹⁸ Les biais fondés sur la race et le genre dans les données se recoupent également pour reproduire l'oppression des femmes racisées. Récemment, des chercheurs ont montré comment les logiciels de reconnaissance faciale alimentés par l'IA d'IBM, de Microsoft et de Face++ sont moins précis pour les sujets à la peau plus foncée et plus particulièrement les femmes à la peau plus foncée, ce qui entraîne une probabilité plus élevée d'erreur de classification par rapport aux hommes blancs. Encore une fois, c'est survenu parce que les données sous-jacentes ne comportaient pas une représentation diversifiée des races et des genres.¹⁹ Selon le logiciel de reconnaissance faciale utilisé, cette erreur pourrait renforcer la surveillance et l'identification erronée de personnes racisées et, plus particulièrement, des femmes racisées.

Ce ne sont pas seulement les grandes entreprises technologiques qui font face à ce problème de données.

L'IA est également utilisée dans le secteur public comme les services de police. Une étude récente a montré que de nombreux services de police aux États-Unis utilisent des données préjudiciables sur le plan racial dans des systèmes prédictifs et ces systèmes font ensuite des prévisions préjudiciables sur les personnes qui vont commettre des crimes et les lieux où ceux-ci seront commis. Ils pourraient ainsi cibler différemment les communautés de minorités raciales qui sont déjà trop ciblées.²⁰ Ce type de services de police algorithmiques est déjà à l'étape de la conception ou de la mise en œuvre dans plusieurs corps policiers au Canada et dans les aéroports parallèlement à une technologie de surveillance qui recueille et surveille les données des gens en ligne ou les images.²¹

Renforcement des stéréotypes et de la marginalisation lors de la mise en œuvre de l'IA

Les préjudices causés par l'IA sont non seulement dus aux ensembles de données problématiques, mais aussi aux façons dont les entreprises et les organisations ont conçu et utilisé l'IA pour renforcer les stéréotypes, la marginalisation et l'effacement de certains groupes. Par exemple, un logiciel d'analyse faciale alimenté par l'IA a été utilisé pour propager la fausse idée que les personnes ayant certaines caractéristiques faciales sont sujettes à la criminalité et que le logiciel peut identifier ces personnes. Cela ouvre des possibilités dangereuses pour les communautés racisées qui sont déjà stéréotypées comme étant foncièrement criminelles.²² Les chercheurs ont également montré comment les représentations courantes de l'IA, comme les images génériques et d'autres représentations de robots et de robotiques, ont tendance à être racisées comme blanches et à avoir une apparence et une voix eurocentriques. Ils pensent que la blancheur reproduit les conceptions de l'intelligence, du professionnalisme et du pouvoir. Dans la mesure où l'IA et les machines intelligentes sont souvent créées pour prendre en charge des « emplois salissants, monotones ou dangereux » que les personnes racisées à faible revenu occupent de façon disproportionnée, la blancheur de l'IA peut aussi être perçue comme un moyen d'effacer les personnes racisées et leur travail.²³ Autre exemple : les assistants numériques alimentés par l'IA comme Alexa d'Amazon, Siri d'Apple et Cortana de Microsoft ont un nom et un genre féminin. Les chercheurs ont discuté de la façon dont l'attribution d'un genre à cette technologie réaffirme la division du travail selon le genre, qui place les femmes dans des rôles de subordonnées en prestation de soins et de services pour accomplir des tâches ménagères.²⁴ Ces assistants numériques féminisés agissent à la fois comme assistants et compagnons pour assurer le bien-être des utilisateurs de façon amicale et empathique, ce qui enchâsse davantage les stéréotypes sur les femmes dans des rôles de subordonnées.²⁵

Dans certains cas, le renforcement des stéréotypes par l'entremise de l'IA est explicitement lié à l'obtention de profits. Une récente vérification indépendante de la diffusion algorithmique des offres d'emploi par Facebook a révélé qu'elle perpétue la ségrégation des emplois selon le genre en fonction des répartitions actuelles des genres dans différentes catégories d'emploi : par exemple, une offre

d'emploi en vente de voitures a été montrée à plus d'hommes que de femmes et l'inverse était vrai pour une offre d'emploi en vente de bijoux.²⁶ Les développeurs pourraient ajuster l'algorithme de diffusion publicitaire pour compenser les biais dans les données ou éliminer complètement la diffusion algorithmique des offres d'emploi. Cette correction ne cadrerait toutefois pas avec les motivations de profits à court terme des entreprises technologiques qui sont basées sur le nombre de clics sur des publicités.²⁷ Il faudra donc que les dirigeants s'engagent à apporter des changements pour éliminer ces biais. La recherche s'est également penchée sur la façon dont les algorithmes intégrés aux recherches sur Google, dont beaucoup considèrent les résultats comme neutres et factuels, renforcent les récits racistes et sexistes sur les femmes noires. Dans son livre *Algorithms of Oppression*, la chercheuse Safiya Noble décrit comment sa recherche sur Google avec l'expression « filles noires » [traduction] a donné en première page des résultats montrant la pornographie et l'objectification sexuelle des filles et des femmes noires. Cela montre comment les algorithmes de Google renforcent l'oppression en montrant dans les premiers résultats de recherche ceux qui stimuleront les profits.²⁸

Renforcement des relations de pouvoir à l'échelle mondiale

De nouvelles recherches sont également en cours sur les répercussions de l'IA sous le prisme des relations de pouvoir mondiales. Les travaux de la chercheuse Kate Crawford portent plus particulièrement sur la façon dont l'IA élargit les asymétries de pouvoir. Elle démontre que l'IA est une « industrie extractive » qui « dépend de l'exploitation des ressources énergétiques et minérales de la planète, de la main-d'œuvre bon marché et des données à grande échelle ». ²⁹ Par exemple, la création de systèmes et de logiciels d'IA exige une forte consommation d'énergie et de minéraux, de pétrole et de charbon et les personnes qui créent les composantes matérielles peuvent travailler dans des conditions dangereuses et hautement surveillées dans les usines. Ces coûts sont assumés par les ouvriers dans les communautés minières ou les fabricants en sous traitance, tandis que les entreprises technologiques accumulent la richesse et le pouvoir.³⁰ De cette façon, l'IA peut également être considérée comme un facteur contribuant aux inégalités à l'échelle mondiale.

Il est également important de noter que l'IA est actuellement utilisée pour prendre des décisions sur les demandes d'immigration et de réfugiés par le gouvernement fédéral du Canada, par exemple pour évaluer les demandes et automatiser les activités connexes. Même s'il est possible qu'elle crée des prédictions plus exactes pour les demandeurs d'asile, ces chercheurs pensent que les gouvernements doivent s'assurer que ces décisions peuvent être traitées adéquatement par l'IA compte tenu de la complexité des cas des réfugiés. Ils ont soutenu que l'IA utilisée dans de tels appareils étatiques représente non seulement une menace pour la vie privée, mais qu'elle pourrait aussi entraver le droit à la liberté et le droit à la protection contre la détention arbitraire, ce qui peut avoir des répercussions importantes sur la vie d'immigrants, de migrants et de réfugiés.³¹

Défis liés à l'élimination des iniquités en IA

La recherche suggère différentes explications pour les mécanismes par lesquels l'IA peut perpétuer la dynamique des inégalités et des iniquités. Pour contrer les biais qui peuvent être introduits par les données ou les hypothèses utilisées pour élaborer des algorithmes, il faut faire un effort réfléchi pour les corriger. Il est essentiel de se poser certaines questions. Quelles valeurs sociales doivent être intégrées aux machines, qui décide, comment doit on le faire et comment les fabricants et les utilisateurs de l'IA peuvent ils être tenus responsables? Un manque de transparence pour le public et un manque de responsabilisation de ceux qui élaborent et mettent en œuvre l'IA peuvent représenter des scénarios inquiétants pour l'équité. Une autre influence est que l'IA est souvent créée par des entreprises dans lesquelles les groupes marginalisés ne sont pas représentés, donc elles ne connaissent pas les besoins de ces groupes.

Quelles valeurs? Complexité et compromis

Enseigner l'impartialité et l'équité aux algorithmes est d'abord difficile parce qu'il s'agit de concepts très complexes qui doivent être expliqués à une machine. C'est à dire qu'il est difficile de savoir comment il faut montrer aux algorithmes à adhérer aux normes et aux valeurs sociales, dont beaucoup impliquent des structures complexes comme la loi ou la culture. À l'instar des réflexions des chercheurs, est il possible de programmer l'IA de manière à ce que le fait de poser certains gestes répréhensibles lui impose une pénalité, tout comme les humains évitent de poser certains gestes pour éviter des sanctions sociales comme la honte ou la culpabilité?³² D'autres ont suggéré que la supervision des algorithmes peut servir de « boussole morale » aux algorithmes. Il serait alors possible de surveiller les biais et de les modifier en conséquence.³³ Quelles valeurs faut il prioriser et dans quels cas? Ces questions sont actuellement sous la loupe de personnes qui se penchent sur des enjeux d'éthique en l'IA.

Deuxièmement, il peut y avoir des compromis avec l'exactitude lors de la programmation de telles valeurs dans l'IA.³⁴ Par exemple, un algorithme faisant des prédictions sur les personnes qui seront en défaut de paiement d'un prêt devrait être explicitement programmé pour réduire la disproportion raciale, mais cela pourrait donner lieu à des prédictions moins précises. Pourtant, le fait de ne pas le faire renforcerait l'iniquité compte tenu des antécédents de privation et d'oppression qui ont conduit à des taux accrus de pauvreté et d'insécurité financière pour les communautés racisées. Cette réflexion amène ensuite à se demander si l'utilisation de l'IA devrait être limitée et, le cas échéant, dans quel domaine et si les humains qui font les mêmes décisions et prédictions sont plus ou moins susceptibles de perpétuer un biais. Comme l'ont fait remarquer les chercheurs : « À l'ère des données et de l'apprentissage automatique, la société devra accepter les compromis entre l'équité et l'exactitude des modèles et prendre des décisions à ce sujet [...]. En fait, ces compromis ont toujours été implicitement présents dans la prise de décisions de l'humain.

L'ère algorithmique axée sur les données vient de les mettre au premier plan. »³⁵ Il est essentiel d'aborder ces problèmes et de relier l'IA aux contextes sociaux et historiques, car l'IA devient de plus en plus omniprésente.

Qui décide? Diversité des équipes en IA

Une représentation accrue des groupes marginalisés dans le développement de l'IA pourrait mener à des résultats plus équitables en IA. Bien qu'il existe peu d'études sur les répercussions d'équipes diversifiées sur la création de produits plus équitables en particulier,³⁶ il existe dans divers secteurs des exemples d'équipes dirigées par des femmes, des personnes racisées et d'autres groupes marginalisés qui ont créé des produits et des services délibérément inclusifs. Par exemple, Fenty Beauty, une entreprise de cosmétiques fondée par la vedette de la pop barbadienne Rihanna, crée des teintures de maquillage pour les personnes ayant la peau plus foncée et qui ont souvent été exclues des gammes de cosmétiques. AccessNow, une application dont la fondatrice vit avec un handicap, indique aux utilisateurs l'accessibilité de différents endroits dans une source d'information centrale.³⁷ Dans le cas de l'IA, les produits qui ont été testés et créés par un groupe homogène peuvent logiquement ne pas tenir compte des besoins ou des perspectives des autres.³⁸ Si une équipe diversifiée sur le plan racial concevait un logiciel de reconnaissance faciale, on peut s'imaginer qu'elle serait susceptible de remarquer la possibilité d'une mauvaise classification fondée sur la race.

Partout dans le monde, les femmes sont sous représentées dans les domaines de l'informatique et du génie informatique. À l'échelle mondiale, seulement 22 % des professionnels en IA sont des femmes. Au Canada, malgré une concentration relativement élevée de professionnels en IA comparativement à d'autres pays, seulement 24 % sont des femmes. Les femmes représentent seulement 15 % du personnel de recherche en IA chez Facebook. Ce pourcentage descend à 10 % chez Google.⁴⁰ Les données montrent également que les diplômés en STGM (sciences, technologie, génie et mathématiques) faisant partie d'une minorité visible au Canada sont beaucoup moins susceptibles que les diplômés en STGM ne faisant pas partie d'une minorité visible de travailler dans le domaine des STGM.⁴¹ De plus, seulement 2,5 % des travailleurs à temps plein de Google sont noirs. Ce pourcentage grimpe à 4 % chez Microsoft.⁴²

Des recherches ont montré comment la ségrégation des professions selon le genre et la race empêchent des femmes et des personnes racisées d'occuper des emplois influents en technologie. Les femmes et les groupes racisés dans les professions des STGM ont déclaré se sentir constamment exclus de la culture organisationnelle et travailler dans des milieux qui n'offrent aucune flexibilité pour répondre à leurs besoins en ce qui concerne les personnes à leur charge.⁴³ Les milieux de travail dans les domaines des technologies peuvent aussi reléguer les femmes à des rôles axés sur les personnes plutôt qu'à des postes techniques, les détournant ainsi d'emplois qui auraient une incidence sur la façon de développer les technologies.⁴⁴ Les femmes et les groupes

racisés ont également déclaré faire face à des biais flagrants, à l'effacement et à la marginalisation au travail.⁴⁵ En décembre 2020, Mme Timnit Gebru, chercheuse scientifique de premier plan en IA chez Google, a fait les manchettes des médias lorsqu'elle a été congédiée en raison d'un article qu'elle a rédigé sur les risques et les méfaits des modèles langagiers (c. à d. l'IA alimentée par des données texte). Son congédiement a mis en lumière les tactiques abusives employées par l'entreprise, dont la manipulation, le rejet et le discrédit, soit des tactiques couramment utilisées contre les femmes noires qui veulent promouvoir la justice dans les domaines des technologies et dans toutes les sphères de la société.⁴⁶

Comment pouvons-nous éliminer les biais dans les données?

L'intelligence artificielle fonctionne par « apprentissage » à partir d'ensembles de données : des algorithmes sont créés pour extraire des données, les analyser, cerner les tendances et faire des prédictions. Les ensembles de données peuvent provenir de diverses sources, notamment des livres, des photos, des données sur la santé, des données d'organismes gouvernementaux ou des profils Facebook. Les biais sociétaux et les inégalités sont souvent intégrés à ces données et l'IA ne promulguera pas de valeurs sociales telles que l'équité à moins d'être directement programmée pour le faire. Ainsi, si un système d'embauche appuyé par l'IA est fondé sur les données des précédentes embauches et que peu de femmes ont été embauchées par le passé, l'algorithme perpétuera cette tendance.⁴⁷ De même, les groupes racisés et à faible revenu sont plus susceptibles de faire l'objet d'une surveillance, par exemple si leur quartier est surveillé étroitement par les policiers. Par conséquent, l'IA utilisée pour les services de police prédictifs sera plus susceptible de prédire la criminalité dans un secteur où il y a déjà eu un contrôle policier dans le passé que dans des secteurs où il n'y en a pas eu. Comme le font remarquer les chercheurs, « les données renforcent la marginalité [des gens] lorsqu'elles sont utilisées pour les cibler afin de les rendre suspects et de les soumettre à une surveillance accrue. Ces groupes considérés comme non méritants font l'objet de politiques publiques punitives et d'une surveillance plus intense, et le cycle recommence », ce qui crée selon leurs mots une « boucle d'injustice ».⁴⁸

Par ailleurs, les données peuvent aussi être biaisées en raison d'omissions. Par exemple, les données utilisées pour former des systèmes d'IA sur une langue peuvent provenir de textes publics gratuits, par exemple des livres qui n'entrent dans le domaine public que soixante dix ans après le décès de leur auteur. Étant donné que ce canon littéraire est basé sur des livres écrits principalement par des hommes blancs occidentaux, certaines perspectives et certains mots sont omis.⁴⁹ Un autre exemple : les systèmes d'IA formés pour reconnaître le genre ont tendance à avoir peu ou pas de données sur les personnes transgenres et non binaires, ce qui pourrait mener à des erreurs de classification du genre.⁵⁰ Les ensembles de données peuvent également omettre des populations entières qui n'ont jamais utilisé Internet ou été dans les médias sociaux ou qui n'ont jamais eu de carte de crédit ou de dossier de santé électronique, ce qui fausse les

résultats. Ces personnes omises sont souvent des personnes racisées, des personnes à faible revenu et d'autres personnes en marge.⁵¹

Assurer l'équité des données utilisées pour l'IA est un problème complexe compte tenu de la façon dont l'inégalité et l'iniquité influencent la vie des gens de façon complexe et chevauchante. Il n'est pas efficace de simplement supprimer des variables comme le genre et la race pour éviter la discrimination par les algorithmes, car les variables de substitution peuvent finir par créer les mêmes impacts.⁵² Récemment, la carte de crédit d'Apple a fait les manchettes parce que son algorithme semblait donner des marges de crédit plus petites aux femmes qu'aux hommes, même à celles qui étaient mariées et partageaient des actifs. Au départ, Apple et ses partenaires bancaires ont déclaré que les résultats ne pouvaient être biaisés parce que le genre n'était pas une variable dans l'algorithme et que la notation du crédit se faisait à l'insu du genre. En fin de compte, bien qu'une enquête sur la carte d'Apple ait conclu qu'elle ne discriminait pas les femmes, les experts ont noté que la création d'un algorithme sans distinction de genre n'empêcherait pas la discrimination fondée sur le genre de se produire par inadvertance.⁵³ Autre exemple : un outil qu'un service à la famille utilise pour analyser le risque de préjudice à l'égard des enfants ne tient pas compte de la race, mais d'autres variables incluses, comme les niveaux de pauvreté, sont fortement corrélées à la race. Le biais racial ne peut ainsi pas être éliminé même si la race n'est pas explicitement mesurée.^{54 55}

Comment pouvons-nous assurer la responsabilisation ?

Les recherches suggèrent qu'en général, il y a un manque de responsabilisation à l'égard des personnes qui subissent des préjudices à cause des systèmes d'IA. Autrement dit, la portée des répercussions de l'IA ainsi que la personne responsable de les créer et de les atténuer sont souvent nébuleuses. Cela suggère la nécessité d'effectuer davantage d'évaluations et de vérifications sur ce que les produits et services alimentés par l'IA signifient pour les gens, y compris des évaluations de leur équité.⁵⁶

Le premier défi en matière de responsabilisation est la transparence. Il y a souvent un manque de transparence concernant les objectifs des systèmes d'IA, leurs algorithmes et les données qui les alimentent. C'est ce qu'on appelle parfois le problème de la « boîte noire », car l'impénétrabilité de ces systèmes peut empêcher le public de comprendre les risques et les répercussions.⁵⁷ Si les gens ne savent pas comment les algorithmes sont utilisés à leur égard, il est alors impossible de remettre en question ou de modifier leurs prédictions et leurs décisions.⁵⁸ Même lorsque le secteur public utilise l'IA pour des processus aussi vastes que la surveillance et les décisions en matière d'immigration, le public n'en a peut-être pas connaissance ou ne possède pas les données. Par conséquent, certains chercheurs ont proposé une transparence complète de l'IA : les algorithmes ou les données et les résultats qu'ils visent à obtenir devraient être disponibles pour un examen public. Aider le public à comprendre comment les algorithmes et l'IA

influencent leur vie peut être une étape vers l'atténuation des résultats potentiellement préjudiciables.⁵⁹

Parallèlement, il y a des débats sur la mesure dans laquelle les systèmes d'IA peuvent faire preuve de transparence. Certains chercheurs suggèrent qu'exiger une telle transparence empêcherait les entreprises d'innover parce que la propriété intellectuelle ne serait pas protégée. De plus, étant donné que le code algorithmique est généralement incompréhensible pour le commun des mortels, la transparence ne peut pas nécessairement accroître la confiance des gens à l'égard de l'IA ni diminuer les préjudices de celle-ci.⁶⁰ Les logiciels sont également exclusifs et la transparence peut ne pas être possible pour des raisons de sécurité ou légales.⁶¹ Enfin, les gouvernements ou les entreprises qui utilisent des algorithmes pourraient ne pas vouloir les divulguer par crainte que les gens trouvent comment les contourner ou manipuler les résultats.⁶² Ainsi, la transparence au profit du public devra être équilibrée avec les avantages de la propriété intellectuelle et de l'innovation.

Le deuxième défi en matière de responsabilisation est l'absence de structures de gouvernance appropriées. Les chercheurs travaillent actuellement sur des structures de gouvernance et des procédures de vérification qui peuvent être mises en place au sein d'entreprises technologiques pour évaluer explicitement un système d'IA au chapitre des avantages sociaux et des valeurs.⁶³ Même si de nombreuses entreprises effectuent déjà des vérifications de leur IA, elles ne sont pas réglementées ni normalisées. Il est donc difficile pour les utilisateurs de s'assurer que les résultats des vérifications sont utilisés pour modifier les algorithmes.⁶⁴ En outre, les chercheurs tiers qui effectuent des vérifications doivent souvent faire face à de nombreux défis : des entreprises comme Google et Facebook créent des obstacles aux vérifications externes en interdisant la création de faux profils à des fins de recherche⁶⁵ et en empêchant la divulgation de données nécessaires. La fourniture d'un tel accès exige un équilibre entre la protection des renseignements personnels et la vérifiabilité. La vérification par des tiers est également coûteuse et nécessite beaucoup de temps et d'efforts.⁶⁶

Certains chercheurs ont proposé que l'adoption de mécanismes réglementaires pour veiller à ce que les entreprises et les gouvernements soient tenus responsables des répercussions inévitables et injustes. Si une loi ou une politique faisait en sorte que les créateurs et les propriétaires des algorithmes soient tenus directement responsables de ses résultats, elle pourrait contribuer à faire en sorte que l'IA soit conçue en tenant compte au préalable de ses répercussions sociales plutôt que dans le cadre d'efforts après sa mise en œuvre pour éliminer les préjudices après leur survenue.⁶⁷ Les politiques qui permettent au public et à la société civile d'intervenir dans le développement de l'IA et son utilisation dans le secteur public, par exemple au moyen de consultations auprès de personnes à l'extérieur du secteur de la technologie, peuvent également permettre une plus grande responsabilisation.⁶⁸

IA, automatisation et emploi

Outre les conversations sur les biais dans l'IA, il y a une discussion parallèle sur les effets potentiellement inéquitables de l'IA et de l'automatisation sur l'emploi. Il convient de souligner que l'IA et l'automatisation ne sont pas identiques : bien que l'automatisation ait tendance à désigner toutes les tâches effectuées par des machines, en particulier celles qui sont banales et répétitives, l'IA désigne spécifiquement le travail effectué par des machines qui imitent l'intelligence humaine et qui font des prédictions ou prennent des décisions. Toutefois, l'intelligence artificielle peut être une forme d'automatisation ou être directement utilisée dans celle-ci de sorte que les données probantes sur l'automatisation et l'iniquité socioéconomique sont importantes à ce chapitre.

Des recherches ont suggéré que l'automatisation pourrait entraîner la restructuration ou la suppression de nombreux emplois. Selon une étude, environ 47 % des emplois aux États-Unis présentent un risque d'automatisation⁶⁹ ; une autre estime que ce risque est de 42 % au Canada.⁷⁰ Des chercheurs ont également estimé qu'aux États-Unis, chaque « robot » supplémentaire (machine programmable entièrement autonome) par millier de travailleurs réduit le taux d'emploi de 0,2 % et les salaires de 0,42 %.⁷¹ En outre, les analyses suggèrent qu'au cours des 30 dernières années, la suppression d'emplois aux États-Unis en raison de l'automatisation s'est élevée à 16 %, alors qu'il y a eu une augmentation de seulement 10 % des nouvelles possibilités. Ces nouvelles possibilités profitaient souvent aux travailleurs hautement qualifiés, ce qui entraînait la perte d'emplois et la stagnation des salaires chez les travailleurs peu qualifiés.⁷² Les travailleurs dont les emplois ont tendance à être normalisés et comportent des tâches routinières, comme celles dans les usines, les magasins de vente au détail ou dans les bureaux, sont plus susceptibles d'être supprimés par la technologie que les emplois nécessitant une dextérité manuelle, des compétences technologiques ou un travail créatif ou émotionnel.⁷³

Certaines études suggèrent donc que les femmes et les groupes racisés seront relativement plus touchés en raison de leur forte présence dans des emplois et des industries spécifiques. Une étude récente émanant du Royaume-Uni révèle que les femmes occupent 70 % des emplois à risque élevé d'automatisation.⁷⁴ Les femmes sont surreprésentées en tant que caissières, secrétaires, commis à la tenue de livres, réceptionnistes et comptables et parmi d'autres professions à risque élevé d'automatisation. De plus, les femmes sont moins susceptibles que les hommes d'occuper les emplois bien rémunérés que nécessitera l'automatisation accrue, par exemple ceux d'informaticiens. La race et l'indigénité peuvent aussi avoir un impact. Au Canada, environ 250 000 emplois occupés par des Autochtones sont à risque d'automatisation. L'emploi chez les Autochtones est plus concentré dans des industries comme l'hébergement et les services de restauration, le commerce de détail, la construction et le transport, par comparaison à l'emploi chez les personnes qui ne sont pas Autochtones.⁷⁵ De même, une étude récente émanant des États-Unis révèle que 31 % des travailleurs latino-américains et 27 % des travailleurs noirs se retrouvent dans 30 professions à risque élevé d'automatisation par comparaison à 24 % des

travailleurs blancs.⁷⁶ L'automatisation du travail pourrait donc désavantager particulièrement les groupes qui font déjà face à la pauvreté et à la marginalisation à des taux disproportionnés et exacerber l'iniquité économique et sociale existante.⁷⁷

Par ailleurs, l'automatisation peut entraîner une croissance de l'emploi ou compléter le travail de diverses façons. Par le passé, l'automatisation dans les industries du textile, de l'acier et de l'automobile a entraîné une augmentation des emplois dans ces secteurs en réduisant les coûts et en augmentant ainsi la demande.⁷⁸ Toutefois, les emplois n'augmentent habituellement pas aussi rapidement que les revenus. Dans tous les secteurs, les gens travaillent déjà avec des machines automatisées, notamment dans le domaine de la santé, de l'éducation et du droit.⁷⁹ La question de savoir si des emplois seront remplacés par l'automatisation est quelque peu difficile à prédire : si l'efficacité accrue créée par la technologie peut réduire le nombre d'emplois, la croissance des ventes découlant de cette efficacité accrue peut accroître la demande d'emplois. On peut le constater dans le cas d'Amazon. Elle utilise de plus en plus de robots dans ses centres de gestion des commandes pour augmenter l'efficacité du personnel, mais elle embauche toujours des milliers de nouveaux employés pour répondre à l'augmentation de la demande. Les chercheurs ont également montré que les plateformes numériques alimentées par l'IA peuvent accroître le bassin d'employeurs et de travailleurs en éliminant les obstacles qui empêchent les employeurs de trouver des travailleurs, en réduisant les coûts des transactions et en améliorant le jumelage entre employeurs et employés.⁸⁰

Ce qui est nouveau avec l'IA prédictive, c'est qu'elle peut de plus en plus remplacer de nombreux emplois de cols blancs qui nécessitent de faire preuve de jugement et d'effectuer des prévisions, notamment les personnes dans le domaine immobilier, les conseillers en investissements, les professionnels du droit et les développeurs de logiciels. En effet, les compétences requises pour faire des tâches comme l'examen de documents, la lecture, l'écriture, le codage et même l'enseignement sont de plus en plus couramment automatisées.⁸¹ ⁸² Les taxis et d'autres services de conduite - un secteur dominé par les hommes - ont aussi été « déconstruits » et transformés par des entreprises technologiques comme Uber et Lyft. Les répercussions sont ambiguës, car cette substitution de tâches comportant des prédictions peut également accroître le besoin de main-d'œuvre pour réaliser des tâches complémentaires qui sont en amont ou en aval de la tâche automatisée. Par exemple, l'utilisation de l'automatisation en radiologie pour interpréter les résultats d'imagerie accélère cette tâche, mais elle peut également entraîner un besoin accru d'effectifs pour communiquer les résultats et décider des mesures à prendre à la suite de la prédiction faite à l'aide de l'IA.⁸³

Des recherches portent également sur la façon dont des employeurs utilisent les algorithmes pour contrôler et diriger les travailleurs. Cela a des répercussions sur la façon dont le travail est effectué et sur les nouveaux types de main-d'œuvre. Des employeurs ont utilisé des algorithmes pour consigner le comportement des travailleurs en temps réel, comme calculer le temps qu'ils prennent pour réaliser les tâches ou surveiller

la façon dont ils communiquent. Ils peuvent également servir à mesurer la productivité, à prévoir et à évaluer le rendement des travailleurs, à formuler des recommandations à l'intention des employés lorsqu'ils font leurs tâches, etc. Un bon exemple est Uber, qui utilise des algorithmes pour suggérer aux conducteurs de se reposer si leur conduite semble erratique ou les retirer de la plateforme s'ils reçoivent de nombreuses notes faibles. Cette utilisation d'algorithmes peut mener à de nouvelles professions, notamment des analystes pour passer en revue les nouvelles données recueillies par algorithmes, ainsi que des ingénieurs, des développeurs et des employés de soutien à la technologie pour concevoir et gérer des algorithmes. Simultanément, cette utilisation d'algorithmes peut avoir des répercussions négatives sur les travailleurs, y compris une atteinte à la vie privée, un stress et une frustration accrus et une méconnaissance des raisons du congédiement ou de la perte de salaire.⁸⁴

Il sera donc important d'étudier comment ces nouvelles technologies touchent diverses professions, ainsi que d'assurer l'adaptabilité des travailleurs aux nouvelles normes et la capacité des gens à accéder à de nouvelles compétences et à de la formation.⁸⁵ C'est à dire que la main-d'œuvre doit posséder les compétences nécessaires pour prendre en charge un travail nouveau et différent créé par l'automatisation, car ceux qui n'ont pas les compétences nécessaires pour s'adapter aux changements technologiques seront peut-être les plus touchés.⁸⁶

Ces changements en IA et en automatisation recourent maintenant l'impact de la pandémie de COVID 19. Ce sont surtout les femmes, les groupes racisés et les groupes à faible revenu qui ont perdu leur emploi et souffert de l'instabilité économique découlant des retombées de la pandémie de COVID 19 au Canada.^{87 88} Les responsabilités de proche aidant ont nui davantage à la participation des femmes au marché du travail depuis le début de la pandémie et peuvent nuire à leur capacité de se recycler ou de se perfectionner pour s'adapter aux nouvelles technologies.⁸⁹ Les décideurs et les entreprises ont donc l'occasion de se concentrer sur des stratégies et des politiques qui garantiront un emploi rémunérateur et décent pour tous, surtout dans le contexte d'une économie numérique en évolution rapide.⁹⁰

Répercussions sur la recherche, les politiques et la pratique

Compte tenu de ces considérations, la recherche, les politiques et les pratiques peuvent être mobilisées pour comprendre et contrer les répercussions de l'IA et s'assurer qu'elle est mise en œuvre de façon équitable et juste.

Réglementation et politiques

Il est largement reconnu que les gouvernements ont du rattrapage à faire pour veiller à ce que l'IA soit développée et utilisée de manière à réduire les préjudices pour les groupes marginalisés. Une réglementation accrue n'est pas une solution fourre tout parce qu'elle peut souvent prendre du retard sur le rythme rapide de développement de l'IA. Certains chercheurs ont donc suggéré que

la réglementation sera efficace seulement si les développeurs et d'autres acteurs de l'industrie de la technologie font des efforts concertés pour réduire l'iniquité et l'injustice lorsqu'ils créent des algorithmes.⁹¹ Néanmoins, de nouvelles lois ou politiques encadrant l'IA pourraient aider à protéger les gens et à bâtir un avenir plus inclusif, où l'IA peut « tirer parti des forces humaines » et compléter les humains plutôt que de chercher à les remplacer.⁹²

Pour établir une plus grande responsabilisation, de nouvelles politiques ou lois pourraient faire en sorte de clarifier qui crée, qui possède et qui contrôle l'IA, attribuant ainsi une responsabilité là où il y en a actuellement peu.⁹³ D'autres ont suggéré que les vérifications et les évaluations des répercussions de l'IA devraient être obligatoires et effectuées avant et pendant la mise en œuvre de l'IA.⁹⁴ De plus, bien qu'il y ait des débats au sujet de la transparence, des processus normalisés « d'explicabilité » peuvent encore être mis en place afin que les entreprises justifient les décisions prises au sujet de l'IA et par l'entremise de celle-ci (y compris son objectif, sa conception et ses ensembles de données) et divulguent les risques, par exemple dans les documents publics et les rapports publiés.^{95 96} Ensemble, ces idées suggèrent la nécessité d'une surveillance plus rigoureuse de l'IA. Les gouvernements doivent créer des organismes de réglementation ou des cadres qui tiennent compte des contextes sociaux et historiques de l'IA et réglementent de façon exhaustive les pratiques.⁹⁷

Dans de nombreux pays, grâce au travail d'organisations axées sur l'éthique de l'IA, de tels cadres et initiatives sont déjà mis en œuvre. Aux États Unis, une loi sur la responsabilisation algorithmique (Algorithmic Accountability Act) a été introduite en 2019 pour amener les grandes entreprises à évaluer leurs algorithmes afin de détecter les biais et les risques pour les utilisateurs.⁹⁸ En 2021, l'UE a créé une proposition de loi sur l'intelligence artificielle, soit « le tout premier cadre juridique sur l'IA ».⁹⁹ De plus, certains chercheurs et défenseurs ont recommandé l'adoption d'un cadre global exhaustif qui régira largement l'utilisation de l'IA comme la déclaration universelle des droits de la personne des Nations Unies.¹⁰⁰ La Déclaration de Toronto de 2018, dirigée par Amnistie internationale et Access Now, est un exemple d'une déclaration exhaustive d'appels à l'action pour défendre les droits de la personne en IA.¹⁰¹

Le Canada est en train d'élaborer ses propres politiques et cadres. À la suite d'un investissement de 125 M\$ dans une stratégie pancanadienne sur l'intelligence artificielle en 2017, le gouvernement fédéral a élaboré une directive sur la prise de décisions automatisée et une évaluation publique des répercussions des algorithmes.¹⁰² Il a également créé le Conseil consultatif en matière d'intelligence artificielle en 2019, bien que ce conseil ait été particulièrement critiqué pour son manque de représentation des groupes racisés et marginalisés.¹⁰³ Le Commissariat à la protection de la vie privée du Canada a également récemment formulé des recommandations sur la mise à jour de la Loi sur la protection des renseignements personnels et les documents électroniques afin de mieux réglementer l'IA.¹⁰⁴ En Ontario, au moment de la rédaction du présent rapport, des consultations publiques sont en cours pour créer un cadre de l'intelligence artificielle de confiance.¹⁰⁵ Les effets de ces efforts stratégiques sur l'équité, la transparence et la responsabilisation restent à voir. Il convient toutefois de noter qu'en date de juin 2021, il n'y a toujours aucun projet de loi sur la gouvernance de l'IA au Canada.¹⁰⁶

Normes de l'industrie

Au-delà de la réglementation, les chercheurs et les défenseurs s'emploient à conceptualiser une IA qui met l'accent sur des considérations sociales plutôt que la découverte et la résolution de problèmes après coup. En effet, l'IA peut être développée et créée de façon à s'aligner sur les objectifs de réduction des inégalités et des iniquités systémiques, mais comme il est mentionné précédemment, ce n'est pas une mince tâche de programmer dans un système d'IA des normes et des valeurs complexes que les humains comprennent lorsqu'ils font des prédictions et prennent des décisions.¹⁰⁷ Une autre question se pose : si un tel système d'IA n'a pas encore été développé de façon efficace, quelles sont les circonstances dans lesquelles l'IA ne devrait pas être utilisée et quelles sont les meilleures solutions?

Cela devient aussi une question morale comportant des compromis et des valeurs. L'alignement délibéré de l'IA avec les valeurs sociales signifie que les entreprises devront prioriser l'équité et d'autres considérations sociales plutôt que le profit ou l'efficacité. Un tel changement peut nécessiter beaucoup de temps et d'argent, par exemple les coûts de la recherche sur les répercussions sociales ou les pertes de revenus potentielles découlant de l'abandon de la mise en œuvre d'un nouveau système d'IA pour des raisons éthiques. Il faut donc une coopération de l'industrie et une action collective comprenant l'établissement de normes afin que les systèmes d'IA sécuritaires et responsables soient acceptés en tant que norme.¹⁰⁸

Représentation

Si le manque de représentation des communautés marginalisées dans le développement de la technologie nuit à la création de systèmes d'IA inclusifs et équitables, ce problème pourrait être atténué par des embauches et des promotions plus équitables. De nombreuses études ont été réalisées sur des solutions pour rendre les milieux de travail, y compris les entreprises du secteur de la technologie, plus inclusifs à l'égard des femmes et des groupes racisés. Il s'agit notamment de faire preuve de plus de souplesse à l'égard des travailleurs qui doivent prioriser la prestation de soins (habituellement des femmes); de transformer des pratiques d'embauche et de recrutement non inclusives qui favorisent certains candidats (comme les jeunes hommes ou les personnes qui ont suivi une formation dans des écoles d'élite) et de travailler à l'instauration de politiques et d'une culture antiracistes et antisexistes.¹⁰⁹ Les écoles qui enseignent les STGM peuvent également transformer utilement leur culture, car des études ont montré que les jeunes femmes peuvent être mal traitées par leurs camarades masculins en génie et avoir l'impression de ne pas y avoir leur place.¹¹⁰

Une autre possibilité qui permettrait aux équipes d'être mieux outillées pour régler les problèmes d'équité serait d'assurer la participation de différentes disciplines au développement de l'IA. Des chercheurs ont suggéré que l'embauche de chercheurs spécialisés en sciences humaines et en sciences sociales qui ont une compréhension approfondie des inégalités socioéconomiques, de l'histoire et de la théorie critique touchant le genre, la race et d'autres identités sociales pourrait aider les entreprises à cerner et à résoudre les problèmes d'éthique liés à la mise en œuvre de l'IA.¹¹¹

Recyclage et perfectionnement des compétences

La pandémie de COVID 19 a entraîné la décimation des industries où les contacts sont fréquents comme le tourisme, l'hôtellerie, la restauration et le commerce de détail. Ce sont aussi des industries qui emploient des femmes de façon disproportionnée. Parallèlement, certaines recherches suggèrent que les emplois où les contacts sont fréquents et qui sont plus susceptibles d'être remplacés par l'automatisation ont tendance à être occupés par des femmes ainsi que par des groupes racisés, autochtones et à faible revenu. Mettre l'accent sur l'acquisition et le perfectionnement de compétences à l'échelle des entreprises et des gouvernements pourrait régler ces problèmes parce que certains groupes ne seraient pas laissés pour compte.¹¹² Ces initiatives pourraient comprendre la conception de nouvelles possibilités de formation; des partenariats entre des gouvernements, des entreprises et des établissements postsecondaires afin d'assurer un accès équitable aux programmes de recyclage ou de perfectionnement des compétences et un soutien gouvernemental et organisationnel à l'égard d'initiatives et d'organisations qui essaient d'intégrer des groupes en quête d'équité dans les STGM.¹¹³ Elles pourraient également inclure un solide financement dans le secteur des soins, c'est à dire les services de garde d'enfants et les soins aux aînés. Cela pourrait créer des centaines de milliers de nouveaux emplois peu sujets à l'automatisation, notamment pour les femmes, et faciliterait les services dont les familles canadiennes ont grand besoin compte tenu de la pandémie.¹¹⁴

Vers l'équité en IA

Les recherches actuelles sur l'IA suggèrent qu'il s'agit d'un moment important pour les dirigeants, les décideurs et les chercheurs de prévenir le renforcement des inégalités et des iniquités par le truchement de la technologie. Beaucoup d'efforts sont déjà déployés par des entreprises partout au pays et dans le monde pour promouvoir l'équité en IA. Dans le milieu universitaire, il existe plusieurs pistes de recherche, dont certaines sont énumérées ci-dessous:

- Réaliser des travaux interdisciplinaires sur l'IA – Les chercheurs techniques en IA peuvent collaborer avec des chercheurs en sciences sociales et en sciences humaines pour mieux comprendre les répercussions de l'IA sur les groupes confrontés à la marginalisation, y compris à l'échelle mondiale, ainsi que la façon dont il est possible de les modifier. Des liens plus solides peuvent également être établis entre les groupes marginalisés et les personnes qui créent la technologie qui les touche.^{115 116}
- Créer une nouvelle IA conforme aux valeurs sociales – Des réflexions et des recherches novatrices se poursuivent afin de comprendre comment mettre en œuvre des algorithmes et des systèmes d'IA qui sont alignés sur des valeurs comme l'équité et qui permettent de corriger les biais et d'autres préjudices. Des chercheurs ont exploré comment tout miser sur les profits lorsqu'il est question d'IA contribue au renforcement de la marginalisation. Des recherches ultérieures pourraient contribuer à élaborer d'autres solutions axées sur d'autres considérations.¹¹⁷

- Étudier comment la réglementation et les politiques peuvent mieux atténuer les répercussions préjudiciables – Il est possible de poursuivre les recherches sur les méthodes optimales de réglementation de l'IA (c. à d. les lois et les politiques) et de s'assurer que ces règlements gèrent la tension potentielle entre l'atténuation et la marginalisation, tout en n'étouffant pas l'innovation.

L'IA a changé l'économie et la société. Bien qu'elle ait le potentiel d'améliorer de nombreuses vies, elle peut entraîner des préjudices importants. Étant donné que la technologie est créée à l'intérieur de contextes et d'histoires d'inégalité et de pouvoir, il est facile, quoique peut-être involontaire, de les reproduire dans les systèmes d'IA. De nombreux exemples ont montré que l'IA appliquée sans attention à de tels contextes peut renforcer la discrimination et les préjugés à l'égard des femmes, des communautés racisées et d'autres personnes aux prises avec des iniquités et des inégalités. L'IA a aussi le potentiel de placer les communautés marginalisées davantage à risque du point de vue socioéconomique en supprimant ou en restructurant des emplois, surtout en considérant que la stabilité économique de nombreux moyens de subsistance a déjà été mise à mal par la COVID 19. Ces constatations suggèrent la nécessité de travaux continus dans les domaines suivants pour les entreprises, les décideurs et les chercheurs:

- Les entreprises technologiques et les gouvernements peuvent se concentrer sur des initiatives visant à assurer une représentation équitable dans le développement de l'IA.
- Les créateurs, les chercheurs et les responsables de la mise en œuvre de l'IA peuvent prioriser l'alignement de l'IA sur des valeurs sociales comme l'impartialité et l'équité, malgré les compromis liés à l'efficacité et aux profits.
- Les gouvernements peuvent créer des politiques et des lois sur l'IA qui accordent la priorité à la responsabilisation et à la transparence et exigent l'adhésion des entreprises technologiques à ces principes.
- Les gouvernements et les entreprises peuvent travailler de sorte à contribuer à la sécurité économique des travailleurs qui sont doublement touchés par les nouvelles technologies et la pandémie mondiale en portant une attention aux programmes de recyclage ou de perfectionnement des compétences.
- Les chercheurs universitaires peuvent approfondir leurs connaissances sur l'IA et l'iniquité, notamment en poursuivant des travaux interdisciplinaires sur les répercussions sociales, politiques et environnementales de l'IA et en concevant des solutions de rechange nouvelles et différentes qui priorisent l'atténuation des préjudices.

Les recherches suggèrent que la prévention du renforcement de l'iniquité par l'IA nécessite un travail intersectoriel et interdisciplinaire, des gouvernements aux universités, en passant par les entreprises. Autrement dit, les solutions nécessiteront une combinaison de règlements et de politiques, de nouvelles activités de recherche et de développement pour accroître l'équité de l'IA, un changement de normes en ce qui concerne les personnes qui élaborent et prennent des décisions au sujet de l'IA et une responsabilisation à l'égard des personnes les plus touchées. Sans efforts concertés, le renforcement des préjugés et de la discrimination systémiques se perpétuera par le truchement de ces systèmes technologiques qui deviennent omniprésents.

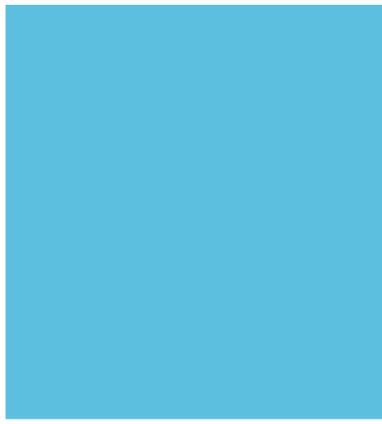
Notes

1. Russell, S.J. et P. Norvig. « Artificial Intelligence: A Modern Approach », Pearson, 2013.
2. Stone, P., R. Brooks, E. Brynjolfsson, R. Calo, O. Etzioni, G. Hager, J. Hirschberg, S. Kalyan Krishnan, E. Kamar, S. Kraus, K. Leyton-Brown, D. Parkes, W. Press, A. Saxenian, J. Shah, M. Tambe et A. Teller. « Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence », Université de Stanford, 2016. Consulté le 19 mars 2020 à https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fnl_singles.pdf
3. Jaume-Palasi, L. « Why We Are Failing to Understand the Societal Impact of Artificial Intelligence », *Social Research*, vol. 86, no 2, p. 477-498, 2019.
4. Birhane, A. et J. van Dijk. « Robot Rights?: Let's Talk about Human Welfare Instead », travaux de la conférence de l'AAAI/ACM sur l'IA, l'éthique et la société, 2020.
5. Kabeer, N., S. Razavi et Y. van der Meulen Rodgers. « Feminist Economic Perspectives on the COVID-19 Pandemic », *Feminist Economics*, 2021. Consulté le 2 mars 2021 à <https://doi.org/10.1080/13545701.2021.1876906>
6. Cet aperçu de la recherche est fondé sur un examen de la recherche universitaire existante et d'autres rapports portant sur les questions suivantes : Quelles sont les répercussions de l'IA sur les groupes qui font face à la marginalisation et à l'iniquité et pourquoi ces répercussions se produisent-elles? Quel est l'avenir de la main-d'œuvre canadienne dans le contexte de l'IA et de l'automatisation? Comment notre société peut-elle créer efficacement une économie inclusive et équitable de concert avec l'intensification de l'IA? Quelles sont les lacunes en matière de recherche et comment peut-on les combler pour créer une IA plus équitable? La présente synthèse vise à étudier ces questions d'une manière générative et informative plutôt qu'exhaustive. Étant donné que de nombreux groupes hors du milieu universitaire travaillent sur les répercussions éthiques et sociétales de l'IA, les auteurs sont allés au-delà des publications universitaires pour inclure des rapports stratégiques et d'autres articles produits par des chercheurs et des défenseurs de l'IA équitable. Pour cerner la documentation à examiner et à résumer, nous avons cherché des recherches et des documents dans des bases de données comme JStor, EBSCOHost, Google Scholar et Recherche Google et nous avons examiné des citations pour trouver d'autres sources. Les termes de recherche comprenaient les divers termes utilisés pour décrire l'IA et ses processus, notamment « intelligence artificielle », « apprentissage automatique » et « algorithmes », ainsi que des termes qui traiteraient de ses répercussions sur l'égalité et l'équité, notamment « biais », « genre », « race », « stéréotypes », « inégalité » et « emploi ». Comme il s'agit d'un domaine de recherche en rapide évolution, nous avons limité les recherches aux cinq dernières années. Nous avons également consulté un certain nombre d'experts universitaires sur les répercussions sociales de l'IA afin de trouver d'autres sources de recherche et d'obtenir des commentaires sur les versions antérieures.
7. Sunstein, C.R. « Algorithms, Correcting Biases », *Social Research*, vol. 86, no 2, p. 499-511, 2019.
8. Kleinberg, J., J. Ludwig, S. Mullainathan et C.R. Sunstein. « Discrimination in the Age of Algorithms », *Journal of Legal Analysis*, vol. 10, p. 113-174, 2018.
9. Sunstein, C.R. « Algorithms, Correcting Biases », *Social Research*, vol. 86, no 2, p. 499-511, 2019.
10. Hurley, D. « Can an Algorithm Tell When Kids Are in Danger? », *New York Times Magazine*, 2018. Consulté le 28 avril 2021 à www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html
11. Cameron, H.E., A. Goldfarb et L. Morris. « Artificial intelligence for a reduction of false denials in refugee claims », document de travail, 2021.
12. Eubanks, V. « Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor », St. Martin's Press, 2018.
13. Angwin, J., J. Larson, S. Mattu et L. Kirchner. « Machine Bias », *ProPublica*, 2016. Consulté le 28 avril 2021 à www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
14. Albright, A. « If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions » https://thelittledataset.com/about_files/albright_judge_score.pdf
15. Birhane, A. et J. van Dijk. « Robot Rights?: Let's Talk about Human Welfare Instead », travaux de la conférence de l'AAAI/ACM sur l'IA, l'éthique et la société, 2020.
16. Garcia, M. « Racist in the Machine: The Disturbing Implications of Algorithmic Bias », *World Policy Journal*, 2016-2017.
17. Dastin, J. « Amazon scraps secret AI recruiting tool that showed bias against women », Reuters, 2018. Consulté le 23 mars 2020 à www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scrap-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G
18. Noor, P. « Can we trust AI not to further embed racial bias and prejudice? », *BMJ*, 368, 2020.
19. Buolamwini, J. et T. Gebru. « Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification », *Proceedings of Machine Learning Research*, vol. 81, p. 115, 2018.
20. Richardson, R., J.M. Schultz et K. Crawford. « Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice », *New York University Law Review Online*, vol. 94, no 15, 2019.
21. Roberson, K., C. Khoo et Y. Song. « To Surveil and Predict: A Human Rights Analysis of Algorithmic Policing in Canada », The Citizen Lab, Université de Toronto, 2020. Consulté le 12 février 2021 à <https://citizenlab.ca/wp-content/uploads/2020/09/To-Surveil-and-Predict.pdf>
22. Agüera y Arcas, B., M. Mitchell et A. Todorov. « Physiognomy's New Clothes », 2017. Consulté le 18 février 2021 à <https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a>
23. Cave, S. et K. Dihal. « The Whiteness of AI », *Philosophy & Technology*, vol. 33, p. 685-703, 2020.
24. Schiller, A. et J. McMahon. « Alexa, Alert Me When the Revolution Comes: Gender, Affect, and Labor in the Age of Home-Based Artificial Intelligence », *New Political Science*, vol. 41, no 2, *New Political Science*
25. Costa, P. « Conversing with Personal Digital Assistants: on Gender and Artificial Intelligence », *Journal of Science and Technology of the Arts*, vol. 10, no 3, p. 59-72, 2018.
26. Imana, B., A. Korlova et J. Heidemann. « Auditing for discrimination in Algorithms Delivering Job Ads », travaux de la conférence Web 2021, 19 au 23 avril 2021. Consulté le 5 mai 2021 à <https://ant.isi.edu/datasets/addelivery/Discrimination-Job-Ad-Delivery.pdf>
27. Hao, K. « Facebook's ad algorithms are still excluding women from seeing jobs », *MIT Technology Review*, 2021. Consulté

- le 1er mai 2021 à www.technologyreview.com/2021/04/09/1022217/facebook-ad-algorithm-sex-discrimination/
28. Noble, S.U. « Algorithms of oppression: How search engines reinforce racism », New York University Press, New York, 2018.
 29. Crawford, K. « The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence », Yale University Press, 2021.
 30. Crawford, K. et V. Joler. « Anatomy of an AI System », 2018. Consulté le 1er mai 2021 à <https://anatomyof.ai/>
 31. Molnar, P. et L. Gill. « Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System », The Citizen Lab, Université de Toronto, 2018. Consulté le 4 avril 2020 à <https://ihrp.law.utoronto.ca/sites/default/files/media/IHRP-Automated-Systems-Report-Web.pdf>
 32. Hadfield-Menell, D.H. et G.K. Hadfield. « Incomplete Contracting and AI Alignment », travaux de la conférence de l'AAAI/ACM sur l'IA, l'éthique et la société, 2019. Consulté le 20 mars 2021 à <https://dl.acm.org/doi/10.1145/3306618.3314250>
 33. Wellner, G. et T. Rothman. « Feminist AI: Can We Expect Our AI Systems to Become Feminist? », *Philosophy and Technology*, vol. 33, p. 191-205, 2020.
 34. Kearns, M. et A. Roth. « The Ethical Algorithm: The Science of Socially Aware Algorithm Design », Oxford University Press, 2020.
 35. Kearns, M. et A. Roth. « The Ethical Algorithm: The Science of Socially Aware Algorithm Design », Oxford University Press, 2020.
 36. La recherche sur les effets de l'utilisation d'équipes de travail diversifiées sur leurs produits et processus incite à penser qu'ils dépendent des circonstances, par exemple des types de tâches confiées aux équipes. Certaines études suggèrent que les équipes diversifiées sur le plan culturel ou racial parviennent à des niveaux plus élevés d'innovation dans les produits et affichent une performance et une réflexion plus créatives. D'autres études ont constaté que cela peut entraîner des coûts de communication ou des conflits qui réduisent la capacité d'innovation des équipes.
 37. Voir « Creating a More Accessible World » : Séance de questions et réponses avec Maayan Ziv, fondatrice et chef de la direction, AccessNow, 2018. [<https://news.microsoft.com/en-ca/2018/12/03/creating-a-more-accessible-world-qa-with-maayan-ziv-founder-and-ceo-accessnow/>]; « Highlighting Fenty Beauty's Diversity-Based Business Model », 2020. [www.gendereconomy.org/fenty-beauty-diversity-based-business-model/] et « Backstage Capital: Investing in Black women founders », 2020. [www.gendereconomy.org/backstage-capital-investing-in-black-women-founders/].
 38. Gebru, T. « The Oxford Handbook of Ethics of AI », chapitre sur la race et le genre (dir. M.D. Dubber, F. Pasquale et S. Das), 2020. [<https://arxiv.org/abs/1908.06165>]
 39. World Economic Forum (2018). « The Global Gender Gap Report », 2018. Consulté le 4 mars 2021 à www3.weforum.org/docs/WEF_GGGR_2018.pdf
 40. West, S.M., M. Whittaker et K. Crawford. « Discriminating Systems: Gender, Race and Power in AI », AI Now Institute, 2019. Consulté le 24 mars 2020 à <https://ainowinstitute.org/discriminatingystems.pdf>
 41. Frank, K. Parcours professionnel des titulaires d'un diplôme en STGM au Canada : analyse comparative entre les sexes, Statistique Canada, 2019. Consulté le 19 mars 2020 à www150.statcan.gc.ca/n1/pub/11f0019m/11f0019m2019017-fra.htm#a4_2
 42. West, S.M., M. Whittaker et K. Crawford. « Discriminating Systems: Gender, Race and Power in AI », AI Now Institute, 2019. Consulté le 24 mars 2020 à <https://ainowinstitute.org/discriminatingystems.pdf>
 43. Seron, C., S. Silbey, E. Cech et B. Rubineau. « I am Not a Feminist, but...: Hegemony of a Meritocratic Ideology and the Limits of Critique Among Women in Engineering », *Work and Occupations*, vol. 45, no 2, 2018.
 44. Alegria, S. « Escalator or Step Stool? Gendered Labor and Token Processes in Tech Work », *Gender & Society*, vol. 33, no 5, p. 722-745, 2019.
 45. Seron, C., S. Silbey, E. Cech et B. Rubineau. « Persistence is Cultural: Professional Socialization and the Reproduction of Sex Segregation », *Work and Occupations*, vol. 43, no 2, 2016.
 46. Turner, K., D. Wood et C. D'Ignazio. « The Abuse and Misogynoir Playbook », *The State of AI Ethics*, Montréal, AI Ethics Institute, 2021. Consulté le 2 février 2021 à <https://montrealaiethics.ai/wp-content/uploads/2021/01/The-State-of-AI-Ethics-Report-January-2021.pdf>
 47. Williams, B.A., C.F. Brooks et Y. Shmargad. « How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications », *Journal of Information Policy*, vol. 8, p. 78-115, 2018.
 48. Eubanks, V. « Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor », St. Martin's Press, 2018.
 49. Levendowski, A. « How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem », *Washington Law Review*, p. 579, 2018.
 50. Gebru, T. « The Oxford Handbook of Ethics of AI », chapitre sur la race et le genre (dir. M.D. Dubber, F. Pasquale et S. Das), 2020. [<https://arxiv.org/abs/1908.06165>]
 51. Katal, S.K. « Artificial Intelligence, Advertising, and Disinformation », *Advertising and Society Quarterly*, vol. 20, no 4, 2019.
 52. Wellner, G. et T. Rothman. « Feminist AI: Can We Expect Our AI Systems to Become Feminist? », *Philosophy & Technology*, vol. 33, p. 191-205, 2020.
 53. Knight, W. « The Apple Card Didn't 'See' Gender—and That's the Problem », *Wired*, 19 novembre 2019. Consulté le 29 juin 2021 à www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/
 54. Hirschman, D. et E.A. Bosk. « Standardizing Biases: Selection Devices and the Quantification of Race », *Sociology of Race and Ethnicity*, vol. 115, no 6, p. 974-1001, 2019.
 55. Kearns, M. et A. Roth. « The Ethical Algorithm: The Science of Socially Aware Algorithm Design », Oxford University Press, 2020.
 56. Donahoe, E. et M. MacDuffee Metzger. « Artificial Intelligence and Human Rights », *Journal of Democracy*, vol. 30, no 2, 2019.
 57. Coglianese, C. et D. Lehr. « Transparency and Algorithmic Governance », *bourse d'études, Penn Law*, 2123.
 58. O'Neil, C. « Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy », Crown, 2016.
 59. Wellner, G. et T. Rothman. « Feminist AI: Can We Expect Our AI Systems to Become Feminist? », *Philosophy & Technology*, vol. 33, p. 191-205, 2020.
 60. De Fine Licht, K. et J. de Fine Licht. « Artificial intelligence, transparency, and public decision-making », *AI & Society*, 2020.
 61. Askeff, A., M. Brundage et G. Hadfield. « The Role of Cooperation in Responsible AI Development », document de travail, 2019. Consulté le 24 mars 2021 à <https://arxiv.org/abs/1907.04534>
 62. Brauneis, R. et E.P. Goodman. « Algorithmic Transparency for the Smart City », *Yale J.L. & Tech*, 2018.
 63. Raji, I.D., A. Smart, R.N. White, M. Mitchell, T. Gebru, B. Hutchinson, J. Smith-Loud, D. Theron et P. Barnes. « Closing the AI Accountability Gap », travaux de la conférence de 2020 sur l'équité, la responsabilisation et la transparence, p. 33-44, 2020.
 64. Ng, A. « Can Auditing Eliminate Bias from Algorithms? », *The Markup*, 2021. Consulté le 2 mars 2021 à <https://themarkup.org/ask-the-markup/2021/02/23/can-auditing-eliminate-bias-from-algorithms>
 65. O'Neil, C. « Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy », Crown, 2016.

66. Imana, B., A. Korlova et J. Heidemann. « Auditing for discrimination in Algorithms Delivering Job Ads », travaux de la conférence Web 2021, 19 au 23 avril 2021. Consulté le 5 mai 2021 à <https://ant.isi.edu/datasets/addelivery/Discrimination-Job-Ad-Delivery.pdf>
67. Pasquale, F. « New Laws of Robotics: Defending human expertise in the age of AI », Belknap Press, 2020.
68. McKelvey, F. et M. MacDonald. « Artificial Intelligence Policy Innovations at the Canadian Federal Government », Canadian Journal of Communication Policy Portal, vol. 44, p. 43 50, 2019.
69. Frey, C.B. et M.A. Osborne. « The future of employment: How susceptible are jobs to computerisation? », Technological Forecasting and Social Change, vol. 114, p. 254 280, 2017.
70. Lamb, C. « The Talented Mr. Robot: The impact of automation on Canada's workforce », Brookfield Institute for Innovation + Entrepreneurship, 2016. Consulté le 23 mars 2021 à https://brookfieldinstitute.ca/wp-content/uploads/TalentedMrRobot_BIIE-1.pdf
71. Acemoglu, D. et P. Restrepo. « Robots and Jobs: Evidence from US Labor Markets », Journal of Political Economy, vol. 128, no 6, p. 2188 2244, 2020.
72. Acemoglu, D. et P. Restrepo. « Unpacking Skill Bias: Automation and New Tasks », AEA Papers and Proceedings, no 110, p. 356 361, 2020.
73. Dahlin, E. « Are Robots Stealing Our Jobs? », Socius: Sociological Research for a Dynamic World, no 5 (1 14), 2019.
74. Roberts, C., H. Parkes, R. Statham et L. Rankin. « The Future is Ours: Women, Automation and Equality in the Digital Age », The Centre for Economic Justice, 2019. Consulté le 25 mars 2020 à www.ippr.org/files/2019-07/the-future-is-ours-women-automation-equality-july19.pdf
75. Différences numériques : Les répercussions de l'automatisation sur l'économie autochtone au Canada, Institut de la diversité, Conseil canadien pour l'entreprise autochtone et Centre des compétences futures. Consulté le 8 février 2021 à <https://fsc-ccf.ca/fr/recherche/differences-numeriques-les-repercussions-de-lautomatisation-sur-leconomie-autochtone-au-canada/>
76. Broady, K. « Race and Jobs at High Risk to Automation », Joint Center for Political and Economic Studies, 2017. Consulté le 8 février 2021 à https://jointcenter.org/wp-content/uploads/2017/12/Race-and-Jobs-at-High-Risk-to-Automation-12-18-17-11_30-am.docx-2_o.pdf
77. Hegewisch, A., C. Childers et H. Hartmann. « Women, Automation, and the Future of Work », Institute for Women's Policy Research, 2019. Consulté le 19 mars 2020 à http://iwpr.org/wp-content/uploads/2019/03/C476_Automation-and-Future-of-Work.pdf
78. Bessen, J. « Automation and Jobs: When Technology Boosts Employment », Boston University School of Law, Law & Economics Paper No. 17-09, 2019.
79. Pasquale, F. « New Laws of Robotics: Defending human expertise in the age of AI », Belknap Press, 2020.
80. Roos, G. et Z. Shroff. « What will happen to the jobs? Technology-enabled productivity improvement - good for some, bad for others », Labour & Industry, no 3, p. 165 192, 2017.
81. Baldwin, R.E. « The Globotics Upheaval: Globalization, Robotics and the Future of Work », Oxford University Press, 2019.
82. Schlesinger, J. « Why everyone from students to NASA is using this Canadian invention », The Globe and Mail, 30 juin 2021. Consulté le 4 juillet 2021 à www.theglobeandmail.com/business/technology/article-why-everyone-from-students-to-nasa-is-using-this-canadian-invention/
83. Agrawal, A., J.S. Gans et A. Goldfarb. « Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction », Journal of Economic Perspectives, vol. 33, no 2, p. 31 50, 2019.
84. Kellogg, K.C., M.A. Valentine et A. Christin. « Algorithms At Work: The New Contested Terrain of Control », Academy of Management Annals, vol. 14, no 1, p. 366 410, 2020.
85. Latham, S. et B. Humberd. « Four Ways Jobs Will Respond to Automation », MIT Sloan Management Review, 2018. Consulté le 3 mai 2021 à <https://sloanreview.mit.edu/article/four-ways-jobs-will-respond-to-automation/>
86. Acemoglu, D. et P. Restrepo. « Artificial Intelligence, Automation and Work », série de documents de travail, NBER, 2018.
87. Lim, J. « Survey of labour force shows racialized unemployment gap due to COVID 19 », 2020. Consulté le 22 mars 2021 à <https://ipolitics.ca/2020/08/07/july-labour-force-survey-shows-racialized-gap-in-unemployment-due-to-covid-19/>
88. Desjardins, D. et C. Freestone. « COVID Further Clouded the Outlook for Canadian Women at Risk of Disruption », RBC Economics, 2021. Consulté le 10 mars 2021 à <https://thoughtleadership.rbc.com/covid-further-clouded-the-outlook-for-canadian-women-at-risk-of-disruption/>
89. Jivani, A.I. « Gender Lens to the Future of Work », Daedalus, vol. 149, no 1, 2020.
90. Roos, G. et Z. Shroff. « What will happen to the jobs? Technology-enabled productivity improvement - god for some, bad for others », Labour & Industry, no 3, p. 165 192, 2017.
91. Kearns, M. et A. Roth. « The Ethical Algorithm: The Science of Socially Aware Algorithm Design », Oxford University Press, 2020.
92. Pasquale, F. « New Laws of Robotics: Defending human expertise in the age of AI », Belknap Press, 2020.
93. Pasquale, F. « New Laws of Robotics: Defending human expertise in the age of AI », Belknap Press, 2020.
94. O'Neil, C. « Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy », Crown, 2016.
95. De Fine Licht, K. et J. de Fine Licht. « Artificial intelligence, transparency, and public decision-making », AI & Society, 2020.
96. Coglianese, C. et D. Lehr. « Transparency and Algorithmic Governance », bourse d'études, Penn Law, 2123, 2019.
97. Collett, C. et S. Dillon. « AI and Gender: Four Proposals for Future Research », Cambridge, The Leverhulme Centre for the Future of Intelligence, 2019. Consulté le 20 février 2021 à http://lcfi.ac.uk/media/uploads/files/AI_and_Gender__4_Proposals_for_Future_Research_210619_p8qAu8L.pdf
98. H.R.2231 - Algorithmic Accountability Act de 2019 (2019). Consulté le 24 mars 2021 à www.congress.gov/bill/116th-congress/house-bill/2231
99. Commission européenne (2021). Proposition d'un règlement prévoyant des règles harmonisées sur l'intelligence artificielle (législation sur l'intelligence artificielle). Consulté le 12 mai 2021 à <https://eur-lex.europa.eu/legal-content/FR/T/?qid=1623335154975&uri=CELEX%3A52021PC0206>
100. Donahoe, E. et M. MacDuffee Metzger. « Artificial Intelligence and Human Rights », Journal of Democracy, vol. 30, no 2, 2019.
101. Déclaration de Toronto. Consulté le 2 mars 2021 à www.torontodeclaration.org/declaration-text/francais/
102. Gouvernement du Canada (2019). Directive sur la prise de décisions automatisée. Consulté le 3 mars 2021 à www.tbs-sct.gc.ca/pol/doc-fra.aspx?id=32592
103. Poetranto, I., V. Heath et P. Molnar. « Canada's Advisory Council on AI lacks diversity », 2019. Consulté le 1er mars 2021 à www.thestar.com/opinion/contributors/2019/05/29/canadas-advisory-council-on-ai-lacks-diversity.html
104. Commissariat à la protection de la vie privée du Canada (2020). Un cadre réglementaire pour l'IA : Recommandations pour la réforme de la LPRPDE. Récupéré le 1er mars 2021 à www.priv.gc.ca/fr/about-the-opc/what-we-do/consultations/completed-consultations/consultation-ai/reg-fw_202011/

105. Gouvernement de l'Ontario (2021). Consultations sur le cadre de l'intelligence artificielle (IA) de confiance de l'Ontario. Consulté le 1er mai 2021 à www.ontario.ca/fr/page/consultations-sur-le-cadre-de-lintelligence-artificielle-ia-de-confiance-de-lontario
106. McKelvey, F. et M. MacDonald. « Artificial Intelligence Policy Innovations at the Canadian Federal Government », *Canadian Journal of Communication Policy Portal*, vol. 44, p. 43-50, 2019.
107. Jaume-Palasi, L. « Why We Are Failing to Understand the Societal Impact of AI », *Social Research: An International Quarterly*, vol. 86, no 2, 2019.
108. Askill, A., M. Brundage et G. Hadfield. « The Role of Cooperation in Responsible AI Development », document de travail, 2019. Consulté le 24 mars 2021 à <https://arxiv.org/abs/1907.04534>
109. Wynn, A.T. et S.J. Correll. « Puncturing the pipeline: Do technology companies alienate women in recruiting sessions? », *Social Studies of Science*, vol. 48, no 1, 2018.
110. Seron, C., S. Silbey, E. Cech et B. Rubineau. « Persistence is Cultural: Professional Socialization and the Reproduction of Sex Segregation », *Work and Occupations*, vol. 43, no 2, 2016.
111. Noble, S.U. « Algorithms of oppression: How search engines reinforce racism », New York University Press, New York, 2018.
112. Illanes, P., S. Lund, M. Mourshed, S. Rutherford et M. Tyreman. « Retraining and reskilling workers in the age of automation », McKinsey Global Institute, 2018. Consulté le 8 avril 2020 à www.mckinsey.com/featured-insights/future-of-work/retraining-and-reskilling-workers-in-the-age-of-automation
113. Dessanti, C. « The She-Covary Project: Confronting the Gendered Economic Impacts of COVID 19 in Ontario », Ontario Chamber of Commerce, 2020. Consulté le 12 février 2021 à <https://occ.ca/wp-content/uploads/OCC-shecovery-final.pdf>
114. De Henau, J. et S. Himmelweit. « A Care-Led Recovery From Covid 19: Investing in High-Quality Care to Stimulate and Rebalance the Economy », *Feminist Economics*, 2021. Consulté le 20 mars 2021 à <https://doi.org/10.1080/13545701.2020.1845390>
115. Collett, C. et S. Dillon. « AI and Gender: Four Proposals for Future Research », Cambridge, The Leverhulme Centre for the Future of Intelligence, 2019. Consulté le 20 février 2021 à http://lcfi.ac.uk/media/uploads/files/AI_and_Gender__4_Proposals_for_Future_Research_210619_p8qAu8L.pdf
116. Noble, S.U. « Algorithms of oppression: How search engines reinforce racism », New York University Press, New York, 2018.
117. Noble, S.U. « Algorithms of oppression: How search engines reinforce racism », New York University Press, New York, 2018



Une perspective d'équité en matière d'intelligence artificielle

Carmina Ravanera et Sarah Kaplan

Institute for Gender and the Economy,
Rotman School of Management,
Université de Toronto
Le 15 août 2021



Rotman School of Management
UNIVERSITY OF TORONTO



www.gendereconomy.org/artificial-intelligence