Review

Science

Is That Reliable and Feasible for AI as a Peer Reviewer?

Ethan Michael Reynolds*
University of California, Berkeley, 110 Sproul Hall #5800, Berkeley, CA 94720-5800, USA
*: All correspondence should be sent to: Dr. Ethan Michael Reynolds.
Author's Contact: Dr. Ethan Michael Reynolds, PhD., E-mail: emreynolds@berkeley.edu
DOI: https://doi.org/10.15354/si.26.re126
Funding: No funding source declared.
COI: The author declares no competing interest.
AI Declaration: The author affirms that artificial intelligence did not contribute to the process of preparing the work.

Peer review is the cornerstone of scientific quality control, yet it faces growing challenges including reviewer fatigue, bias, inconsistency, and escalating submission volumes. Artificial intelligence has recently been proposed as a potential tool—or even substitute—for human peer reviewers. This review article critically examines whether AI can reliably and feasibly function as a peer reviewer in scholarly publishing. Drawing on developments in natural language processing, machine learning, and automated evaluation systems, the article analyzes current capabilities, limitations, ethical concerns, and structural constraints. Rather than asking whether AI can replace human reviewers outright, this review evaluates where AI meaningfully contributes to review processes and where human judgment remains indispensable. The analysis suggests that AI shows promise in technical screening, methodological consistency checks, and bias reduction, but remains limited in conceptual novelty assessment, epistemic judgment, and ethical reasoning. Ultimately, AI is better positioned as a co-reviewer or decision-support system rather than an autonomous arbiter of scientific merit.

Keywords: Artificial Intelligence; Peer Review; Scientific Publishing; Research Integrity; Automation

Science Insights, January 30, 2026; Vol. 48, No. 1, pp.2123-2126.

PEER REVIEW occupies a paradoxical position in modern science. It is universally acknowledged as essential, yet widely criticized for being slow, inconsistent, opaque, and vulnerable to bias. As scientific output increases exponentially and the pool of qualified reviewers struggles to keep pace, the peer review system is under unprecedented strain (Obermeyer et al., 2019). Against this backdrop, artificial intelligence has emerged as a provocative candidate for reform. Can AI reliably and feasibly act as a peer reviewer? Or does this proposal misunderstand both the nature of peer review and the current limits of artificial intelligence? This review examines the question not as a binary choice, but as a spectrum of possible roles AI might play within the peer review ecosystem.

At its core, peer review serves multiple functions simultaneously. It evaluates technical correctness, methodological rigor, originality, relevance, clarity, ethical compliance, and potential impact. These dimensions are not purely objective, nor are they entirely subjective (Bedeian, 2004). Instead, they sit at the intersection of formal rules, disciplinary norms, and tacit expertise. Any assessment of AI as a peer reviewer must therefore ask not only whether AI can perform specific tasks, but whether it can engage meaningfully with the epistemic values that underlie scientific judgment.

From a technical perspective, AI has made remarkable progress in analyzing scientific texts. Modern language models can summarize manuscripts, detect plagiarism, identify statistical inconsistencies, flag missing methodological details, and assess adherence to reporting guidelines (Gerstner & König, 2023). Automated tools already assist editors by screening submissions for formatting issues, ethical disclosures, image manipulation, and basic statistical errors. In these domains, AI has demonstrated reliability that often exceeds human performance, particularly in consistency and speed. This has fueled optimism that AI could shoulder a larger share of peer review responsibilities.

Reliability, however, depends on what is being reviewed. AI excels at pattern recognition and rule-based evaluation, especially when clear standards exist. It can check whether sample sizes are reported, whether confidence intervals match p-values, or whether references are properly cited. It can compare manuscripts against vast corpora to identify similarities or anomalies (Birhane et al, 2021). In this sense, AI enhances reliability by reducing human oversight errors and variability. Human reviewers may miss details due to fatigue or time constraints, whereas AI systems do not tire.

Yet peer review is not merely a checklist exercise. One of its most valued functions is assessing novelty and significance (Blank, 1991). Determining whether a finding advances knowledge requires contextual understanding of a field's trajectory, unresolved debates, and conceptual blind spots. While AI can map citation networks and identify topical similarity, it struggles to distinguish incremental work from genuinely transformative ideas. Novelty often involves breaking existing patterns rather than conforming to them—precisely where pattern-based systems are weakest. This limitation raises concerns about over-reliance on AI reinforcing conservatism in science.

Feasibility also depends on transparency and explainability. Peer review is not only about making decisions, but about justifying them. Authors expect feedback that explains why a manuscript is flawed, how it can be improved, or why it may not be suitable for a particular journal. While AI can generate fluent critiques, the reasoning behind its judgments is often opaque. Without clear interpretability, AI-based reviews risk becoming authoritative yet unaccountable, undermining trust in the process. Reliability without explainability is insufficient in scholarly evaluation.

Bias presents a more complex picture. Human peer review is known to suffer from biases related to gender, institution, geography, language, and theoretical orientation. AI systems are often proposed as neutral alternatives (Horbach & Halffman, 2018). In practice, however, AI inherits biases from training data, which reflect historical inequities in publishing. If not carefully designed, AI reviewers may amplify existing biases by favoring dominant paradigms, prestigious institutions, or frequently cited topics. Reliability therefore hinges not on AI's objectivity, but on deliberate bias auditing and governance.

Ethical evaluation is another domain where feasibility is limited. Peer reviewers are expected to assess ethical considerations such as participant consent, data integrity, dual-use risks, and potential societal harm (COSE, 2018). These judgments often require moral reasoning, contextual sensitivity, and value-based deliberation. While AI can flag missing ethics statements or detect image manipulation, it cannot meaningfully evaluate whether a study poses ethical concerns beyond predefined rules. Delegating such judgment to AI risks reducing ethics to compliance rather than reflection.

The feasibility of AI peer review also depends on disciplinary differences. In fields with standardized methods and reporting norms, such as certain areas of biomedical research or physics, AI assistance is more readily applicable (Bouter, 2018). In contrast, disciplines that value theoretical interpretation, qualitative analysis, or creative synthesis pose greater challenges. A one-size-fits-all AI reviewer is neither realistic nor desirable. Feasibility increases when AI tools are tailored to specific disciplinary contexts rather than imposed universally.

Another critical consideration is the social function of peer review. Peer review is not only a filter but a dialogue among scholars. Reviewers often act as mentors, shaping manuscripts through constructive critique. This relational aspect fosters community norms, shared standards, and intellectual development. AI, lacking professional stake or scholarly identity, cannot fully participate in this communal function. Even if technically reliable, an AI-only review process may erode the social fabric that sustains scientific communities.

That said, AI can significantly improve feasibility by reducing reviewer burden. One of the most compelling use cases is AI as a pre-review or co-reviewer (Resnik et al., 2008). By handling routine checks and generating preliminary assessments, AI can free human reviewers to focus on conceptual depth and interpretive judgment. Editors could receive structured AI reports highlighting strengths, weaknesses, and potential red flags, improving decision efficiency without replacing human oversight. This hybrid model balances reliability with feasibility.

Concerns about accountability further limit the feasibility of fully autonomous AI reviewers. When a human reviewer makes an error, responsibility is traceable. With AI, accountability becomes diffuse: does it lie with developers, publishers, editors, or institutions? Without clear governance frameworks, AI-driven decisions risk becoming ethically and legally ambiguous (Kovanis et al., 2016). For peer review—a process that can determine careers, funding, and reputations—this ambiguity is unacceptable. Feasibility therefore depends on establishing robust accountability structures.

Data privacy and intellectual property present additional challenges. Manuscripts under review contain unpublished ideas and data (Lipworth et al., 2011). Using them to train or operate AI systems raises concerns about data leakage, unauthorized reuse, and competitive disadvantage. Trust in peer review depends on confidentiality. Any AI system involved must guarantee strict data isolation and transparency about data use. Without this, feasibility collapses regardless of technical capability.

The question of reliability must also be examined longitudinally. Peer review is iterative; standards evolve as fields advance. Human reviewers adapt through ongoing engagement with research and debate. AI systems require retraining and updating to reflect these shifts (Marcus & Davis, 2019). If AI models lag behind current discourse, they may enforce outdated standards. Reliability over time thus requires continuous human supervision and recalibration, reinforcing the argument that AI cannot function independently.

Importantly, AI's role in peer review raises philosophical questions about what peer review is for. Is it primarily a quality control mechanism, or is it a collective epistemic practice? If the former, automation appears attractive. If the latter, replacing human judgment with algorithmic evaluation risks hollowing out the very process it seeks to improve. Feasibility depends not only on technical readiness but on alignment with the values of science.

Current evidence suggests that AI performs best when its role is clearly bounded. Tools that detect plagiarism, statistical anomalies, image manipulation, or reporting gaps are already improving reliability. Automated reviewer suggestions based on expertise matching can enhance efficiency. Language-editing AI can reduce linguistic bias against non-native English speakers. These applications are feasible, valuable, and largely uncontroversial. The controversy arises when AI is positioned as a substitute rather than an assistant.

In reviewing the state of the field, it becomes clear that the question is not whether AI can be a peer reviewer, but whether peer review can remain credible without human judgment. AI can evaluate form, consistency, and probability; humans evaluate meaning, value, and implication. Reliability emerges from combining these strengths, not from replacing one with the other. Feasibility lies in integration, not delegation.

In conclusion, AI as a peer reviewer is partially reliable and conditionally feasible. It excels at technical screening, consistency checks, and efficiency enhancement, but falls short in evaluating novelty, ethical complexity, and conceptual significance. Fully autonomous AI peer review is neither realistic nor desirable at present. The most sustainable future lies in hybrid models where AI augments human reviewers, improves transparency, and reduces workload while preserving human responsibility and epistemic judgment. The success of AI in peer review will ultimately depend not on how much authority we grant it, but on how wisely we constrain it. ■

Received: October 19, 2025 | Revised: December 10, 2025 | Accepted: January 05, 2026

References

Bedeian, A. G. (2004). Peer review and the social construction of knowledge in the management discipline. Academy of Management Learning & Education, 3(2), 198–216. DOI: https://doi.org/10.5465/amle.2004.13500506

Birhane, A., Prabhu, V., & Kahembwe, E. (2021). Multimodal datasets: Misogyny, pornography, and malignant stereotypes. arXiv. https://arxiv.org/abs/2110.01963

Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from The American Economic Review. American Economic Review, 81(5), 1041–1067.

Bouter, L. M. (2018). Fostering responsible research practices is a shared responsibility of multiple stakeholders. Journal of Clinical Epidemiology, 96, 143–146. DOI: https://doi.org/10.1016/j.jclinepi.2017.12.016

COSE (Council of Science Editors). (2018). Ethics in scientific publishing (2nd ed.). University of Chicago Press.

Gerstner, K., & König, J. (2023). Artificial intelligence in peer review: Opportunities, risks, and governance challenges. Research Integrity and Peer Review, 8, 18. DOI: https://doi.org/10.1186/s41073-023-00133-7

Horbach, S. P. J. M., & Halffman, W. (2018). The changing forms and expectations of peer review. Research Integrity and Peer Review, 3, 8. DOI: https://doi.org/10.1186/s41073-018-0051-5

Kovanis, M., Porcher, R., Ravaud, P., & Trinquart, L. (2016). The global burden of journal peer review in the biomedical literature: Strong imbalance in the collective enterprise. PLOS ONE, 11(11), e0166387. DOI: https://doi.org/10.1371/journal.pone.0166387

Lipworth, W., Kerridge, I., & Carter, S. M. (2011). Journal peer review in context: A critical review of the literature. Health Sociology Review, 20(3), 297–314.DOI: https://doi.org/10.5172/hesr.2011.20.3.297

Marcus, G., & Davis, E. (2019). Rebooting AI: Building artificial intelligence we can trust. Pantheon Books.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. DOI: https://doi.org/10.1126/science.aax2342

Resnik, D. B., Gutierrez-Ford, C., & Peddada, S. (2008). Perceptions of ethical problems with scientific journal peer review: An exploratory study. Science and Engineering Ethics, 14(3), 305–310. DOI: https://doi.org/10.1007/s11948-008-9059-4