



IDÉE 5

L'UTILISATION DE L'IA POUR PRÉDIRE LE COMPORTEMENT HUMAIN POURRAIT S'AVÉRER INEFFICACE

Si l'IA fait parfois des prédictions impressionnantes sur le comportement humain, nombre d'entre elles sont inexactes. Prendre des décisions sur la base de ces prédictions peut avoir des conséquences désastreuses pour les personnes. Il pourrait être impossible d'améliorer la technologie à un niveau tel que ses avantages l'emportent sur les coûts.

AUJOURD'HUI

De plus en plus de gouvernements et d'institutions utilisent l'IA pour prédire le comportement humain et prendre des décisions concernant les personnes. Par exemple, plus de 500 écoles aux États-Unis utilisent un modèle d'IA appelé Navigate pour prédire la réussite des élèves¹. Aux États-Unis, les travailleurs sociaux ont utilisé l'IA pour prédire quels appels de protection de l'enfance doivent faire l'objet d'une enquête plus approfondie². Tous deux sont des exemples d'« **optimisation prédictive** »³. D'éminents ingénieurs en IA ont affirmé que les algorithmes d'optimisation prédictive reposaient sur une science erronée, les prédictions de l'IA n'étant que légèrement plus précises que le hasard d'un tirage à pile ou face⁴. Malgré cela, ils continuent d'être utilisés parce qu'ils externalisent des tâches complexes telles que l'élaboration de règles de prise de décision (p. ex., quels sont les critères à prendre en compte pour rechercher des comportements frauduleux ou comment décider si un enfant risque d'être victime d'abus). Les règles de prise de décision générées par l'humain peuvent sembler subjectives et inexactes par rapport à celles des modèles prédictifs d'IA, qui prétendent représenter des modèles objectifs dans le monde réel.

Optimisation prédictive

L'utilisation de l'IA pour prédire des résultats futurs sur la base de données historiques, afin de prendre des décisions concernant des personnes.



Les modèles prédictifs ne sont pas toujours justes. Les modèles d'IA prédictifs sont confrontés à de nombreux problèmes, notamment des erreurs dues à une inadéquation entre les données d'entraînement et les données de déploiement. Comme l'IA prédictive doit être formée sur des données antérieures, elle ne peut pas prendre en compte les variables émergentes et complexes du monde, et des comportements humains individuels. Les modèles peuvent être incapables de prendre en compte des facteurs nouveaux et inattendus. En outre, l'IA ne peut pas filtrer les effets des pratiques racistes dans le monde réel, telles que le maintien de l'ordre disproportionné dans les quartiers ou les communautés noirs, qui entraîne une augmentation des fausses arrestations⁵. Cela a conduit à des prévisions inexactes pour les personnes vulnérables⁶.

Les modèles d'IA prédictifs ne peuvent pas comprendre pourquoi le comportement dans le monde réel diffère de leurs prédictions. Les modèles peuvent supposer que les personnes agiront de manière rationnelle ou suivront les mêmes règles et modèles humains dans l'ensemble. Les modèles peuvent ne pas tenir compte des facteurs structurels qui expliquent les différences entre les comportements prévus et les comportements réels. L'accent mis sur la prédiction peut empêcher la découverte de processus susceptibles d'engendrer de nouveaux comportements, comme lorsque la simplification du langage utilisé dans les convocations au tribunal a diminué le taux de personnes ne se présentant pas au tribunal⁷.

Bien que cela soit parfois justifié par des économies de coûts, certains gouvernements ont ressenti des répercussions importantes après avoir utilisé des modèles d'optimisation prédictive. Par exemple, en 2021, le gouvernement néerlandais a démissionné à la suite d'un scandale impliquant l'adoption par l'administration fiscale d'une IA autoapprenante pour prédire les fraudes aux allocations de garde d'enfants⁸. L'IA a identifié à tort des dizaines de milliers de familles comme ayant des dettes excessives envers l'administration fiscale. Plus d'un millier d'enfants ont été placés dans des familles d'accueil et certaines victimes se sont suicidées. Les amendes infligées aux agences gouvernementales et les indemnités versées aux victimes pourraient atteindre plusieurs centaines de millions d'euros.



AVENIRS

À l'avenir, l'optimisation prédictive pourrait être utilisée dans certaines juridictions, mais pas dans d'autres. Elle pourrait être interdite dans certaines juridictions, en particulier là où les gouvernements ont dû faire face à des coûts élevés et à un examen minutieux en raison d'échecs. Cela pourrait encore permettre au secteur privé d'étendre ses utilisations actuellement opaques de l'optimisation prédictive⁹. D'autres juridictions peuvent continuer à utiliser des algorithmes d'optimisation prédictive malgré les risques. Cela peut s'expliquer par le fait que les personnes concernées sont moins à même de saisir la justice ou que leurs gouvernements ne sont pas liés par des normes démocratiques. D'autres peuvent considérer l'optimisation prédictive comme un outil inévitablement imparfait, mais dont l'utilisation peut être justifiée par des économies de coûts. Les institutions, y compris les gouvernements, qui adoptent l'IA pour l'optimisation prédictive et constatent que les coûts dépassent les avantages pourraient maintenir les systèmes en service bien plus longtemps qu'elles ne le devraient ou ne le voudraient, en raison des montants élevés déjà investis ou des difficultés liées à l'annulation d'un déploiement. Certains pourraient considérer l'IA prédictive comme inacceptable d'un point de vue éthique pour la prise de décision, et préférer travailler sur des interventions visant à réduire autant que possible les résultats négatifs prévus.



IMPLICATIONS

- ▶ Les gouvernements et les entreprises qui utilisent l'optimisation prédictive sans être transparents sur les règles de prise de décision de l'IA **pourraient être considérés comme indignes de confiance**
- ▶ Si les institutions utilisent l'IA pour l'optimisation prédictive alors que la charge de la preuve pour contester des prédictions inexactes incombe aux personnes concernées, **les populations déjà vulnérables risquent de voir leur situation s'aggraver**. Cela pourrait créer de nouveaux goulets d'étranglement bureaucratiques et engorger les tribunaux avec des litiges portant sur des préjudices algorithmiques, y compris des affaires liées aux droits de l'humain ou à des violations de la Charte
- ▶ Les tentatives de sacrifier les droits individuels au profit de gains collectifs peuvent **bénéficier aux populations privilégiées au détriment des plus vulnérables**, créant ainsi de plus grandes divisions socioéconomiques
- ▶ L'adoption de modèles d'optimisation prédictive pourrait créer **des économies initiales qui cèdent rapidement la place à de nouveaux coûts** : pour lutter contre les litiges liés à des prédictions inexactes, pour résigner des contrats avec des fournisseurs afin de recycler et de réajuster les modèles, et pour créer de nouvelles voies pour les plaintes et l'indemnisation des dommages
- ▶ Si le processus décisionnel de l'IA punit préventivement les personnes sur la base d'hypothèses biaisées, il pourrait **réduire l'autonomie des populations vulnérables** et dresser de nouveaux obstacles dans leur parcours de vie

Notes de fin de texte

- 1 Feathers, Todd. « [Texas A&M Drops “Race” from Student Risk Algorithm Following Markup Investigation](#) ». The Markup, 30 mars 2021.
- 2 PBS News. « [AP Report: DOJ Examining AI Screening Tool Used by Pa. Child Welfare Agency](#) ». PBS NewsHour, 31 janvier 2023.
- 3 Wang, Angelina; Kapoor, Sayash; Barocas, Solon et Narayanan, Arvind. « [Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms That Optimize Predictive Accuracy](#) ». SSRN Scholarly Paper. Rochester (New York) 4 octobre 2022.
- 4 Wang, Angelina; Kapoor, Sayash; Barocas, Solon et Narayanan, Arvind. « [Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms That Optimize Predictive Accuracy](#) ». SSRN Scholarly Paper. Rochester (New York) 4 octobre 2022.
- 5 Knox, Dean; Lowe, Will et Mummolo, Jonathan. « [Administrative Records Mask Racially Biased Policing](#) ». *American Political Science Review*, vol. 114, no 3 (août 2020) : p. 619-637.
- 6 Wang, Angelina; Kapoor, Sayash; Barocas, Solon et Narayanan, Arvind. « [Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms That Optimize Predictive Accuracy](#) ». SSRN Scholarly Paper. Rochester (New York) 4 octobre 2022.
- 7 Fishbane, Alissa; Ouss, Aurelie et Shah, Anuj K. « [Behavioral Nudges Reduce Failure to Appear for Court](#) ». *Science*, vol. 370, no 6517 (6 novembre 2020) : eabb6591.
- 8 Brenninkmeijer, Alex et Seldam, Björn ten. « [The Dutch Benefits Scandal: A Cautionary Tale for Algorithmic Enforcement](#) ». Application du droit de l'UE (blogue), 30 avril 2021.
- 9 Rhea, Alene K.; Markey, Kelsey; D'Arinzo, Lauren; Schellmann, Hilke; Sloane, Mona; Squires, Paul; Arif Khan, Falaah et Stoyanovich, Julia. « [An external stability audit framework to test the validity of personality prediction in AI hiring](#) ». *Data Mining and Knowledge Discovery*, vol. 36, no 6 (2022) : p. 2153-2193.

© Sa Majesté le Roi du chef du Canada, 2025

Pour obtenir des informations sur les droits de reproduction : <https://horizons.gc.ca/fr/contactez-nous/>

PDF : PH4-215/2025F-PDF

ISBN : 978-0-660-76890-8

Also available in English under the title: Using AI to predict human behaviour may not work.

AVERTISSEMENT

Horizons de politiques Canada (Horizons de politiques) est le centre d'excellence en prospective du gouvernement du Canada. Notre mandat est de doter le gouvernement du Canada d'une perspective et d'un état d'esprit tournés vers l'avenir afin de renforcer la prise de décisions. Le contenu de ce document ne représente pas nécessairement le point de vue du gouvernement du Canada ou des agences et ministères participants.