



IDÉE 3

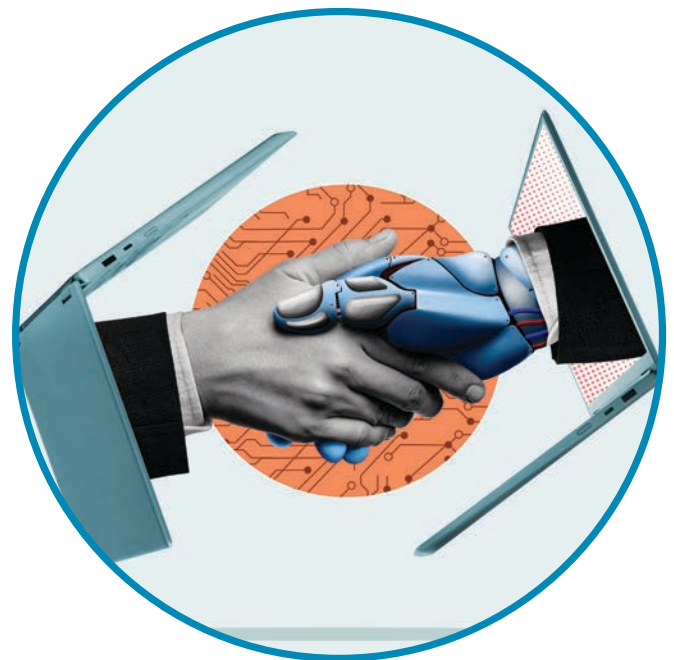
LE MANQUE DE CONFIANCE DANS L'IA POURRAIT ENTRAVER SON ADOPTION



On ne sait pas comment évoluera la confiance dans l'IA. Les défaillances fréquentes ou non corrigées des systèmes d'IA, ou une seule défaillance importante, pourraient éroder la confiance et entraver l'adoption, mettant en péril des secteurs entiers. Les nouvelles formes de certification, de vérification et d'efforts pour réparer les préjudices pourraient encourager la confiance des utilisatrices et utilisateurs et l'adoption du système.

AUJOURD'HUI

La confiance est essentielle à l'acceptation de l'IA et, au Canada, la confiance dans l'IA est en déclin¹. L'indice CanTrust montre que la confiance des Canadiennes et Canadiens dans l'IA a diminué de 6 % entre 2018 et 2024². L'étude mondiale IPSOS AI Monitor montre que l'anglosphère, y compris le Canada, fait moins confiance à l'IA que d'autres régions : par exemple, 63 % des Canadiennes et Canadiens sont nerveux à l'égard des produits et services qui utilisent l'IA, contre seulement 25 % des Japonaises et Japonais³.



La confiance dans l'IA dépend du contexte dans lequel elle est utilisée. Par exemple, la confiance est plus élevée pour des tâches simples telles que le réglage d'un thermostat, et plus faible pour des tâches liées à la sécurité personnelle telles que les voitures autopilotées⁴. La confiance du public dans les voitures autonomes est faible et en baisse. En 2023, seuls 22 % des Canadiennes et Canadiens déclarent faire confiance aux voitures autonomes et autres moyens de transport sans conducteur basés sur l'IA⁵; contre 37 % des Américaines et Américains, soit une baisse par rapport aux 39 % de 2022 et aux 41 % de 2021⁶.

Malgré la baisse de confiance, l'utilisation des outils d'IA au Canada augmente. Selon un sondage de 2024 de Legar, 30 % des Canadiennes et Canadiens utilisent désormais l'IA, contre 25 % il y a un an. Les jeunes utilisent davantage l'IA que les personnes plus âgées; 50 % des 18-35 ans déclarent utiliser l'IA, contre seulement 13 % des 55 ans et plus⁷.

Les risques et les échecs liés aux technologies de l'IA ont fréquemment attiré l'attention du public au cours de l'année écoulée. Dans certains cas, la mise au point et les tests de nombreux outils d'IA ont été effectués après le lancement public, ce qui contraste fortement avec les essais de médicaments cliniques qui nécessitent de longues périodes de test avant d'être rendus publics. De nouvelles initiatives ont vu le jour pour recenser les incidents liés à l'IA et en rendre compte, comme la base de données des incidents liés à l'IA et le moniteur des incidents liés à l'IA de l'OCDE^{8,9}.

L'adoption de l'IA peut sembler forcée, plutôt que choisie par la personne. La volonté actuelle d'intégrer l'IA partout peut signifier que les préoccupations valables concernant la sécurité des données, l'équité, les conséquences environnementales et la sécurité de l'emploi sont minimisées¹⁰. Forcer les gens à adopter l'IA dans leur vie quotidienne sans faire d'efforts pour rendre la technologie plus digne de confiance peut limiter les répercussions transformatrices potentielles de la technologie¹¹. La réaction actuelle contre l'utilisation croissante de la technologie de reconnaissance faciale dans les aéroports est un exemple de l'interaction entre l'adaptation forcée et l'absence de confiance¹².





AVENIRS

L'amélioration des technologies, des pratiques et des systèmes pourrait contribuer à renforcer la confiance dans l'IA. Par exemple, de nouvelles capacités telles que l'IA neuro-symbolique, qui combine des réseaux neuronaux avec un traitement symbolique basé sur des règles, promettent d'améliorer la transparence et l'explicabilité des modèles d'IA. L'adoption par les entreprises de nouveaux modèles d'étiquetage, de certification ou d'assurance pourrait compenser une partie de la méfiance à l'égard de l'IA^{13, 14}. Certains fournisseurs mettent actuellement au point des moyens d'évaluer la sécurité et la fiabilité des modèles d'IA, en offrant des garanties pour vérifier leurs performances^{15, 16}. À l'avenir, les systèmes d'IA pourraient fournir un intervalle de confiance pour tout, des résultats de recherche aux véhicules autopilotés, aidant les utilisateurs à évaluer les risques et les incertitudes¹⁷.

Un déploiement plus stratégique et réfléchi de l'IA pourrait renforcer la confiance. À l'avenir, l'IA deviendra probablement la bonne solution à certains problèmes, mais pas à d'autres. La confiance dans l'IA pourrait être renforcée si les gens perçoivent qu'elle leur facilite la vie¹⁸, plutôt que de remplacer des tâches qu'ils apprécient ou de sembler être une solution à la recherche d'un problème. La familiarité individuelle avec l'IA peut renforcer la confiance dans un domaine du travail ou de la vie, sans nécessairement se traduire par des niveaux de confiance accrus dans l'ensemble de l'écosystème de l'IA¹⁹.

Les échecs très médiatisés et l'appréciation croissante des risques pourraient éroder la confiance. Le scepticisme et la méfiance pourraient s'accroître à mesure que les risques de l'IA sont mieux connus et documentés et que davantage de tâches à fortes répercussions sont déléguées à l'IA. Les groupes qui subissent les effets négatifs de l'IA s'opposent activement à son utilisation dans certains domaines, comme les écrivaines, écrivains et artistes qui s'organisent collectivement pour limiter ce qu'ils considèrent comme le pouvoir destructeur de la technologie²⁰. La méfiance pourrait être alimentée non seulement par des récits décrivant l'IA comme une menace d'extinction, mais aussi par son association avec l'accroissement des inégalités²¹. De même, les échecs technologiques très médiatisés pourraient faire planer l'ombre d'une perte de confiance en l'avenir. Par exemple, la confiance du public et le soutien à l'énergie nucléaire au Canada ont considérablement diminué à la suite de l'accident nucléaire de Fukushima Daiichi en 2011, et les préoccupations du public concernant la sûreté nucléaire ont entravé la croissance du secteur pendant des années²². Une perte de confiance semblable dans les technologies de l'IA telles que les voitures autonomes pourrait mettre en péril non seulement une entreprise, mais aussi des secteurs entiers.



IMPLICATIONS

- Le manque de confiance pourrait être un **obstacle majeur** à l'intégration de l'IA dans certains secteurs
- Un seul **incident aberrant et très médiatisé** impliquant un système d'IA bien établi pourrait nuire de manière disproportionnée à la confiance dans l'IA et à son adoption; par exemple, une crise financière déclenchée par un contenu généré par l'IA et le commerce algorithmique à haute fréquence
- Les gens pourraient faire confiance à l'IA pour effectuer certaines tâches **plus qu'ils ne font confiance à d'autres humains**
- **Les différents niveaux de confiance dans l'IA** selon les groupes ou les cas d'utilisation pourraient **unir les gens au-delà des divisions sociétales habituelles** ou les polariser d'une nouvelle manière
- **Une confiance excessive** dans certains résultats de l'IA **pourrait accroître la désinformation**, avec des conséquences pour la démocratie et la cohésion sociétale
- **Une mauvaise expérience avec un système d'IA pourrait conduire à une perte de confiance dans d'autres systèmes d'IA**, alors qu'une expérience positive avec un outil d'IA pourrait conduire à une confiance accrue dans d'autres applications d'IA
- **La jurisprudence et la législation** qui déterminent la responsabilité des décisions prises par ou avec l'IA **pourraient influencer la confiance et l'adoption**
- L'émergence de **nouveaux labels et certifications** pourrait nuire à la confiance des consommateurs dans l'IA, tels que les labels d'avertissement, ou les labels analogues à ceux du commerce équitable ou des produits biologiques²³
- **Les régimes d'imputabilité et de responsabilité seront clarifiés**, et de nombreux systèmes devront déterminer qui est responsable des défaillances de l'IA

Notes de fin de texte

- 1 Gillespie, Nicole; Steven, Lockey; Caitlin Curtis; Javad Pool et Ali Akbari. « [Trust in Artificial Intelligence: A Global Study](#) ». The University of Queensland and KPMG Australia, 2023.
- 2 Proof Strategies. « [L'Indice CanTrust 2024](#) ». Proof Strategies, 13 février 2024.
- 3 Carmichael, Matt et Jamie Stinson. « [The Ipsos AI Monitor 2024: Changing Attitudes and Feelings about AI and the Future It Will Bring](#) » Ipsos AI Monitor. Paris : Ipsos, 6 juin 2024.
- 4 Karadeglija, Anja. « [Poll Finds More Canadians Using AI despite 'deep-Rooted' Fears](#) ». Le National Post, 9 février 2024.
- 5 Chhim, Chris et Sanyam Sethi. « [Canadians Among Least Likely to Believe Artificial Intelligence Will Make Their Lives Better](#) ». Paris : Ipsos, 14 janvier 2022.
- 6 Stilgoe, Jack. « What Does It Mean to Trust a Technology? » *Science*, vol.382, no 6676 (s.d.) : 9782.
- 7 « [Utilisation d'outils d'IA](#) ». Léger, 5 février 2024.
- 8 L'Observatoire OCDE des politiques de l'IA. « [AIM: The OECD AI Incidents Monitor, an Evidence Base for Trustworthy AI](#) ». Consulté le 7 août 2024.
- 9 « [Bienvenue dans la base de données des incidents d'IA](#) ». Consulté le 7 août 2024.
- 10 O'Gorman, Marcel. « [Opinion: When It Comes to AI, It Feels like We're Doing More Adaptation than Adoption These Days. This Is Not a Good Feeling](#) ». *The Globe and Mail*, 21 juin 2024.
- 11 Marr, Bernard. « [Will AI Really Revolutionize Every Industry? A Critical Analysis](#) ». Forbes, 23 juillet 2024.
- 12 Vermes, Jason. « [Airports want to scan your face to make travelling easier. Privacy experts caution it's not ready for takeoff](#) ». *CBC News*. 3 mars 2024.
- 13 Schuman, Leslie. « [Leading Corporations Introduce Data Provenance Standards](#) ». *Businesswire*, 30 novembre 2023.
- 14 The Data & Trust Alliance. « [The Data & Trust Alliance](#) ». Consulté le 7 août 2024.
- 15 « [LLM Safety Leaderboard - a Hugging Face Space by AI-Secure](#) ». Consulté le 7 août 2024.
- 16 Masse, Bryson. « [Armilla Offers Verification and Warranties for Enterprises Using AI Models](#) ». *VentureBeat* (blogue), 3 octobre 2023.
- 17 Sanz Sáiz, Beatriz. « [How your organization can have confidence in the opportunities AI brings](#) ». *Ernst & Young* (blogue), 15 janvier 2024.
- 18 Gillespie, Nicole; Lockey, Steven; Curtis, Caitlin; Pool, Javad et Ali Akbari. « [Trust in Artificial Intelligence: A global study](#) ». *The University of Queensland* et *KPMG*. (2023).

- 19 Horowitz, Michael C., Lauren Kahn, Julia Macdonald et Jacquelyn Schneider. « [Adopting AI: How Familiarity Breeds Both Trust and Contempt](#) ». *AI & SOCIETY*, 12 mai 2023.
- 20 Schneider, Michael. « [SXSW Audiences Loudly Boo Festival Videos Touting the Virtues of AI](#) ». *Variety*, 13 mars 2024.
- 21 Perrigo, Billy. « [AI Poses Extinction-Level Risk, State-Funded Report Says](#) ». *Time Magazine*, 11 mars 2024.
- 22 Personnel de la Presse canadienne. « [Parti Québécois Government to Close Gentilly-2 Nuclear Power Plant | Globalnews.Ca](#) ». *Global News*, 12 septembre 2012.
- 23 Huet, Ellen. « [AI Certification Program Verifies Systems Are 'Fairly Trained' - Bloomberg](#) ». *Bloomberg*. Consulté le 13 février 2024.

© 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 : PH4-213/2025F-PDF

ISBN : 978-0-660-76886-1

Also available in English under the title: Lack of trust in AI could impede its adoption.

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.