

Nouvelle politique sur l'utilisation des outils d'intelligence artificielle dans les manuscrits soumis au *JAMC*

Matthew B. Stanbrook MD PhD, Meredith Weinhold MA, Diane Kelsall MD MEd

■ Citation : *CMAJ* 2023 July 24;195:E958-9. doi : 10.1503/cmaj.230949-f

Voir la version anglaise de l'article ici : www.cmaj.ca/lookup/doi/10.1503/cmaj.230949

Les récentes avancées majeures dans le domaine des grands modèles de langage — des algorithmes d'intelligence artificielle (IA) générative sophistiqués entraînés à l'aide d'un volume colossal de données sur la langue — ont donné plus largement accès à des outils de rédaction tels que ChatGPT, le populaire agent conversationnel d'OpenAI, capables d'analyser un texte et de produire du nouveau contenu d'après les requêtes fournies par les personnes qui les utilisent. Cette technologie a des implications immédiates et sérieuses pour les universitaires qui rédigent des articles et pour les responsables des revues qui les publient. En mai 2023, le Comité international des rédacteurs de revues médicales (ICMJE, pour *International Committee of Medical Journal Editors*) a mis à jour ses recommandations sur les travaux de recherche publiés dans des revues médicales pour y ajouter des directives particulières concernant les technologies assistées par IA¹. Comme le *JAMC* suit les recommandations de l'ICMJE, ces nouvelles recommandations seront désormais appliquées à tous les manuscrits qui lui sont soumis.

Les grands modèles de langage ont une puissante capacité de recherche et d'exploitation des informations contenues dans leurs données d'entraînement, qu'ils peuvent reformuler dans une panoplie de styles et de formats, au gré des requêtes. Ces outils peuvent générer des idées ou un plan pour la rédaction d'un manuscrit universitaire, voire des articles entiers. Puisque les outils d'IA modernes peuvent acquérir une capacité impressionnante d'imitation du discours et des styles rédactionnels humains, les textes générés peuvent paraître très semblables au fruit du travail d'un être humain et véhiculer un semblant d'exactitude et d'autorité en plus de susciter un lien émotionnel.

Cependant, cette impression n'est qu'une illusion. En raison du mécanisme de compression et de stockage des immenses ensembles de données utilisés par les ordinateurs, les grands modèles de langage procèdent essentiellement à une estimation des informations qu'ils récupèrent et qu'ils compilent, plutôt qu'à une reproduction rigoureuse de ces dernières, un peu comme si on essayait de reconstituer le texte exact d'une conférence sans y

avoir assisté en se basant sur les notes en style télégraphique de quelqu'un d'autre. De ce fait, les textes générés par ces grands modèles de langage sont susceptibles de contenir beaucoup d'erreurs, une grande partie du contenu, comme les références, est parfois créé de toutes pièces, et les passages qui sont reproduits avec exactitude peuvent être considérés comme du plagiat².

Les algorithmes d'IA actuels sont incapables de distinguer correctement le vrai du faux dans les données avec lesquelles ils ont été entraînés ni de reconnaître les cas où ils fabriquent de l'information (on parle alors d'« hallucination ») plutôt que de la reproduire fidèlement, contribuant au manque de fiabilité des textes générés. De plus, comme les versions actuelles des grands modèles de langage ne permettent pas la mise à jour en temps réel de leurs données, les textes produits sont susceptibles de contenir de l'information caduque. Par exemple, ChatGPT a été entraîné à partir de l'information présente sur Internet en 2021³, ce qui remonte pour ainsi dire à la préhistoire dans certains domaines de recherche où l'état des connaissances change rapidement.

Enfin, tous les arguments susmentionnés présupposent que l'intention des auteurs et des autrices est de viser l'exactitude et l'honnêteté. Cela dit, l'IA s'accompagne d'un potentiel sans précédent d'inconduite scientifique commise par des personnes peu scrupuleuses, qui pourraient l'utiliser pour rédiger des articles contenant des informations fausses ou fabriquées présentées de façon convaincante. Une telle approche serait toutefois bien imprudente, puisque la mise en marché d'outils capables de détecter l'utilisation des technologies assistées par IA pour la rédaction d'articles scientifiques, déjà en développement, paraît imminente⁴.

Pour toutes ces raisons, l'ICMJE et le *JAMC* exigent désormais que les auteurs et autrices qui ont eu recours à des technologies assistées par IA pour la rédaction d'une quelconque partie d'un manuscrit en fassent la déclaration (encadré 1). Cette dernière doit se retrouver à la fois sur le manuscrit et sur la lettre d'accompagnement au moment de la soumission, et les personnes ayant participé à la rédaction pourraient avoir à fournir des détails concernant la façon dont l'IA a été utilisée. Les technologies assistées par IA ne

Encadré 1 : Exigences concernant la déclaration de l'utilisation de technologies assistées par intelligence artificielle dans les manuscrits soumis au JAMC (selon les recommandations du Comité international des rédacteurs de revues médicales [International Committee of Medical Journal Editors]¹)

- Le JAMC exige que les auteurs et autrices, au moment de soumettre leur manuscrit, déclarent tout recours à des technologies assistées par intelligence artificielle (IA) (p. ex., grands modèles de langage, agents conversationnels ou générateurs d'images par IA) dans tout aspect de la création du travail soumis. Les personnes ayant participé à la rédaction devront décrire la nature de l'utilisation qui a été faite de ces outils dans la lettre d'accompagnement ainsi que dans le manuscrit lui-même.
- Les technologies assistées par IA ne doivent pas figurer en tant qu'autrices ou co-autrices d'un manuscrit.
- Les technologies assistées par IA ne doivent pas être citées parmi les références ou autres sources primaires ni parmi les auteurs et autrices d'une référence.
- Les auteurs et autrices du genre humain assument l'entière responsabilité de tout matériel soumis provenant de l'utilisation de technologies assistées par IA et se portent garants notamment de l'exactitude, de l'intégralité et de la fidélité des informations présentées.
- Les auteurs et autrices doivent être en mesure de certifier que l'article soumis ne contient pas de contenu plagié, y compris dans le texte et dans les images générés par les technologies assistées par IA, et doivent veiller à ce que toutes les sources soient citées de façon appropriée, y compris pour les citations complètes, le cas échéant.
- Les pairs réviseurs ne doivent pas téléverser de manuscrits du JAMC sur un logiciel ni sur tout autre outil assisté par IA, dans tous les contextes où le respect de la confidentialité ne peut être garanti.
- Dans les cas où des technologies assistées par IA peuvent être utilisées sans risque de nuire à la confidentialité, les pairs réviseurs doivent en faire la déclaration et préciser au JAMC la nature de l'utilisation de ces outils. Ces personnes demeurent responsables d'assurer l'exactitude, l'intégralité et l'objectivité de tout contenu généré par une IA figurant dans leur évaluation.

doivent pas figurer parmi les noms d'auteurs et d'autrices, car elles sont inaptes à répondre à plusieurs critères de l'ICMJE à ce sujet, notamment être en mesure d'assumer la responsabilité des travaux publiés, de déclarer les intérêts concurrents et d'adhérer aux ententes sur le droit d'auteur et aux contrats de licence¹. Des êtres humains doivent assumer l'entière responsabilité de confirmer que le contenu généré par IA est fidèle et exempt d'erreurs, d'informations fabriquées ou de plagiat. De même, les technologies assistées par IA ne peuvent en aucun cas être citées en tant que sources primaires, puisque l'IA reprend uniquement (souvent de façon inexacte) des informations tirées d'autres sources, dont l'origine et la qualité sont parfois inconnues.

Sachant que les technologies assistées par IA se révèlent très efficaces pour traiter et analyser des textes, les personnes qui acceptent de réviser des manuscrits en tant que pairs réviseurs pourraient être tentées d'utiliser ces outils pour les aider dans le processus d'examen. Toutefois, les recommandations de l'ICMJE l'interdisent dans de nombreux cas. Plusieurs outils d'IA conservent

les données du contenu téléversé pour leur développement continu. Le fait de téléverser un manuscrit inédit en tout ou en partie sur un tel outil constitue une infraction aux exigences en matière de confidentialité que les pairs réviseurs sont tenus de respecter.

Le recours occasionnel à des outils d'IA courants moins élaborés, comme les outils de correction des erreurs d'orthographe et de grammaire dans de nombreux logiciels de traitement de texte modernes, n'ont pas à être déclarés. De plus, les recherches destinées à étudier la nature et les effets de l'IA sont, bien entendu, permises, moyennant une description appropriée de la méthodologie adoptée.

Le but de ces nouvelles recommandations est de faire preuve de prudence et de remédier immédiatement à certains préjudices anticipés de l'avènement des technologies assistées par IA dans les publications scientifiques et savantes. Bien que ces outils puissent présenter des avantages théoriques pour la pratique clinique, les tâches administratives et la vie de tous les jours, les adopter trop rapidement comporte un risque de conséquences graves et imprévisibles⁵. La tendance des versions actuelles des technologies assistées par IA à générer des erreurs et de fausses informations est inadmissible dans le domaine de la science, qui repose fondamentalement sur l'exactitude, la précision et la reproductibilité; tous les chercheurs, les chercheuses et les universitaires qui seraient tentés de recourir à de tels outils dans leur travail doivent en tenir compte. La transparence, la responsabilisation et la prudence sont essentielles dans le contexte du développement et de l'utilisation des technologies assistées par IA. Peu importe le degré de sophistication atteint, ces outils demeurent fondés sur des algorithmes conçus par des êtres humains; ils ne pourront jamais remplacer la créativité, la curiosité et l'ingéniosité humaines, qui sont la pierre angulaire de la science et de l'érudition.

Références

1. *Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals*. International Committee of Medical Journal Editors; mis à jour mai 2023. Accessible ici : <https://www.icmje.org/recommendations/> (consulté le 28 juin 2023).
2. Dien J. Editorial: Generative artificial intelligence as a plagiarism problem. *Biol Psychol* 2023 June 24 [cyberpublication avant impression]. doi : 10.1016/j.biopsycho.2023.108621.
3. Chow JCL, Sanders L, Li K. Impact of ChatGPT on medical chatbots as a disruptive technology. *Front Artif Intell* 2023;6:1166014.
4. Kirmani AR. Fantastic (AI) bots and how to catch them. *MRS Bull* 2023;48:314-8.
5. Pujari S, Reis A, Zhao Y, et al. Artificial intelligence for global health: cautious optimism with safeguards. *Bull World Health Organ* 2023;101:364-364A.

Intérêts concurrents : www.cmaj.ca/staff

Affiliations : www.cmaj.ca/staff

Propriété intellectuelle du contenu : Il s'agit d'un article en libre accès distribué conformément aux modalités de la licence Creative Commons Attribution (CC BY-NC-ND 4,0), qui permet l'utilisation, la diffusion et la reproduction dans tout médium à la condition que la publication originale soit adéquatement citée, que l'utilisation se fasse à des fins non commerciales (c.-à-d., recherche ou éducation) et qu'aucune modification ni adaptation n'y soit apportée. Voir : <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.fr>

Correspondance : Rédaction du JAMC, editorial@cmaj.ca