EDITORIAL

AI IN MEDICINE

Artificial Intelligence in Peer Review

Roy H. Perlis, MD, MSc; Dimitri A. Christakis, MD, MPH; Neil M. Bressler, MD; Dost Öngür, MD, PhD; Jacob Kendall-Taylor, BA; Annette Flanagin, RN, MA; Kirsten Bibbins-Domingo, PhD, MD, MAS

Like David Foster Wallace's tale of 2 fish that don't notice the water they're swimming in because it's all around them,[1] the medical community has at times taken for granted peer review and its centrality to evidence-based medicine. But the peer review process is under increasing pressure: the number of journals seeking peer reviewers and the number of manuscripts requiring peer review continue to grow. The pressures on peer reviewers themselves have grown in parallel, with many complaining of "reviewer fatigue" or simply opting out of reviewing altogether.[2] Journal editors have acknowledged longstanding critiques that the peer review process is inefficient, slows publication of important work, introduces potential for reviewer bias and inconsistency, and fails to prevent publication of poor-quality or fraudulent research.[3,4]

Strategies to Address Problems With Peer Review

One response has been to increase educational and experiential training in the fundamentals of peer review through mentoring opportunities, a strategy embraced by *JAMA* and the JAMA Network journals[5] and many other journals and publishers.[6-8] Other options have been to use software to match reviewers with manuscripts via keywords or other metadata. Still, these innovations may only marginally increase the pool of qualified and available peer reviewers and are not likely to accelerate the review process or address its other weaknesses.

Another strategy is to apply the capacity of artificial intelligence (AI) to assist with the peer review process. Given the ability of large language models (LLMs) to efficiently summarize text, extract key features, and facilitate an interactive question-and-answer process, AI may help reviewers to be more efficient and streamline some review processes such as identifying required elements in manuscripts.[9,10] Augmenting this process has the potential to enhance quality, but testing and standards are needed to confirm these benefits. Recognizing the potential risks of the use of AI models in the review process, we and other journal editors have elected to tread cautiously. In 2023, the JAMA Network issued guidance for author use of AI, prohibiting the inclusion of AI tools as authors and requiring authors to disclose use of such tools and take responsibility for such use.[11] This policy was updated to include use of AI by peer reviewers.[12] The policy prohibits reviewers from uploading confidential manuscripts to AI tools that could breach the confidentiality agreement between journals and authors. It requires reviewers to disclose if they use AI as a resource during their review.

The International Committee of Medical Journal Editors (ICMJE) has adopted similar guidance, noting that LLMs cannot be authors of manuscripts because they cannot "be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved."[13] A computer currently cannot be held accountable for anything in the absence of a human to hold it to account, and a computer lacks the capacity for ethical judgment and integrity.[14] In the same vein, ICMJE warns that "[r]eviewers should be aware that AI can generate authoritative-sounding output that can be incorrect, incomplete, or biased."[13]

As AI continues to evolve at a rapid pace, the question no longer is whether AI will enter peer review, but how to efficiently and effectively harness and manage its deployment in a manner that is fair to authors and valuable in the editorial process. Journals have evaluated a panoply of strategies.[15] Some specific reviewer or editorial tasks may be readily automated. For example, LLMs can instantaneously generate summaries of manuscripts, determine completeness of adherence with manuscript submission and publication checklists, flag missing reporting guideline items, and compare data within abstracts, text, and tables for internal consistency. However, these models can produce false-positives (eg, flagging inconsistencies incorrectly) and false-negatives (eg, ignoring certain errors for which they have not been trained), such that they may not yet result in efficiencies for editors and reviewers. Yet, the performance of these models will continue to improve. As with plagiarism-checking software, the fact that such tools are imperfect may not preclude their careful application.

Other key functions currently performed by humans may be more challenging to replace. For example, can AI evaluate a figure to determine whether it synthesizes and represents the data accurately and efficiently for the human reader? Can AI accurately assess the novelty of a research report, which requires prior literature and methodological rigor, or identify a key insight that might be determinative in the editorial decision? Estimating clinical importance is even more challenging because clinical relevance is tethered to clinical context, clinician expertise, and patient need and can differ across disciplines and cultures. At present and at least in the near future, these insights will be difficult to automate.

The JAMA Network journals will explore multiple potential hybrid strategies to incorporate AI into editorial assessment and peer review while retaining accountable humans in the loop. Editorial decisions at JAMA Network have always been made by editors, with support of peer review that provides critical insights to aid in editorial decisions and yield even better manuscripts. Any effort to incorporate AI in the editorial

assessment and peer review processes will focus on how such tools can aid editors in decision-making, reviewers in providing timely assessment and feedback, and ultimately authors in improving their manuscripts. We see our approach as analogous to driver-assistance technologies, beginning with adaptive cruise control or blind spot detection, but when it comes to peer review and editorial decisions, we are not considering fully autonomous driving. Editors and reviewers will not be taking their hands off the wheels or their eyes off the road. Even as AI-augmented workflows give each human reviewer an AI copilot for summarization and checking for quality of reporting, the human reviewer still takes ultimate responsibility for the assessment and decisions.

We believe that, with time, parallel workflows that incorporate an AI-generated review to consider alongside human critiques will shorten time to decision and improve the overall quality of review. In this model, an AI reviewer might serve a specialized purpose (eg, assessing fidelity to a study protocol or identifying incomplete reporting or common methodologic problems), in the same way many journals have biostatistical reviewers to advise on study design and statistical analysis. Here, too, we are not considering the replacement of human statistical reviewers but looking for ways to make their assessments even more valuable and efficient. Meta-review systems may apply AI to synthesize and summarize multiple human reports into a single, structured set of actionable recommendations for editors, flagging inconsistencies and facilitating editorial decision-making. However, ultimate responsibility will rest with the editors.

Challenges With Use of AI in Peer Review

Although authors may welcome faster decisions potentially facilitated by incorporating AI-augmented reviews, AI review also introduces new risks. In some instances, authors may recognize an AI-generated peer review and conclude that an editor was only superficially involved, undermining confidence in the entire process. Another important concern is breach of confidentiality if manuscripts are uploaded to public language models, where materials may be applied to train future models, contribute to leakage of identifiers or unpublished data, or even be hijacked for unintended or nefarious use. Moreover, while some reviewers may have access to secure, walled garden models, or run these models locally, this process still risks creating classes of reviewers who do, or do not, have access to such tools, and editors have no way to assess reviewers' access or expertise. This inequity applies to science in general—some investigators are fortunate to be part of institutions with more resources, others not—but at least journals can avoid exacerbating them.

Another challenge is the need to address confabulation or hallucination, whereby language models yield seemingly correct but fictional references or other obvious or subtle errors. This problem may be worsening: frontier models may exhibit more confabulation than some earlier models.[16] The possibility that an AI reviewer may fabricate details is real and could lead to an increase, rather than a decrease, in human workload if reviewers and editors need to confirm every AI-generated assertion. There is precedent for this concern in the early use of AI in clinical medicine: while some studies have found that AI scribes can reduce clinician workload, others have not found such benefits, likely because more effort is required to check the scribes' work.[17] Continuous quality improvement assessments of these approaches will be needed. Otherwise, by introducing even subtle errors, AI tools risk increasing rather than decreasing workload.

The fairness of LLM-generated reviews also remains to be established. While bias in models is a frequent topic for discussion, including evidence that LLMs use enthusiastic language and exhibit positivity bias, the concern that certain topics or kinds of language may elicit harsher reviews has received less attention. That is, in addition to bias toward groups of people, it is possible that models will exhibit bias toward particular diseases, or methodologies, based on their initial training. However, models may be engineered to reduce bias more readily than human reviewers can control and disclose their own biases. Presently, LLMs hold no personal vendettas, they are not motivated by competition or jealousy, and they do not cling to a prevailing scientific theory simply because it concurs with their own research. They do, however, prefer their own writing to that of other models and often to human writing, a challenge that will also need to be overcome.[18]

There are other concerns. Reliance of generative AI in peer review may reduce a human reviewer's critical thinking, a form of "cognitive offloading" that lessens cognitive burden but also results in less critical assessment, an important hallmark of peer review.[10] Often the most valuable peer review includes an incisive, novel observation by an expert reviewer that would not necessarily be reflected in existing literature. LLMs trained on the existing literature may instead privilege incremental work, nudging science toward monoculture.[19] Moreover, authors may attempt to optimize manuscripts for AI algorithmic reviewers just as they may currently do for human reviewers, a form of reward hacking familiar from reinforcement-learning research, which could lead to competition between automated manuscript optimizers and AI manuscript reviewers. Indeed, a recent report indicates such reward hacking has already occurred in scientific meeting abstracts.[20]

Addressing the Challenges

In addressing these challenges, we, as editors of *JAMA* and the JAMA Network journals, plan to adopt the same perspective that we would bring to any new medical technology: empirical, scientific study to determine which methods yield the greatest quality improvement while maintaining safety. The upcoming 10th International Congress on Peer Review and Scientific Publication, co-organized by the JAMA Network, *The BMJ*, and Meta-research Innovation Center at Stanford (METRICS), will feature presentation and discussion of many studies conducted to address these issues.[21] We are beginning to explore the potential benefits of different approaches to applying AI in the editorial and peer review processes with the aims of improving efficiency, preserving fairness, minimizing risks, and promoting integrity and quality of scientific publication.

Our hope is that automating some aspects of peer review, at first, will help to relieve the need to complete rote tasks,

allowing scarcer human expertise to focus on aspects such as impact and significance, novelty, and clinical relevance. We endeavor to improve both the quality and efficiency of the peer review process, all while keeping our hands on the wheel and our eyes on the road. We believe it will be critical to maintain a human in the loop even as we seek to incorporate the strengths of AI-based review in our editorial process. In so doing, human editors will maintain full oversight, accountability, and responsibility for scientific rigor, standards, and editorial decisions.

ARTICLE INFORMATION

Author Affiliations: Editor in Chief, JAMA+ AI (Perlis); Editor, *JAMA Pediatrics* (Christakis); Editor, *JAMA Ophthalmology* (Bressler); Editor, *JAMA Psychiatry* (Öngür); Director of Editorial Systems, *JAMA* and the JAMA Network (Kendall-Taylor); Executive Managing Editor, *JAMA* and the JAMA Network (Flanagin); Editor in Chief, *JAMA* and the JAMA Network (Bibbins-Domingo).

Corresponding Author: Roy H. Perlis, MD, MSc, *JAMA* (Roy.Perlis@jamanetwork.org).

Published Online: August 28, 2025. doi:10.1001/jama.2025.15827

Conflict of Interest Disclosures: All authors are editors and staff at *JAMA* and the JAMA Network. Dr Perlis reported receiving personal fees for serving as a scientific advisor to Genomind, Circular Genomics, Psy Therapeutics, and Alkermes. Dr Öngür reported receiving personal fees from Neumora Inc, Guggenheim LLC, Boehringer Ingelheim, and Rapport Therapeutics. No other disclosures were reported.

REFERENCES

1. This is water. Kenyon College Alumni Bulletin. Accessed May 14, 2025. http://bulletin-archive.kenyon.edu/x4280.html

2. Fox CW, Albert AYK, Vines TH. Recruitment of reviewers is becoming harder at some journals: a test of the influence of reviewer fatigue at six journals in ecology and evolution. *Res Integr Peer Rev*. 2017;2(1):3. doi:10.1186/s41073-017-0027-x

3. Rennie D, Flanagin A. Three decades of Peer Review Congresses. *JAMA*. 2018;319(4):350-353. doi:10.1001/jama.2017.20606

4. Rennie D. Guarding the guardians: a conference on editorial peer review. *JAMA*. 1986;256(17):2391-2392. doi:10.1001/jama.1986.03380170107031

5. JAMA Network. JAMA Network Peer Review Academy. Accessed August 6, 2025. https://jamanetwork.com/pages/jama-network-peer-review-academy

6. Elsevier Researcher Academy. Certified Peer Reviewer course. Accessed August 6, 2025. https://researcheracademy.elsevier.com/navigating-peer-review/certified-peer-reviewer-course

7. Wiley. Journal reviewers. Accessed August 6, 2025. https://authorservices.wiley.com/Reviewers/journal-reviewers/index.html

8. *The BMJ*. Resources for reviewers. Accessed August 9, 2025. https://www.bmj.com/about-bmj/resources-reviewers

9. Bauchner H, Rivara FP. Use of artificial intelligence and the future of peer review. *Health Aff Sch*. 2024;2(5):qxae058. doi:10.1093/haschl/qxae058

10. Hoyt R, Limon A, Chang A. Generative AI and scientific manuscript peer review. *Intell Based Med*. 2025;11:100246. doi:10.1016/j.ibmed.2025.100246

11. Flanagin A, Bibbins-Domingo K, Berkwits M, Christiansen SL. Nonhuman "authors" and implications for the integrity of scientific publication and medical knowledge. *JAMA*. 2023;329(8):637-639. doi:10.1001/jama.2023.1344

12. Flanagin A, Kendall-Taylor J, Bibbins-Domingo K. Guidance for authors, peer reviewers, and editors on use of AI, language models, and chatbots. *JAMA*. 2023;330(8):702-703. doi:10.1001/jama.2023.12500

13. International Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Accessed August 9, 2025. https://www.icmje.org/recommendations/

14. Bressler NM. What artificial intelligence chatbots mean for editors, authors, and readers of peer-reviewed ophthalmic literature. *JAMA Ophthalmol*. 2023;141(6):514-515. doi:10.1001/jamaophthalmol.2023.1370

15. 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;7(12):e2448609. doi:10.1001/jamanetworkopen.2024.48609

16. Hsu J. AI hallucinations are getting worse—and they're here to stay. New Scientist. Accessed May 14, 2025. https://www.newscientist.com/article/2479545-ai-hallucinations-are-getting-worse-and-theyre-here-to-stay/

17. Haberle T, Cleveland C, Snow GL, et al. The impact of nuance DAX ambient listening AI documentation: a cohort study. *J Am Med Inform Assoc*. 2024;31(4):975-979. doi:10.1093/jamia/ocae022

18. Panickssery A, Bowman SR, Feng S. LLM evaluators recognize and favor their own generators. *arXiv*. Preprint published April 15, 2024. doi:10.48550/arXiv.2404.13076

19. Doshi AR, Hauser OP. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Sci Adv*. 2024;10(28):eadn5290. doi:10.1126/sciadv.adn5290

20. Sugiyama S, Eguchi R. 'Positive review only': researchers hide AI prompts in papers. Nikkei Asia. Accessed July 7, 2025. https://asia.nikkei.com/Business/Technology/Artificial-intelligence/Positive-review-only-Researchers-hide-AI-prompts-in-papers

21. Ioannidis JPA, Berkwits M, Flanagin A, Bloom T. Tenth International Congress on Peer Review and Scientific Publication: call for abstracts. *JAMA*. 2024;332(17):1434-1436. doi:10.1001/jama.2024.18311