Trends in Scholarly Publishing

OPEN ACCESS

Governing the Game-Changer: Reimagining Peer Review in the Age of AI

[1]Lucy Threadgold, [2]Hong Zhou, [3]Simone Ragavooloo, [4]Daniel Stuckey and [5]Maryam Sayab
[1]Peer Review Lead, Emerald Publishing, United Kingdom
[2]KnowledgeWorks Global Ltd., United States of America
[3]Frontiers, Research Integrity, London
[4]Research Integrity and Publishing Ethics Centre of Expertise, Elsevier, United Kingdom
[5]Director of Communications, Asian Council of Science Editors, United Arab Emirate

ABSTRACT

The peer review system, long regarded as the cornerstone of scholarly publishing, is facing unprecedented strain. Rising submission volumes, reviewer fatigue, and increasing instances of research misconduct have revealed deep structural vulnerabilities. Artificial Intelligence (AI) presents transformative opportunities to reinforce this essential process, accelerating reviewer selection, detecting manipulation, and advancing equity in participation. Yet, it also introduces new risks of over-reliance, opacity, bias, and the potential erosion of trust in scientific judgment. This Perspective explores the dual role of AI in peer review, analyzing both its capacity to enhance integrity and its potential to destabilize it. Drawing on insights from leading publishers and emerging governance models, we argue that the future of peer review will be defined not by the sophistication of algorithms but by the strength of the ethical frameworks that govern them. With transparency, accountability, and hybrid human machine collaboration at its foundation, AI can help transform peer review into a more inclusive, efficient, and trustworthy system, provided that integrity remains its guiding principle.

KEYWORDS

Peer review, artificial intelligence, research integrity, scholarly publishing, governance, transparency, accountability, reviewer diversity, human-AI collaboration, publication ethics

INTRODUCTION

Peer review, long regarded as the gold standard of scientific quality assurance, is showing visible signs of strain. Mounting submission volumes, reviewer fatigue, and uneven participation have exposed systemic inefficiencies, challenging the stability of a system once considered resilient[1,2]. Compounding this stress, the growing complexity and interdisciplinarity of research outputs have pushed editorial workflows to their operational limits[3].

Against this backdrop, a new force is emerging, not to replace human expertise but to reinforce it. The AI has surfaced as both a solution and a disruptor, prompting the scholarly community to reconsider how the very foundations of peer review might evolve[4].

ISSN: 2789-9446 (Online)
ISSN: 2789-9438 (Print)
https://doi.org/10.21124/tsp.2025.109.116

Received: 05 Nov. 2025
Accepted: 10 Dec. 2025
Published: 31 Dec. 2025

The AI already performs functions once dependent on human labor. It can screen manuscripts for plagiarism, image manipulation, and statistical anomalies, detect papermill patterns invisible to editors, triage out-of-scope submissions, and suggest qualified reviewers across global networks[5,6]. In a system constrained by reviewer scarcity yet abundant in data, such automation brings welcome relief, offering speed and scalability without necessarily compromising rigor.

Yet the potential of AI is not limited to efficiency but also in its potential to promote diversity, equity, and inclusion. Properly counter linguistic, geographic, and disciplinary biases in reviewer selection, thereby broadening participation. However, this promise depends on intentional design. Adopting automation without first defining ethical goals risks entrenching existing inequities. Publishers must articulate what "good" looks like before layering technology onto flawed human systems. Without such clarity, well-intentioned innovation may reproduce or amplify the very problems it aims to solve.

Successful Integration of AI requires capacity building and critical literacy. Editorial teams must be trained to interpret algorithmic outputs, while all stakeholders, authors, reviewers, and editors alike, need to understand where, why, and how AI tools are deployed. Only through such transparency can AI evolve from a black box into a trusted, collaborative partner in the scholarly process[7]. This article aims to critically examine how Artificial Intelligence (AI) is transforming the peer review process in scholarly publishing. It explores both the opportunities and challenges AI introduces, ranging from improving efficiency, accuracy, and bias detection to raising concerns about transparency, accountability, and ethical oversight. The objective is to propose a reimagined framework for integrating AI responsibly into peer review, ensuring that technological innovation strengthens, rather than compromises, research integrity and editorial standards.

WHEN AUTOMATION THREATENS TRUST

Yet, as with every technological revolution, innovation brings uncertainty. The same tools that promise speed and precision can also introduce risks that strike at the heart of scholarly communication, trust. For peer review, the greatest danger is the quiet erosion of that trust, the foundation upon which the credibility of science rests.

Over-reliance on AI, opaque decision-making, or biased algorithms can quietly undermine confidence in the system. Questions such as "Why was this reviewer suggested?" or "Why was this paper flagged?" must have transparent answers; otherwise, confidence collapses. Emerging threats such as prompt injection and AI-enabled manipulation further underscore the need for governance frameworks that prioritize security, disclosure, and traceability[8,9].

A subtler danger lies in what scholars have termed the "creep effect", the gradual automation of intellectual judgment. If AI begins to draft, review, and validate research without meaningful human oversight, the system risks devolving into an echo chamber of "AI-written, AI-reviewed, AI-approved" science. Such a cycle would dilute human rigor and potentially distort the body of evidence upon which knowledge advances.

The consequences reach beyond academia. When flawed or biased research informs clinical guidelines, public policy, or technological innovation, the harms are tangible, from patient risk to societal misinformation. The papermill crisis illustrates how quickly trust can collapse and how painstakingly it must be rebuilt[6]. With AI, the timeline shortens, and the scale magnifies, making pre-emptive governance not optional but essential.

Rejecting AI is neither feasible nor desirable. The challenge, therefore, is not whether AI belongs in peer review but how it is governed. Accountability, explainability, and transparency must be embedded from the outset. Human judgment must remain central, with clear visibility into data provenance, training processes, and decision logic. Only when responsibility is explicit can trust be preserved.

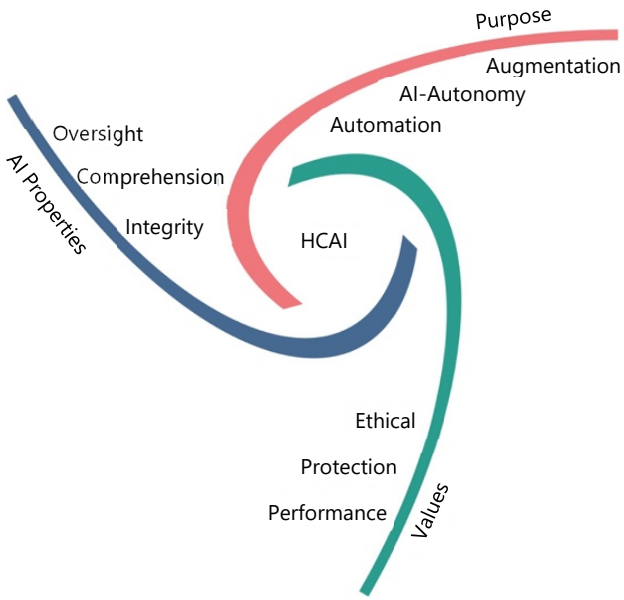

Fig. 1: Human-centered AI framework

To conceptualize this relationship between human oversight, ethical principles, and AI functionalities, we draw on the Human-Centered AI (HCAI) framework proposed by Shneiderma[10], which emphasizes integrity, oversight, and ethical values as the core of AI deployment in critical systems (Fig. 1).

BUILDING GUARDRAILS FOR RESPONSIBLE AI

Addressing these risks requires shifting from anxiety to action. If the risks of AI in peer review are significant, they are still manageable through thoughtful governance and shared accountability. The challenge lies in building robust guardrails, policies, standards, and oversight mechanisms that keep technology aligned with scientific goals. Governance must begin with transparency: Authors, reviewers, and editors deserve clarity on where and how AI tools are used. Journals should disclose AI use and require reviewers to declare any assistance.

As shown in Fig. 1, responsible AI in peer review is not simply about automation but about balancing technological capabilities with human oversight and ethical values. This alignment of purpose, properties, and values forms the foundation for trustworthy integration.

Policy development must be collaborative, not isolated. Effective frameworks depend on cooperation among publishers, institutions, funders, and researchers. Shared global standards for AI use can ensure coherence while allowing disciplinary flexibility, defining expectations for privacy, accountability, and disclosure, a stable base for responsible innovation[11].

Governance extends beyond policy. It requires human expertise to interpret and question AI outputs. Editorial teams must understand both how tools work and why they make certain recommendations. Explainability, the ability to trace an algorithm's reasoning, is essential; without it, oversight weakens and accountability erodes[7].

Finally, governance must evolve with technology. Continuous auditing, evaluation, and recalibration are needed to detect bias or drift, while monitoring human engagement helps prevent over-reliance or misuse. In this sense, governance is not a static checklist but a sustained commitment to ethical stewardship an ongoing dialogue between innovation and integrity.

HUMAN-AI COLLABORATION: THE HYBRID FUTURE

Once strong guardrails are in place, the question shifts from control to collaboration. The future of peer review will not be defined by choosing between humans and machines but by how well they collaborate. The AI is most effective when it manages repetitive and data-intensive tasks, freeing human experts for critical judgment, ethics, and contextual reasoning. This hybrid model, AI as an assistant, not a replacement, offers the most sustainable path forward. Studies suggest hybrid systems can enhance accuracy and inclusivity while maintaining human accountability[5-12].

Hybrid peer review combines human and machine strengths. Algorithms flag integrity issues, suggest reviewers, and surface relevant literature, while editors and reviewers assess novelty, rigor, and ethical soundness. This division of labor enhances efficiency and preserves scholarly quality, ensuring final decisions rest on human judgment rather than algorithmic output.

Keeping humans in the loop safeguards accountability. The AI may recommend reviewers or identify data concerns, but editors retain responsibility for outcomes. Such shared authorship preserves credibility and trust[13,14].

Hybrid models can also advance inclusion. By automating administrative work, AI allows editors more time for mentoring and thoughtful evaluation. Smarter reviewer matching expands participation and diversity, making peer review not only faster but fairer when guided by ethical oversight and intentional design. Thus, collaboration is not simply about efficiency; it is about restoring balance between technological capability and human judgment.

REFORMING PEER REVIEW BEYOND TECHNOLOGY

Still, even collaboration cannot mask the deeper reality: Technology alone cannot heal a strained system. While AI holds transformative potential, technology alone cannot fix the systemic weaknesses of peer review. Structural and cultural reforms must evolve alongside automation[2].

Standardized review templates and structured feedback forms can enhance clarity, comparability, and accountability. Collaborative review models, where authors and reviewers interact transparently, can improve constructive dialogue. Introducing incentives for high-quality reviews and expanding reviewer pools based on expertise can mitigate fatigue and accelerate turnaround times.

Equally essential is education. The AI literacy, understanding both the capabilities and limitations of AI, must become a core component of reviewer and editor training. Yet technical familiarity must be matched by ethical and critical competence. Training in research integrity, reporting standards, and bias recognition ensures that technology complements, rather than compromises, scholarly rigor[15,16].

Ultimately, true reform depends on cultural change. The AI can accelerate the process, but only human values can sustain it. Peer review will thrive not through automation alone but through renewed commitment to transparency, fairness, and accountability, principles that no algorithm can replicate[17].

INTEGRITY AT THE CORE

Each of these reflections leads to a central truth: Integrity remains the bedrock of credible science. The integration of AI into peer review is not a passing trend but a fundamental shift in how science is accessed and communicated. Its impact on strengthening or weakening the system depends on choices made today. As Tennant and Ross-Hellauer note, sustaining trust in peer review requires not just new tools but renewed ethical commitment. The opportunities are clear: speed, quality, broader participation, and better detection of misconduct. Yet so are the risks of opacity, bias, and loss of trust[1].

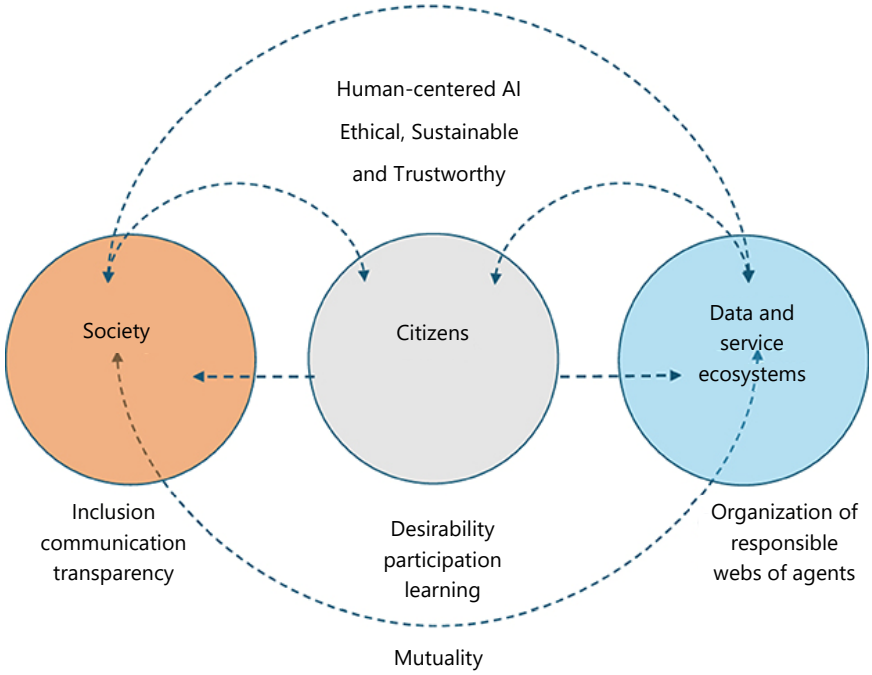

Fig. 2: Human-centered AI governance model
Reproduced from Licensed under CC BY 4.0[19]

The goal is not to embrace AI blindly or reject it outright, but to govern it wisely. This requires clear objectives, training, transparency, and shared standards. Automation must be paired with human judgment, and innovation must be accompanied by ethical reflection. Integrity, not efficiency, must remain the guiding principle that anchors all technological progress in scholarly publishing[18].

TOWARD A RESPONSIBLE FUTURE

If integrity is the compass, governance is the path forward. For AI to become a force for strengthening, not destabilizing, peer review, the scholarly community must act with shared purpose and foresight. The way forward begins with collaboration. Publishers, funders, institutions, and researchers must work together to establish global standards that articulate clear principles of transparency, accountability, privacy, bias mitigation, and explainability. Without such coordination, governance will remain fragmented and reactive[7-10].

As illustrated by Sigfrids *et al.*[19] an effective governance framework involves multiple interconnected layers society, citizens, and data ecosystems all operating under principles of mutuality, transparency, and inclusion (Fig. 2). Applying such a model to peer review underscores that technology alone is insufficient; coordinated stewardship across the ecosystem is essential.

Transparency and disclosure must also become standard practice across all stages of the review process. Authors, editors, and reviewers deserve clarity about when and how AI tools are applied. Publicly disclosing tool usage and requiring reviewers to declare AI assistance can normalize responsible practice and sustain trust.

At the same time, human capacity must grow alongside technological capability. Editorial teams, reviewers, and authors need more than access to tools; they need the competence to interpret them critically. Building AI literacy across the scholarly ecosystem is as important as developing the technology itself.

Yet no matter how sophisticated AI becomes, human oversight must remain the moral and intellectual center of peer review. Algorithms can support judgment but cannot replace it. The AI should augment human reasoning, never substitute it, ensuring that decisions remain contextually and ethically grounded.

Governance, moreover, should not be static. Responsible innovation requires continuous evaluation, auditing, and recalibration to detect bias or drift and to ensure tools perform as intended. The system must be flexible enough to adapt as technology evolves.

Building on this ecosystem perspective (Fig. 2), governance must extend beyond technical standards. It should embed shared values into every layer of peer review, from reviewer selection algorithms to disclosure requirements and auditing. By anchoring peer review in a human-centered governance model, we can enable responsible adoption of AI without eroding trust in scientific communication.

Finally, true reform extends beyond technology. Strengthening incentives for high-quality reviewing, diversifying reviewer pools, improving feedback structures, and embedding ethical education into reviewer training will help ensure that peer review evolves in both capability and integrity.

Together, these actions define not merely a policy agenda but a cultural commitment-one that balances innovation with accountability, and technological progress with enduring human values.

Priority actions and intended outcomes:

- **Global standards:** Establish shared principles for AI governance across publishers and funders
  Intended outcome: Consistent expectations and reduced ambiguity
- **Transparency and disclosure:** Require open reporting of AI tool use and reviewer declarations
  Intended outcome: Public trust and accountability
- **AI literacy:** Train editors and reviewers in interpreting AI outputs
  Intended outcome: Informed oversight and ethical use
- **Human oversight:** Keep humans central in all decision-making loops
  Intended outcome: Preserved accountability and judgment
- **Continuous auditing:** Evaluate and update AI systems regularly
  Intended outcome: Mitigation of bias and model drift
- **Structural reform:** Strengthen incentives and standardize review frameworks
  Intended outcome: Sustainable improvement beyond technology

CONCLUSION

The rise of AI in peer review represents both an opportunity and a test of collective responsibility. Its arrival is no longer a distant prospect; it is already reshaping how scholarly credibility is built and maintained. Whether it becomes a game-changer or a governance challenge will depend on how intentionally and collaboratively the publishing community responds. If automation proceeds unchecked, bias and opacity may deepen. Still, if guided by transparency, accountability, inclusivity, and sustained human oversight, AI can transform peer review into a process that is faster, fairer, and more trusted.

The path ahead, therefore, is one of balance. The future of peer review will not be authored by machines alone, nor preserved by humans alone. It will be co-created through collaboration, humans and algorithms working in partnership under shared ethical principles and anchored in integrity.

As AI continues to redefine how knowledge is evaluated and disseminated, the scholarly community faces a defining choice: To let technology dictate the evolution of peer review, or to govern it wisely so that innovation strengthens rather than supplants human judgment. The time to act is now, and the legacy of this transformation will be measured not by efficiency, but by how steadfastly integrity remains at the center of science.

SIGNIFICANCE STATEMENT

This study discovered the evolving role of artificial intelligence in peer review that can be beneficial for strengthening transparency, efficiency, and integrity across scholarly publishing systems. By examining how AI can support reviewer selection, detect manipulation, and enhance equitable participation, the study highlights pathways for responsible innovation. This study will help researchers uncover the critical areas of AI-driven peer-review governance that many were not able to explore. Thus, a new theory on balanced human AI collaboration in editorial decision-making may be arrived at.

REFERENCES

1. Tennant, J.P. and T. Ross-Hellauer, 2020. The limitations to our understanding of peer review. Res. Integrity Peer Rev., Vol. 5. 10.1186/s41073-020-00092-1.
2. Horbach, S.P.J.M. (Serge) and W. (Willem) Halffman, 2018. The changing forms and expectations of peer review. Res. Integrity Peer Rev., Vol. 3. 10.1186/s41073-018-0051-5.
3. Huisman, J. and J. Smits, 2017. Duration and quality of the peer review process: The author's perspective. Scientometrics, 113: 633-650.
4. Zhang, G. and F. Wei, 2020. Analysing the research performance of province level administrative regions in China. Learned Publ., 33: 395-409.
5. Schulz, R., A. Barnett, R. Bernard, N.J.L. Brown, J.A. Byrne *et al*., 2022. Is the future of peer review automated? BMC Res. Notes, Vol. 15. 10.1186/s13104-022-06080-6.
6. Abalkina, A., 2023. Publication and collaboration anomalies in academic papers originating from a paper mill: Evidence from a Russia based paper mill. Learned Publ., 36: 689-702.
7. Haibe-Kains, B., G.A. Adam, A. Hosny, F. Khodakarami and T. Shraddha *et al*., 2020. Transparency and reproducibility in artificial intelligence. Nature, 586: E14-E16.
8. Bucci, E.M., 2018. Automatic detection of image manipulations in the biomedical literature. Cell Death Dis., Vol. 9. 10.1038/s41419-018-0430-3.
9. Zhuang, H., T.Y. Huang and D.E. Acuna, 2021. Graphical integrity issues in open access publications: Detection and patterns of proportional ink violations. PLoS Comput. Biol., Vol. 17. 10.1371/journal.pcbi.1009650.
10. Shneiderman, B., 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. Int. J. Hum.-Comput. Interact., 36: 495-504.
11. Peroni, S. and D. Shotton, 2020. OpenCitations, an infrastructure organization for open scholarship. Quant. Sci. Stud., 1: 428-444.
12. Bravo, G., F. Grimaldo, E. López-Iñesta, B. Mehmani and F. Squazzoni, 2019. The effect of publishing peer review reports on referee behavior in five scholarly journals. Nat. Commun., Vol. 10. 10.1038/s41467-018-08250-2.
13. Ross-Hellauer, T. and E. Görögh, 2019. Guidelines for open peer review implementation. Res. Integrity Peer Rev., Vol. 4. 10.1186/s41073-019-0063-9.
14. Ross-Hellauer, T., 2017. What is open peer review? A systematic review. F1000Research, Vol. 6. 10.12688/f1000research.11369.2.
15. Willis, J.V., J. Ramos, K.D. Cobey, J.Y. Ng and H. Khan *et al*., 2023. Knowledge and motivations of training in peer review: An international cross-sectional survey. PLoS ONE, Vol. 18. 10.1371/journal.pone.0287660.
16. Buser, J.M., K.L. Morris, V.M. Dzomeku, T. Endale, Y.R. Smith and E. August, 2023. Lessons learnt from a scientific peer-review training programme designed to support research capacity and professional development in a global community. BMJ Global Health, Vol. 8. 10.1136/bmjgh-2023-012224.

17. Singh, S., A.C. Sharma, P.K. Chaurasia, V. Kumar, S.L. Bharati and A.Y.F. Allam, 2024. Prospects and Importance of Training Needs in Peer Review Models. In: Scientific Publishing Ecosystem, Joshi, P.B., P.P. Churi and M. Pandey (Eds.), Springer Nature, Singapore, ISBN: 978-981-97-4060-4, pp: 347-365.
18. Lund, B., Z. Orhan, N.R. Mannuru, R.V.K. Bevara, B. Porter, M.K. Vinaih and P. Bhaskara, 2025. Standards, frameworks, and legislation for artificial intelligence (AI) transparency. AI Ethics, 5: 3639-3655.
19. Sigfrids, A., J. Leikas, H. Salo-Pöntinen and E. Koskimies, 2023. Human-centricity in AI governance: A systemic approach. Front. Artif. Intell., Vol. 6. 10.3389/frai.2023.976887.