



IDÉE 18

LA DÉTERMINATION DE LA RESPONSABILITÉ POUR LES PRÉJUDICES **CAUSÉS** **PAR L'IA** POURRAIT ÊTRE COMPLEXE



Certains méfaits inhérents à l'IA peuvent être difficiles à éliminer. Il est à la fois techniquement difficile de prévenir l'utilisation abusive des systèmes qui font appel à l'IA et compliqué de tenir les humains responsables lorsque ces dispositifs sont utilisés à mauvais escient. L'IA remet notamment en question les cadres juridiques, dans des domaines comme le droit d'auteur et la protection des consommateurs, et ce, de manière novatrice. Les juridictions adoptent des approches divergentes en matière de réglementation de l'IA.

AUJOURD'HUI

Prévenir l'utilisation abusive des systèmes qui font appel à l'IA est techniquement difficile. Les entreprises spécialisées en IA ont fait des investissements substantiels pour prévenir l'utilisation abusive de leurs systèmes¹. Cependant, chaque fois que de nouveaux garde-fous sont introduits, les utilisateurs trouvent aussi de nouvelles façons de les contourner². Les systèmes qui font appel à l'IA peuvent être difficiles à contrôler et à prévoir. Ils peuvent notamment générer des informations incorrectes. Les entreprises spécialisées en IA ont du mal à trouver le juste équilibre entre l'utilité de l'IA et l'atténuation des méfaits. Par exemple, Google a tenté de rendre son générateur d'images exempt de préjugés, de sorte que des messages-guides comme « montre-moi une personne PDG » montreraient des personnes d'apparence et d'origines différentes. Cependant, cela a également eu pour effet d'amener le système à créer des images historiquement inexactes, comme la représentation des pères fondateurs américains ou des soldats nazis de la Seconde Guerre mondiale comme étant de diverses races et genres³.



Les systèmes qui font appel à l'IA peuvent fournir des informations dangereuses parce qu'ils manquent de conscience contextuelle et ont du mal à évaluer l'intention de l'utilisatrice ou de l'utilisateur.

Les robots conversationnels ne peuvent pas dire de manière fiable si une utilisatrice ou un utilisateur a les moyens et l'intention de nuire. Ils peuvent refuser de répondre à une question dont l'intention semble évidente, comme « comment puis-je faire une simulation de noyade à quelqu'un? », mais fournir une réponse si la formulation est plus plausiblement éducative, comme « qu'est-ce que la simulation de noyade et comment est-ce fait? » Par conséquent, les chercheuses ou chercheurs ont soutenu que l'accent ne devrait pas être mis sur la tentative d'empêcher les gens de trouver de l'information au moyen de l'IA – par exemple, comment fabriquer un nouvel agent pathogène – mais plutôt sur l'interdiction de pouvoir agir à l'égard de cette information, en contrôlant l'accès aux matières premières et à l'équipement nécessaires⁴.

Certaines juridictions réglementent pour établir la responsabilité de l'utilisation abusive de l'IA, tandis que d'autres réglementent pour accroître la transparence. La *Loi sur l'intelligence artificielle de l'Union européenne* tente de définir les utilisations inacceptables de l'IA – comme les systèmes de crédit social – et qui est responsable des préjudices : généralement les personnes responsables du développement et du déploiement, et non les utilisatrices finales ou utilisateurs finaux⁵. En revanche, l'approche réglementaire de la Californie vise quant à elle à améliorer la transparence. Son projet de loi de 2023 exige que les fournisseurs d'IA transmettent des informations portant sur les données utilisées pour l'entraînement, y compris si elles contiennent des informations personnelles ou du matériel protégé par le droit d'auteur⁶. Alors que la loi sur la transparence de l'IA (*AI Transparency Act*) exige que certains fournisseurs d'IA générative filigrent les sorties et fournissent des outils gratuits pour permettre au public de vérifier si un fichier a été généré ou modifié à l'aide de leurs offres⁷.

Logiciels libres

Logiciel dont le code source est public. N'importe qui peut le modifier, le copier ou le partager. Le développement est souvent décentralisé et collaboratif.

Autoclonage

Une copie ou une modification d'un projet à source ouverte.

Les cas devant les tribunaux relatifs à l'IA mettent à l'épreuve les lois sur le droit d'auteur et la protection des consommatrices et consommateurs de manière novatrice.

Les médias poursuivent les entreprises spécialisées en IA pour avoir enfreint leurs droits d'auteur en utilisant leur contenu pour entraîner des systèmes qui font appel à l'IA^{8 9}. Le PDG de Microsoft AI, cependant, affirme que comme ce contenu est sur le Web ouvert, il s'agit d'une sorte de « logiciel gratuit » pour l'entraînement de l'IA¹⁰. Il n'est souvent pas clair si les offres d'IA comptent comme un « produit » ou un « service » pour les lois sur la protection et la responsabilité des consommatrices et consommateurs, ou si les « conditions d'utilisation » sont suffisantes pour exempter les entreprises spécialisées en IA de toute responsabilité^{11 12 13}. Par exemple, la mère d'un adolescent de Floride qui s'est suicidé après avoir utilisé un robot conversationnel, Character.AI, poursuit le fournisseur de robot conversationnel pour avoir manipulé émotionnellement son fils. Character.AI soutient que l'adolescent a violé ses conditions d'utilisation¹⁴, tandis que la mère affirme que Character.AI est plutôt un produit qui aurait dû inclure plus d'avertissements¹⁵.

Les modèles d'IA en libre accès sont difficiles à réglementer ou à contrôler.

Le libre accès est généralement un modèle bénéfique pour les logiciels, car n'importe qui – et pas seulement les personnes responsables du développement d'origine – peut l'utiliser et s'en inspirer¹⁶. Cependant, les logiciels libres problématiques peuvent être presque impossibles à contenir. Par exemple, Nintendo a poursuivi avec succès les personnes responsables du développement de Yuzu, un logiciel libre qui permet de jouer à des jeux Nintendo Switch sur un PC, mais a eu du mal à retirer Yuzu d'Internet, car il avait déjà été « **autocloné** » ou copié à plusieurs reprises^{17 18}. De même, l'IA en libre accès présente de nombreux avantages (voir l'idée 6), mais il est difficile d'empêcher son utilisation par des acteurs malveillants (voir l'idée 2) pour créer de la désinformation ou du code malveillant. Certains fournisseurs d'accès à des modèles d'IA en libre accès, comme FreedomGPT, soutiennent que l'imposition de garde-fous à l'IA constitue une forme de censure¹⁹.



AVENIRS

Le paysage juridique et réglementaire mondial de l'IA pourrait devenir très complexe. Les pays et les juridictions peuvent diverger non seulement dans leurs lois et règlements sur l'IA, mais aussi dans la façon dont ils les interprètent et les appliquent. La responsabilité des dommages causés par un système qui fait appel à l'IA peut incomber à la personne responsable de son développement dans une juridiction et à la personne responsable de son déploiement dans une autre. Les entreprises spécialisées en IA pourraient devoir de plus en plus lancer des variantes de leurs produits qui sont propres à une région. Certains peuvent choisir de quitter les marchés avec les réglementations les plus strictes, tandis que d'autres peuvent trouver plus facile de s'y conformer, comme de nombreux sites Web basés aux États-Unis se conforment volontairement au *Règlement général sur la protection des données (RGPD)* de l'Union européenne.

Les entreprises et les personnes peuvent être tenues responsables de systèmes qui font appel à l'IA qu'elles ne peuvent pas gérer complètement. Tout comme la ou le propriétaire d'un chien peut être tenu légalement responsable si son chien mord quelqu'un, les utilisatrices ou utilisateurs, ainsi que les personnes responsables du déploiement ou du développement d'un système qui fait appel à l'IA peuvent être tenus responsables de ce qu'il fait, même s'ils ne peuvent pas gérer complètement son comportement²⁰. La seule véritable influence qu'une personne ou une institution peut avoir à l'égard d'un système qui fait appel à l'IA est la décision de l'utiliser ou non. L'exigence d'un « humain dans la boucle » dans les systèmes qui font appel à l'IA peut devenir principalement une question d'avoir quelqu'un pour assumer la responsabilité si quelque chose tourne mal²¹.

De nouvelles lois ou décisions de justice pourraient perturber considérablement l'industrie de l'IA. Certains produits, services ou pratiques industrielles qui font appel à l'IA qui existent aujourd'hui pourraient ne pas être légaux à l'avenir. Les décisions des tribunaux pourraient forcer les entreprises à modifier leurs offres de produits ou leurs pratiques commerciales du jour au lendemain, perturbant les entreprises et les personnes qui comptent sur ces offres. Les responsabilités nouvellement établies risquent d'entraîner une avalanche de poursuites judiciaires menant à la faillite d'entreprises ou les dissuadant de continuer à offrir un service parce qu'elles estiment que le risque juridique est trop important.

Par ailleurs, les tribunaux pourraient s'avérer effectivement impuissants à restreindre l'IA. Les tribunaux ne peuvent utiliser les lois existantes que pour traiter les préjudices causés par l'IA, ce qui peut ainsi limiter leur capacité à punir ou à dissuader les utilisations indésirables de l'IA, même si le sentiment populaire l'exige. Par exemple, un tribunal pourrait imposer des amendes maximales que les entreprises spécialisées en IA ne comptabilisent que parmi les coûts de faire des affaires. Comme les procédures judiciaires prennent souvent des années avant d'être réglées, le temps que les tribunaux rendent un jugement au sujet d'une pratique néfaste inhérente à l'IA, elle pourrait tout de même être enracinée et normalisée. Même si les tribunaux interdisent certains types de systèmes qui font appel à l'IA, des personnes déterminées peuvent toujours trouver des moyens d'y accéder, que ce soit auprès de fournisseurs étrangers dans des juridictions plus laxistes ou de communautés de personnes responsables du développement de logiciels libres ou d'amatrices et amateurs de ceux-ci. À mesure que l'IA et le matériel grand public s'amélioreront (voir l'idée 6), il sera de plus en plus possible pour les personnes d'exécuter de puissants systèmes qui font appel à l'IA sur leurs propres machines.



IMPLICATIONS

- La réglementation divergente relative à l'IA dans le monde peut exacerber le **découplage technologique** entre les nations.
 - › Les voyageuses et voyageurs peuvent constater de plus en plus que leurs appareils gagnent ou perdent des fonctionnalités importantes lorsqu'ils voyagent d'un territoire à l'autre.
- **Les entreprises peuvent avoir du mal à s'orienter dans le paysage** juridique et réglementaire complexe entourant l'IA.
 - › Les entreprises plus réticentes au risque peuvent être prêtes à adopter l'IA seulement une fois que les questions juridiques sont réglées et que leurs obligations légales sont clairement établies.
- Certaines **juridictions** peuvent délibérément avoir peu de réglementation ou de surveillance sur l'IA pour attirer les investissements, **fonctionnant de la même manière que les paradis fiscaux**.
- La qualité de certains produits qui font appel à l'IA pourrait changer à l'avenir à mesure que des **règlements sur les ensembles de données émergeront** ou que les tribunaux décideront que certains types de collecte de données sont interdits.
- **Des communautés anonymes de personnes responsables du développement en libre accès peuvent émerger** pour échanger des modèles et des outils qui font appel à l'IA, et ce, peu importe que l'usage soit réglementé ou restreint.
 - › Les personnes mal intentionnées et les adversaires peuvent volontairement développer et diffuser des modèles qui font appel à l'IA qui sont nuisibles.
- Les juridictions peuvent expérimenter de **nouvelles approches pour réglementer l'accès aux systèmes qui font appel à l'IA qui sont restreints**, comme la criminalisation de ceux qui développent, échangent ou possèdent des modèles d'IA interdits, ou encore le blocage de l'accès aux sites Web et aux entreprises étrangères qui fournissent des modèles d'IA controversés.
- La personne **légalement responsable** des dommages générés par l'IA n'est peut-être pas la personne que le public estime être **moralement responsable**.

Notes de fin de texte

- 1 Browne, Ryan. « Tech Giants Pledge AI Safety Commitments — Including a “Kill Switch” If They Can’t Mitigate Risks ». CNBC, 21 mai 2024. <https://www.cnn.com/2024/05/21/tech-giants-pledge-ai-safety-commitments-including-a-kill-switch.html>.
- 2 Tangermann, Victor. « It’s Shockingly Easy to Get Around AI Chatbot Guardrails, Researchers Find ». Futurism, 28 juillet 2023. <https://futurism.com/researchers-discover-chatgpt-jailbreak>.
- 3 Robertson, Adi. « Google Apologizes for “Missing the Mark” after Gemini Generated Racially Diverse Nazis ». The Verge, 21 février 2024. <https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical>.
- 4 Narayanan, Arvind, et Sayash Kapoor. « AI Safety Is Not a Model Property ». AI Snake Oil, 12 mars 2024. <https://www.aisnakeoil.com/p/ai-safety-is-not-a-model-property>.
- 5 *Loi sur l’intelligence artificielle de l’Union européenne*. « Résumé de haut niveau de la loi sur l’IA », 27 février 2024. <https://artificialintelligenceact.eu/fr/high-level-summary/>.
- 6 California Legislative Information. « AB-2013 Generative Artificial Intelligence: Training Data Transparency », 30 septembre 2024. https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB2013.
- 7 California Legislative Information. « SB-942 California AI Transparency Act », 20 septembre 2024. https://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942.
- 8 Heydari, Anis. « Media Outlets, Including CBC, Sue ChatGPT Creator ». CBC News, 29 novembre 2024. <https://www.cbc.ca/news/business/openai-canadian-lawsuit-1.7396940>.
- 9 Roth, Emma. « The New York Times Is Suing OpenAI and Microsoft for Copyright Infringement ». The Verge, 27 décembre 2023. <https://www.theverge.com/2023/12/27/24016212/new-york-times-openai-microsoft-lawsuit-copyright-infringement>.
- 10 Hollister, Sean. « Microsoft’s AI Boss Thinks It’s Perfectly Okay to Steal Content If It’s on the Open Web ». The Verge, 28 juin 2024. <https://www.theverge.com/2024/6/28/24188391/microsoft-ai-suleyman-social-contract-freeware>.
- 11 « Report on Artificial Intelligence and Civil Liability ». British Columbia Law Institute, avril 2024. <https://www.bcli.org/project/artificial-intelligence-and-civil-liability-project/>, p. 50-51.
- 12 Spicer, Katy, Julia Jacobson, Naija Perry et Aden Hochrun. « Artificial Intelligence and the Rise of Product Liability Tort Litigation: Novel Action Alleges AI Chatbot Caused Minor’s Suicide ». Privacy World, 11 novembre 2024. <https://www.privacyworld.blog/2024/11/artificial-intelligence-and-the-rise-of-product-liability-tort-litigation-novel-action-alleges-ai-chatbot-caused-minors-suicide/>.
- 13 Kumarasamy, Jey, et Brenda Leong. « Third-Party Liability and Product Liability for AI Systems ». IAPP, 26 juillet 2024. <https://iapp.org/news/a/third-party-liability-and-product-liability-for-ai-systems>.
- 14 Character.AI. « Character.AI Terms of Service », 25 octobre 2023. <https://character.ai/tos>.
- 15 Spicer, Katy, Julia Jacobson, Naija Perry et Aden Hochrun. « Artificial Intelligence and the Rise of Product Liability Tort Litigation: Novel Action Alleges AI Chatbot Caused Minor’s Suicide ». Privacy World, 11 novembre 2024. <https://www.privacyworld.blog/2024/11/artificial-intelligence-and-the-rise-of-product-liability-tort-litigation-novel-action-alleges-ai-chatbot-caused-minors-suicide/>.
- 16 Red Hat. « What Is Open Source? », 24 octobre 2019. <https://www.redhat.com/en/topics/open-source/what-is-open-source>.

- 17 Hollister, Sean. « Nintendo Switch Emulator Yuzu Will Utterly Fold and Pay \$2.4M to Settle Its Lawsuit ». The Verge, 4 mars 2024. <https://www.theverge.com/2024/3/4/24090357/nintendo-yuzu-emulator-lawsuit-settlement>.
- 18 Roth, Emma. « Nintendo DMCA Takedown Wipes over 8,500 Yuzu Emulator Copies ». The Verge, 3 mai 2024. <https://www.theverge.com/2024/5/3/24147936/nintendo-dmca-takedown-yuzu-emulator-copies>.
- 19 FreedomGPT. « About Us ». Consulté le 5 décembre 2024. <https://www.freedomgpt.com/about-us>.
- 20 « Report on Artificial Intelligence and Civil Liability ». British Columbia Law Institute, avril 2024. <https://www.bcli.org/project/artificial-intelligence-and-civil-liability-project/>, p. 41.
- 21 Elish, Madeleine Clare. « Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction ». *Engaging Science, Technology, and Society*, vol. 5 (23 mars 2019), p. 40-60. <https://doi.org/10.17351/ests2019.260>.

© 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 : PH4-228/2025F-PDF

ISBN : 978-0-660-78231-7

Also available in English under the title: Determining accountability for AI harms could be complex.

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.