

3

DÉFIS

L'intégration de l'IA dans le domaine de la recherche juridique soulève plusieurs préoccupations quant à l'exactitude et la qualité des résultats générés par les outils d'IA, de même que des questions relatives à la confidentialité et la propriété intellectuelle. Il est essentiel qu'une revue de droit ait une compréhension approfondie de ces questions et de leurs incidences pour qu'elle puisse s'assurer qu'elles sont correctement prises en compte avant d'intégrer les outils d'IA dans ses processus internes ou d'en autoriser l'utilisation externe.



3.1 Considérations liées à l'exactitude et à la qualité

Notre recherche a révélé trois défis principaux à relever en ce qui a trait à l'exactitude et la qualité que pose l'IA génératrice et son utilisation dans le domaine de la recherche : (1) les préjugés et la discrimination que recèlent les algorithmes des outils d'IA, (2) les hallucinations de faits et de sources, et (3) la transparence quant au recours à l'IA dans les soumissions.

3.1.1 Préjugés et discrimination que recèlent les algorithmes des outils d'IA

Une façon de recourir à l'IA dans le domaine de la recherche juridique est d'utiliser un robot conversationnel comme ChatGPT ou Gemini pour effectuer une recherche. Dans ses premières itérations, l'IA générative « s'acquittait [surtout] de tâches suivant un ensemble de règles établies » [notre traduction]¹⁸. Cependant, les modèles probabilistes actuels d'IA générative se fondent sur des algorithmes complexes et de grandes quantités de données pour générer des réponses aux questions de l'utilisateur. L'exactitude et la validité des réponses dépendent donc des données utilisées et de l'algorithme suivi, puisqu'ils sont le fondement grâce auquel chaque modèle déduit la « réponse appropriée » [notre traduction]¹⁹.

En conséquence, si les données renferment des préjugés ou des stéréotypes, ceux-ci se « refléteront dans les réponses fournies » [notre traduction] par le logiciel²⁰. Les algorithmes procèdent en fonction de la reconnaissance de schémas et « d'une certaine intelligence qui manipule les données pour aboutir à un résultat » [notre traduction]²¹. Des études récentes ont démontré que, suivant leur conception, « les algorithmes peuvent introduire systématiquement des préjugés par inadvertance, renforcer la discrimination historique, écarter certains renseignements ou consolider des pratiques désuètes ou des échecs du passé » [notre traduction]²².

Dans une étude sur les sources de données utilisées par l'IA générative, des journalistes du quotidien Washington Post ont collaboré avec des scientifiques de l'Allen Institute for AI pour analyser l'ensemble des données de Google C4²³. Ils ont découvert que, dans la catégorie « nouvelles et média », la troisième plus importante en termes de nombre de données, une part importante d'entre elles provenaient de Wikipédia, de même que de plusieurs sources « mal cotées sur l'échelle indépendante NewsGuard pour la fiabilité » [notre traduction]²⁴.

La possibilité que des préjugés et de la discrimination découlent de cette situation pose un risque éthique considérable pour les revues de droit. Les auteurs qui se fient exclusivement aux algorithmes des outils d'IA pour effectuer leurs recherches et rédiger leurs soumissions courent le risque de propager de l'information erronée et de produire de la recherche de piètre qualité, ce qui, à son tour, a pour effet de miner la qualité, la crédibilité et l'intégrité d'une revue.

3.1.2 Hallucinations de faits et de sources

Contrairement aux être humains, les robots conversationnels ne sont pas formés pour évaluer la qualité et la crédibilité du contenu qu'ils génèrent. Il est aussi reconnu que ChatGPT et d'autres robots conversationnels « hallucinent » des renseignements et des sources²⁵. Par exemple, en préparant un mémoire de 10 pages déposé auprès d'une cour de district des États-Unis à New York, ChatGPT a inventé trois décisions judiciaires dont les faits et les raisonnements ressemblaient à ceux soumis à la cour²⁶. Une situation similaire s'est produite dans une cause dont était saisie la Cour suprême de la Colombie-Britannique en janvier 2024, dans le cadre de laquelle la demande, qui visait le prononcé d'une ordonnance autorisant les enfants de la demanderesse à se rendre en Chine pour une visite, faisait référence à des causes « inventées »²⁷. Si cette erreur n'avait pas été décelée,

« Dans une étude menée en 2024 et portant spécifiquement sur le domaine juridique, on a constaté un nombre alarmant d'occurrences d'hallucinations — “entre 58 % [...] et 88 % [...] du temps” — lorsque les questions posées au logiciel concernaient la jurisprudence existante. »

ces causes hallucinées auraient pu influencer la décision de la Cour suprême de la Colombie-Britannique. Cela souligne l'importance de la surveillance humaine pour éviter les effets potentiellement préjudiciables de l'utilisation d'outils d'IA générative tels que ChatGPT dans les procédures judiciaires.

Dans une étude menée en 2024 et portant spécifiquement sur le domaine juridique, on a constaté un nombre alarmant d'occurrences d'hallucinations — « entre 58 % [...] et 88 % [...] du temps » [notre traduction] — lorsque les questions posées au logiciel concernaient la jurisprudence existante²⁸. En outre, l'étude a conclu que les modèles linguistiques d'IA générative « fournissent des réponses apparemment authentiques à des questions juridiques dont les prémisses sont fausses de par leur nature même » et qu'ils ne peuvent pas reconnaître qu'ils hallucinent des « faussetés juridiques » [notre traduction]²⁹. Une autre équipe de chercheurs a conclu que « les robots conversationnels ont constamment inventé de l'information » lorsqu'on leur a demandé de résumer des articles de journaux et que les taux d'hallucination ont augmenté lorsqu'on leur a demandé d'effectuer des tâches « allant au-delà du simple résumé » [notre traduction]³⁰. En réponse, plusieurs lignes directrices sur le recours à l'IA exigent que les auteurs assurent « une surveillance et un contrôle humain » de leurs textes en cas de recours à l'IA générative, notamment parce que « [celle-ci] est en mesure de générer des textes faisant apparemment autorité, mais pouvant être incorrects, incomplets ou empreints de préjugés » [notre traduction]³¹.

Il importe de noter que le degré d'hallucinations des systèmes d'IA générative dépend du grand modèle de langage [ci-après « GML »] utilisé et à quelle fin il est utilisé. Dans le contexte de la recherche juridique, par exemple, un robot conversationnel polyvalent doté d'un GML (comme ChatGPT ou Gemini) peut halluciner davantage qu'un outil d'IA génératif conçu spécifiquement pour le contexte des services juridiques (comme Lexis+ AI ou CoCounsel de Thomson Reuters)³², puisque ces derniers se servent de techniques différentes et ont accès à des bases de données distinctes. Ainsi, à mesure que ces nouvelles technologies évolueront et s'amélioreront pour combler ces lacunes, la RDO devra évaluer avec souplesse le bien-fondé du recours à chacun des outils.

3.1.3 Transparence

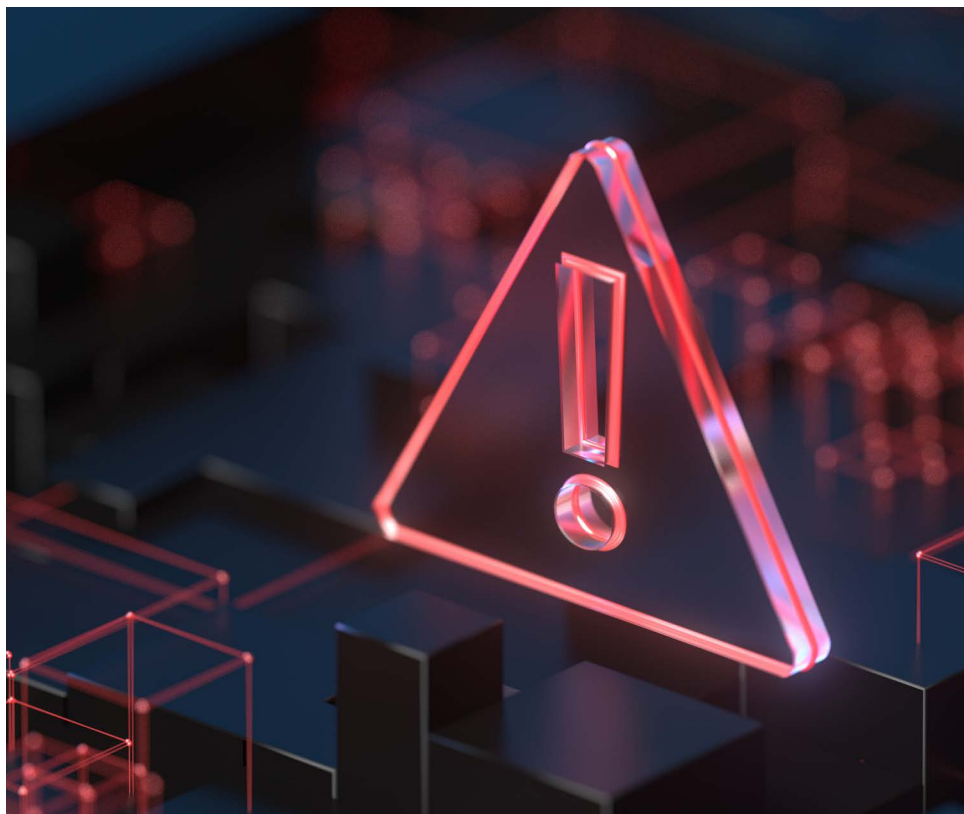
Bien que, comme il en sera question dans la section 4.2.1, le risque d'hallucinations puisse être réduit à un certain degré en recourant à des outils d'IA conçus spécifiquement pour l'industrie juridique, ce risque ne peut être totalement écarté. En conséquence, la transparence des auteurs quant à l'étendue de leur utilisation des technologies d'IA générative dans leurs soumissions aiderait la RDO à préserver l'exactitude et l'intégrité de ses publications.

Comme nous l'avons évoqué dans la section 2.2, de nombreuses revues non juridiques ont traité de la question de la transparence en exigeant la divulgation de l'utilisation de l'IA dans leurs politiques respectives. Certaines d'entre elles imposent aux auteurs de divulguer, dans la section réservée à leur méthodologie, quels outils spécifiques d'IA ils ont utilisés et comment ils ont été utilisés, tandis que d'autres vont jusqu'à exiger que les auteurs téléchargent, à titre de documents complémentaires à leur soumission, la totalité des résultats générés par l'IA. La RDO devrait élaborer une politique similaire pour garantir que les auteurs soient totalement transparents quant aux outils d'IA utilisés, la façon dont ils l'ont été et les mesures prises pour vérifier l'exactitude des résultats ainsi générés.

3.2 Questions juridiques

3.2.1 Préoccupations en matière de confidentialité

L'IA générative soulève plusieurs préoccupations en matière de confidentialité que les revues de droit comme la RDO doivent prendre en considération. Pour générer des textes convaincants, les modèles d'IA ont besoin d'« entraînement » à partir de bases de données volumineuses qui peuvent contenir des quantités importantes de renseignements sensibles et personnels³³. L'obtention de ces données passe souvent par une « extraction de données » massive en ligne, un processus qui peut collecter des renseignements tels des images de visages et des données médicales³⁴. Cette méthode de collecte de données soulève la question de savoir si les revues de droit pourraient recevoir des renseignements sensibles et personnels à partir des résultats générés par les outils d'IA. À l'inverse, si des revues de droit versent elles-mêmes des soumissions dans des outils d'IA dans le cadre de leur processus de révision, elles pourraient contrevenir à leur politique de confidentialité.



Le Commissariat à la protection de la vie privée du Canada [ci-après « CPVP »] a reconnu que l'extraction de données soulève des préoccupations liées aux cyberattaques, à l'usurpation d'identité, au profilage, à la surveillance ou à la collecte non sollicitée de renseignement à des fins politiques³⁵. De plus, les renseignements personnels d'individus appartenant à des groupes vulnérables, comme les enfants, pourraient aussi être extraits pour entraîner les modèles d'IA. Le CPVP a déclenché une enquête conjointe avec ses homologues provinciaux à la suite de plaintes contre OpenAI pour violation des lois sur la protection de la vie privée et des données. Au moment de rédiger le présent rapport, l'enquête est toujours en cours³⁶.

3.2.2 Incidences sur la propriété intellectuelle

L'IA générative soulève également des questions relatives à la propriété intellectuelle, plus spécifiquement des préoccupations en matière de droit d'auteur, en ce qui a trait à la fois à ce qui est versé dans les outils d'IA et à ce qui en ressort. La question à l'étape du téléchargement de données survient lorsque les modèles d'IA copient

illégalement des travaux lors de leur intégration à leur base de données durant l'«entraînement». Au moment de la sortie, la question survient lorsque l'IA enfreint les lois sur le droit d'auteur en générant du contenu qui ressemble à des travaux protégés. Ainsi, la préoccupation principale pour les revues de droit découle du fait que les auteurs pourraient recourir à des outils d'IA qui génèrent des textes contrevenant au régime sur le droit d'auteur du Canada. Des tiers pourraient également être en mesure d'obtenir un accès interdit à des soumissions non publiées et confidentielles si un rédacteur ou rédactrice [ci-après «rédacteur»] ou un évaluateur externe les téléchargeait dans un outil d'IA. En outre, il se pourrait que les outils d'IA copient des publications de la RDO et génèrent des résultats similaires à ses manuscrits protégés et exclusifs.

En 2023, le gouvernement du Canada a lancé une consultation publique sur le droit d'auteur et l'IA générative en vue de possibles modifications à la *Loi sur le droit d'auteur*³⁷. La consultation a porté principalement sur «l'utilisation d'œuvres protégées par le droit d'auteur aux fins de l'entraînement des systèmes d'IA ; la titularité et la propriété des droits en ce qui concerne le contenu produit par l'IA ; la responsabilité, particulièrement si le contenu produit par l'IA viole les droits d'auteur d'œuvres existantes»³⁸. Au moment d'écrire le présent rapport, la consultation est terminée, mais les mémoires détaillés présentés au gouvernement n'ont pas encore été rendus publics. Des modifications au régime canadien sur le droit d'auteur liées à l'IA pourraient avoir une incidence sur les politiques de la RDO sur l'IA. Il est donc recommandé que la RDO continue de rester à l'affût des modifications apportées à la *Loi sur le droit d'auteur* de même que des enquêtes du CPVP pour garantir qu'elle ne recourt pas aux outils d'IA en contravention des lois sur la protection de la vie privée et sur les données ou qu'elle n'autorise pas un tel recours.