

TRIBUNE/TRIBUNE



Intelligence artificielle et évaluation des articles scientifiques: promesses et vigilance

Artificial intelligence and the evaluation of scientific articles: promise and caution

Sébastien LARRÉCHÉ

Dans son éditorial d'avril 2026, Jean-Philippe Chippaux nous alertait sur les limites de l'intelligence artificielle (IA) générative utilisée à des fins de rédaction d'articles scientifiques [1]. Cependant, les auteurs ne sont pas les seuls concernés par cette révolution qui s'impose progressivement dans l'ensemble de l'écosystème de la publication scientifique. Les revues médicales sont désormais confrontées à une question centrale: comment intégrer ces outils sans compromettre la qualité, l'indépendance et l'intégrité du processus éditorial?

Les opportunités offertes par l'IA dans l'évaluation de manuscrits sont multiples, mais concernent des fonctions très différentes. Dans un contexte de difficulté croissante à recruter des relecteurs, elle peut contribuer à identifier des experts pertinents et à détecter d'éventuels conflits d'intérêt. D'autres outils assistent les éditeurs dans la vérification des références bibliographiques ou de la conformité d'un manuscrit aux recommandations internationales (CONSORT, STROBE, STARD, PRISMA, ...). Certains modèles sont capables d'identifier des incohérences méthodologiques ou statistiques tandis que d'autres contribuent à la détection du plagiat, d'images manipulées ou de contenus générés par IA. L'IA peut également apporter une aide linguistique en améliorant la qualité rédactionnelle des rapports d'évaluation ou en reformulant des commentaires éditoriaux. Enfin, elle peut faciliter le travail éditorial en produisant une synthèse des avis des relecteurs afin d'éclairer la décision finale d'acceptation ou de refus d'un manuscrit [2].

Des expériences récentes menées dans des revues internationales ont montré qu'un processus hybride associant relecteurs humains et assistance par IA pouvait raccourcir significativement les délais d'évaluation. Ainsi, NEJM AI est à présent capable de rendre une première décision éditoriale en moins de sept jours [4]. L'IA générative est particulièrement performante pour les tâches répétitives ou techniques: vérification de tableaux, cohérence entre protocole et analyse statistique, repérage d'omissions méthodologiques ou reformulation de commentaires éditoriaux.

In his April 2026 editorial, Jean-Philippe Chippaux warned us about the limitations of generative artificial intelligence (AI) used for writing scientific articles [1]. However, authors are not the only ones affected by this revolution, which is gradually taking hold throughout the entire scientific publishing ecosystem. Medical journals are now faced with a central question: how can they integrate these tools without compromising the quality, independence, and integrity of the editorial process?

AI offers numerous opportunities for manuscript evaluation, but these opportunities involve very different functions. Given the growing difficulty in recruiting reviewers, AI can help identify relevant experts and detect potential conflicts of interest. Other tools assist publishers in verifying bibliographic references or ensuring a manuscript's compliance with international guidelines (CONSORT, STROBE, STARD, PRISMA, etc.). Some models are capable of identifying methodological or statistical inconsistencies, while others help detect plagiarism, manipulated images, or AI-generated content. AI can also provide linguistic support by improving the writing quality of review reports or by rephrasing editorial comments. Finally, it can streamline the editorial process by generating a summary of reviewers' opinions to inform the final decision to accept or reject a manuscript [2].

Recent experiments published in international journals have demonstrated that a hybrid process combining human reviewers and AI assistance can significantly reduce review times. For instance, NEJM AI can now render an initial editorial decision in under seven days [4]. Generative AI is particularly effective at repetitive or technical tasks, such as verifying tables, ensuring consistency between the protocol and statistical analysis, identifying methodological omissions, and rephrasing editorial comments. In an environment where publication timelines are becoming a strategic issue, these efficiency gains cannot be ignored.

Dans un environnement où les délais de publication deviennent un enjeu stratégique, ces gains d'efficacité ne peuvent être ignorés.

Cependant, ces différentes fonctions ne présentent, ni le même niveau de maturité, ni le même degré de fiabilité. Les performances observées dépendent largement des plateformes utilisées, des données sur lesquelles elles ont été entraînées et des modalités de leur déploiement. Les tâches relativement standardisées, comme certaines vérifications formelles ou linguistiques, sont généralement plus robustes que celles nécessitant une interprétation scientifique complexe, une appréciation méthodologique fine ou un jugement éditorial.

Les modèles génératifs ne raisonnent pas comme des experts scientifiques ; ils produisent des réponses probabilistes fondées sur des *corpus* d'entraînement parfois incomplets, biaisés ou obsolètes. L'intégration d'articles scientifiques de faible qualité dans les *corpus* d'apprentissage peut progressivement dégrader les performances des modèles, créant un véritable cercle vicieux. Lorsque l'IA est utilisée dans le cadre du processus éditorial, elle devrait donc reposer sur des environnements sécurisés, privés et maîtrisés par les institutions. La reproductibilité reste insuffisante, le téléversement d'un même manuscrit à plusieurs reprises peut conduire à des rapports d'évaluation différents, voire contradictoires.

La question de la confidentialité est également cruciale. Le processus d'évaluation par les pairs repose sur la protection des manuscrits non publiés et sur la confidentialité des échanges entre auteurs, relecteurs et éditeurs. Téléverser un manuscrit dans un outil d'IA externe peut exposer des données scientifiques inédites, voire des informations sensibles. Cette problématique est particulièrement importante dans le champ de la santé internationale, où les travaux concernent souvent des populations vulnérables et des données issues de terrains sensibles.

Au-delà des enjeux techniques, l'IA soulève une interrogation plus fondamentale sur la nature même de l'évaluation scientifique. La relecture d'un article ne consiste pas uniquement à identifier des erreurs ou à vérifier des critères méthodologiques. Elle mobilise également une expertise contextuelle, une appréciation de la pertinence clinique ou de santé publique, ainsi qu'une capacité à évaluer l'originalité et l'impact potentiel des travaux. Ces dimensions reposent encore largement sur le jugement humain.

Le risque de biais d'automatisation, du fait d'une confiance excessive dans les recommandations d'origine algorithmique, doit également être pris en compte. La pression éditoriale et

However, these functions vary in maturity and reliability. Observed performance largely depends on the platforms used, the data on which they were trained, and the terms of their deployment. Tasks that are relatively standardized, such as certain formal or linguistic checks, are generally more robust than tasks requiring complex scientific interpretation, nuanced methodological assessment, or editorial judgment.

Generative models do not reason like scientific experts; they produce probabilistic responses based on training corpora that are sometimes incomplete, biased, or obsolete. Including low-quality scientific articles in training corpora can gradually degrade model performance, creating a vicious cycle. Therefore, when AI is used as part of the editorial process, it should be based on secure, private environments controlled by institutions. Reproducibility remains insufficient: uploading the same manuscript multiple times can result in different—or even contradictory—evaluation reports.

The issue of confidentiality is also crucial. The peer-review process depends on protecting unpublished manuscripts and maintaining the confidentiality of communications between authors, reviewers, and editors. Uploading a manuscript to an external AI tool can expose unpublished scientific data or sensitive information. This issue is particularly significant in global health research, which often involves vulnerable populations and data from sensitive settings.

Beyond technical challenges, AI raises a more fundamental question about the nature of scientific evaluation itself. Peer review of an article is not merely about identifying errors or verifying methodological criteria. It also requires contextual expertise, an assessment of clinical or public health relevance, and the ability to evaluate the research's originality and potential impact. These aspects still rely heavily on human judgment.

The risk of automation bias, due to excessive reliance on algorithm-generated recommendations, must also be considered. Editorial pressure and the increasing number of submissions could lead to the temptation to gradually delegate certain decisions to automated systems. However, AI tool performance varies greatly depending on the specific function: some applications provide useful assistance, while others are insufficiently reliable or raise significant ethical questions. This could lead to the standardization of scientific evaluation and reduce the diversity of methodological and intellectual approaches. There is a real risk of

la multiplication des soumissions pourraient conduire à la tentation de déléguer progressivement certaines décisions à des systèmes automatisés. Or les performances des outils d'IA sont très hétérogènes selon les fonctions considérées : si certaines applications apportent une aide utile, d'autres restent insuffisamment fiables ou soulèvent des questions éthiques importantes. Une telle évolution pourrait conduire à une standardisation de l'évaluation scientifique et à une réduction de la diversité des approches méthodologiques ou intellectuelles. Il y a là un réel risque d'érosion de l'expertise humaine.

Pour ces raisons, plusieurs grandes maisons d'édition scientifique ont adopté des politiques restrictives concernant l'usage de l'IA dans l'évaluation par les pairs. Certaines interdisent explicitement le téléversement de manuscrits dans des outils génératifs ; d'autres autorisent un usage limité et transparent, à condition que la responsabilité finale demeure entièrement humaine [3]. Un consensus semble aujourd'hui émerger : l'IA peut être un outil d'assistance, mais ne peut constituer un relecteur autonome.

Dans les revues de santé internationale, ces débats prennent une dimension particulière. Les inégalités d'accès aux ressources scientifiques, les barrières linguistiques et la faible disponibilité de relecteurs spécialisés pourraient rendre les outils d'IA particulièrement attractifs. Ils pourraient contribuer à renforcer l'inclusivité de la publication scientifique et à soutenir des équipes disposant de moyens limités. Mais ces bénéfices potentiels ne doivent pas masquer les risques d'une dépendance technologique accrue, ni les biais potentiellement intégrés dans des modèles majoritairement développés dans des contextes occidentaux [5].

L'enjeu des prochaines années ne sera probablement pas de choisir entre intelligence humaine et intelligence artificielle, mais d'organiser une collaboration raisonnée entre les deux. Les outils génératifs peuvent améliorer certaines dimensions du travail éditorial, à condition d'être utilisés avec transparence, esprit critique et supervision humaine constante. Les revues scientifiques ont tout intérêt à privilégier des solutions institutionnelles sécurisées, fonctionnant dans des environnements garantissant la confidentialité des manuscrits et la protection des données. Il est donc nécessaire dès à présent pour *Médecine Tropicale et Santé Internationale* de définir des règles claires pour protéger l'intégrité de l'évaluation par les pairs et promouvoir un usage responsable de l'IA.

eroding human expertise here.

For these reasons, several major scientific publishers have adopted restrictive policies regarding the use of AI in peer review. Some explicitly prohibit uploading manuscripts to generative tools, while others allow limited, transparent use provided that humans remain entirely responsible [3]. A consensus is now emerging: AI can serve as a support tool but cannot function as an autonomous reviewer.

In international health journals, these debates take on a particular dimension. Inequalities in access to scientific resources, language barriers, and the limited availability of specialized reviewers could make AI tools appealing. These tools could enhance the inclusivity of scientific publishing and support teams with limited resources. However, the potential benefits must not obscure the risks of increased technological dependence or the biases that may be embedded in models developed primarily in Western contexts [5].

In the coming years, the challenge will likely not be choosing between human and artificial intelligence, but rather organizing a thoughtful collaboration between the two. Provided they are used with transparency, critical thinking, and constant human oversight, generative tools can improve certain aspects of editorial work. Scientific journals should prioritize secure institutional solutions operating in environments that guarantee manuscript confidentiality and data protection. Therefore, it is necessary for *Médecine Tropicale et Santé Internationale* to establish clear rules to protect the integrity of peer review and promote the responsible use of AI now.

Auteur / Author

Sébastien LARRÉCHÉ

Service de biologie médicale, Hôpital national d'instruction des armées Bégin, 69 avenue de Paris, 94160 Saint-Mandé, France

École du Val-de-Grâce, Académie de santé des armées, 1 place Alphonse Laveran, 75005 Paris, France

UMR-S1144, Université Paris Cité, 4 avenue de l'Observatoire, 75006, Paris, France

ORCID : 0000-0002-8961-7132

slarreche@hotmail.fr

Références / References

1. Chippaux J-P. Apport et limites de l'intelligence artificielle dans la rédaction d'articles scientifiques : attention à l'imposture. *Med Trop Sante Int.* 2026. doi: 10.48327/mtsi.v6i2.2026.843.
2. Hosseini M, Horbach SPJM. Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Res Integr Peer Rev.* 2023 May 18;8(1):4. doi: 10.1186/s41073-023-00133-5.
3. Li ZQ, Xu HL, Cao HJ, Liu ZL, Fei YT, Liu JP. Use of artificial intelligence in peer review among Top 100 medical journals. *JAMA Netw Open.* 2024 Dec 2;7(12):e2448609. doi: 10.1001/jamanetworkopen.2024.48609.
4. Manrai AK, Ouyang D, Hogan JW, Kohane IS. Accelerating science with Human+AI review. *NEJM AI.* 2025 ; 2(12) doi: 10.1056/AIe2501175.
5. Nabavi A, Safari F, Shmoury AH, Tabet S, Perdomo-Luna C, Celi LA. Artificial intelligence in scholarly peer review: a scoping review of applications, risks, and governance challenges. *Int J Med Inform.* 2026 Jul 1;214:106418. doi: 10.1016/j.ijmedinf.2026.106418.