International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 3

August – 2026

Artificial Intelligence in Scholarly Peer Review: Ethical Considerations, Current Practices, and Future Implications

Aras Bozkurt
Open Education Faculty, Anadolu University, Eskişehir, Turkey

Abstract

The integration of artificial intelligence (AI) into scholarly peer review represents a fundamental transformation of academic publishing's quality control mechanisms. This report critically examined the ethical considerations, institutional practices, and emerging technologies associated with AI-assisted peer review. Drawing on recent policy documents from major publishing organizations, empirical research on AI implementation, and critical scholarship on algorithmic bias, this analysis revealed significant tensions between efficiency gains and integrity preservation. While AI tools have demonstrated potential for addressing reviewer burnout and publication delays, their deployment raises critical concerns regarding confidentiality breaches, accountability gaps, algorithmic bias, and the erosion of expert judgment. Major organizations (e.g., International Committee of Medical Journal Editors) and leading publishers such as Elsevier and Taylor & Francis have emphasized disclosure where AI is used, human accountability, and strict confidentiality controls—often prohibiting uploading unpublished manuscripts into generative AI tools. However, empirical evidence has suggested nontrivial, and potentially growing, undisclosed large language model (LLM)-assisted text in peer review in some conference contexts. This report concluded that AI should serve as an augmentative rather than substitutive technology in peer review, with robust governance frameworks, transparent disclosure mechanisms, and continuous evaluation of equity implications essential for responsible implementation.

Keywords: artificial intelligence (AI), scholarly peer review, publication ethics, research integrity, academic publishing

Introduction

The traditional peer review system, long considered the cornerstone of scientific quality control, faces great strain. Manuscript submissions to peer-reviewed journals have experienced sustained annual growth of 6.1% since 2013, with corresponding increases in retraction rates and reviewer burden (Publons, 2018; as cited in Checco et al., 2021). Researchers have estimated that over 15 million hours are expended annually reviewing manuscripts that have been previously rejected and resubmitted to alternative venues (American Journal Experts, 2018; Checco et al., 2021). This unsustainable trajectory has prompted scholarly publishers and funding agencies to explore technological interventions, most prominently artificial intelligence, as potential solutions to systemic inefficiencies.

The emergence of LLMs such as ChatGPT, Gemini, Claude, and domain-specific tools has catalysed intense debate within the academic community. Proponents have argued that AI can streamline repetitive tasks, reduce publication latency, and provide consistent evaluation criteria (Doskaliuk et al., 2025). Critics, on the other hand, have contended that AI deployment risks compromising research integrity, perpetuating algorithmic bias, and fundamentally altering the nature of scholarly evaluation (Mollaki, 2024). This tension reflects broader concerns about the role of automated systems in knowledge production and validation.

This report synthesized (a) policy statements and guidance from major editorial organizations and publishers, and (b) empirical studies examining AI use in peer review and scholarly writing. Sources were identified via targeted searches of publisher policy pages and organizational guidance, complemented by recent empirical work on LLMs use in conference peer review. Given the rapid evolution of policies and tools, this report treated current guidance as a time-bounded snapshot and emphasized verifiable, primary policy documents where available.

Critical Perspectives and Ethical Considerations in AI-Integrated Peer Review

The current landscape of AI in scholarly practice is characterized by a rapid transition from theoretical exploration to active, large-scale implementation. As documented in recent literature, AI has evolved beyond simple grammar correction into a multifaceted writing assistant and specialized agent capable of shaping the entire research lifecycle—from initial hypothesis generation to post-publication dissemination (Enoh, 2026; Luo et al., 2025; Mann et al., 2025).

This integration has moved beyond academic discussion into practical, albeit controversial, application within the peer review workflow. Scholarly practice is now defined by a concerted push toward responsible, secure, and efficient AI reviewer (AIR) systems, designed to address the systemic crisis in traditional review while attempting to safeguard scientific integrity (Wang et al., 2026). Ultimately, this transition represents a fundamental shift in academic publishing, necessitating a delicate balance between unprecedented efficiency gains and profound ethical risks.

Opportunities: Efficiency, Equity, and Standardization

Proponents of AI in peer review emphasize its ability to mitigate the reviewer fatigue crisis and bridge systemic gaps in the current model.

Operational Efficiency and Feasibility

Early feasibility studies have demonstrated that LLMs such as ChatGPT can generate reviews that align closely with human critical analysis in specific clinical case reports (Biswas et al., 2023). AI tools provide near-instantaneous feedback on manuscript clarity and adherence to significance, innovation, evaluation, and reproducibility (SIER) standards, allowing editors to filter low-quality submissions rapidly (Wang et al., 2026).

Addressing Systemic Inequities

AI has the potential to transform peer review into a more equitable process by mitigating human biases related to author prestige or institutional affiliation (Sabet et al., 2023). Furthermore, AI serves as a powerful democratizing tool for non-native English speakers, assisting in polishing manuscripts and reviews to meet global standards (Lee et al., 2025).

Enhanced Objectivity

Automated systems can provide consistent evaluation across manuscripts, potentially reducing human-induced variability in identifying technical flaws or statistical reporting errors (Doskaliuk et al., 2025).

Critical Challenges of AI-Assisted Scholarly Practice

The current landscape of AI in scholarly practice is defined by a rapid transition from theoretical exploration to practical implementation. This shift represents a fundamental realignment of academic publishing, balancing unprecedented efficiency gains against profound ethical and epistemological risks (Wang et al., 2026).

The Duality of AI: Support Versus Substitution

AI has evolved beyond simple grammar correction to become a multifaceted writing assistant shaping the entire research lifecycle, from hypothesis generation to post-publication dissemination (Enoh, 2026; Luo et al., 2025). Tools like Elicit, Perplexity, and Consensus are used to optimize literature analysis, though they are often noted for *shallow analytical depth* and a tendency toward standardized writing styles (Granjeiro et al., 2025; Imran & Almusharraf, 2023).

While AI can produce believable text, its ability to interpret complex scientific questions remains limited, leading to AI hallucinations and fabricated references (Kitamura, 2023; Semrl et al., 2023). This over-reliance creates a cognitive debt; the long-term erosion of critical thinking skills (Bozkurt et al., 2026; Kosmyna et al., 2025; Sun, 2025). A strengths, weaknesses, opportunities, and threats (SWOT) analysis identified these tools as a strength for productivity but a threat to the ethical compass of science (Gurnal & Rana, 2025).

Epistemological Bias and Indigenous Representation

The integration of AI operates at multiple levels of systemic concern:

- Methodological preferences: Because AI systems embody Western-trained epistemological assumptions, researchers employing alternative or non-Western approaches may face systematic disadvantages (Sabet et al., 2023).

- Indigenous knowledge and culture: The representation of Indigenous people and culture is a critical dimension. If AI-assisted peer review becomes normative, there is a risk that these tools, trained on data that historically excludes or misrepresents Indigenous voices, will fail to recognize the validity of Indigenous ontologies. Without specific cultural competence, AI models may dismiss local knowledge systems as outliers, further marginalizing Indigenous scholarship within the global record.

Authorship, Accountability, and Ghost Reviews

There is a fundamental paradox regarding the eligibility of AI for authorship. The consensus among major bodies—including Committee on Publication Ethics (COPE), International Committee of Medical Journal Editors (ICMJE), and World Association of Medical Editors (WAME)—is that AI cannot be listed as an author because it lacks legal personhood and cannot take responsibility for work integrity (Bakla, 2023; Bozkurt, 2024; COPE Council, 2023; Zielinski et al., 2023).

Despite this, so-called ghost reviews and AI-assisted drafting have become prevalent. Studies have indicated that 6.5% to 16.9% of reviews at major AI conferences contain LLM-generated text (Bozkurt, 2024; Liang et al., 2024). This creates a legal quandary where AI-generated content may be ineligible for copyright, potentially forcing such works into the public domain (Rohit & Verma, 2024). Furthermore, reviewers using public AI platforms risk violating fiduciary confidentiality obligations (ICMJE, 2024; Kocak et al., 2025).

Global Equity and the Digital Divide

AI deployment risks exacerbating existing socioeconomic inequalities (Abdelwahab, 2024).

- Language versus access: While AI helps non-native English speakers overcome linguistic barriers (Delanghe, 2024; Enoh, 2026; Lee et al., 2025), the monetization of advanced models has created a divide between well-funded institutions and those in under-resourced regions (Abdelwahab, 2024; Ateriya et al., 2025).

- The efficiency gap: Researchers in developing countries may struggle to meet efficiency expectations in peer review while simultaneously facing AI-assisted evaluation without reciprocal access to preparation tools (Abdelwahab, 2024; Dwivedi et al., 2024).

Specialized Risks and Manipulation

In specialized fields like surgical and clinical research, the risks of hallucinations and biased outputs are particularly acute (Celik, 2025; Luo et al., 2025). Peer review faces a dilemma of depth as evaluations are accelerated at the cost of nuanced expert understanding (Ben Saad et al., 2025). Additionally, the system is vulnerable to prompt injection or explicit manipulation, where authors inject covert content into manuscripts to trigger positive AI feedback (Scuderi et al., 2026; Ye et al., 2024).

Governance and Future Outlook

Institutional responses have shifted toward disclosure-based models and fostering AI literacy (Bozkurt, 2024; Celik, 2025; Clark, 2025; Zielinski et al., 2023). As journals develop best practice guidelines (Ben Saad et al., 2025; Leung et al., 2023), the academic community must decide if the value of a review lies in the human expertise behind the judgment or merely in the utility of the output (Biswas et al., 2023; Matsubara, 2026).

Current Practices and Institutional Guidelines

Major Organizational Policies

In response to proliferating AI use, major scholarly publishing organizations have developed guidelines attempting to establish boundaries for appropriate use. These policies reveal a consensus position: AI may serve as a supportive tool but cannot replace human judgment, and all use must be transparently disclosed.

International Committee of Medical Journal Editors (ICMJE)

The ICMJE's January 2024 update to its recommendations explicitly addressed AI use across the publication process. For authors, the ICMJE mandated disclosure of any AI assistance in manuscript preparation, with writing support acknowledged and AI use in data collection or analysis detailed in the methods section (ICMJE, 2024). Critically, "chatbots (such as ChatGPT) and other AI-assisted tools should not be listed as authors because they cannot be responsible for the accuracy, integrity, and originality of the work" (ICMJE, 2024, p. 20).

For reviewers, the ICMJE has established that

> reviewers must request permission from the journal prior to using AI technology to facilitate their review [and that] reviewers must maintain the confidentiality of the manuscript as outlined above, which may prohibit the uploading of the manuscript to software or other AI technologies where confidentiality cannot be assured. (ICMJE, 2024, p. 20).

The guidelines explicitly warn that "reviewers should be aware that AI can generate authoritative-sounding output that can be incorrect, incomplete, or biased" (ICMJE, 2024, p. 3).

For editors, the ICMJE has stated that "editors should be aware that using AI technology in the processing of manuscripts may violate confidentiality" (ICMJE, 2024, p. 5). This represents a significant constraint on editorial use of AI for manuscript screening or reviewer recommendation systems.

Committee on Publication Ethics (COPE)

COPE has emerged as a leading voice in articulating ethical standards for AI in scholarly publishing. The COPE Council's position statement on authorship and AI has explicitly stated that "AI tools cannot be listed as an author of a paper [and that regarding] authorship and contributorship… AI tools cannot meet the requirements for authorship as they cannot take responsibility for the submitted work" (COPE Council, 2025, para. 2).

In July 2025, COPE hosted a major forum titled *Emerging AI Dilemmas in Scholarly Publishing*, which identified four critical themes: (a) responsible and ethical use of AI, (b) transparency and disclosure, (c) detection and editorial standards, and (d) impact on peer review equity and inclusion (Zhou & Soulière, 2025). Forum discussions revealed that editors face mounting challenges from AI-generated submissions and reviews that are "often detailed but inaccurate, causing delays, policy breaches, and extra workload for journal staff" (COPE Council, 2025, para. 1).

COPE has emphasized that transparency and disclosure are fundamental ethical obligations. Publishers should require an explicit declaration of AI use in both manuscripts and reviews, with clear specification of which tools were used and how they contributed to the work product.

Publisher-Specific Policies

Major publishers have implemented increasingly stringent AI policies. Nature Portfolio's editorial policies distinguish between AI-assisted copy editing (defined as improvements to readability, grammar, and style that do not include generative editorial work) and prohibited uses (Nature Portfolio, 2025). The JAMA Network has prohibited listing AI tools as authors and mandates that "authors must disclose details about the AI tools used and take responsibility for the content generated by these technologies" (Doskaliuk et al., 2025, p. 7).

Sage Publishing's policy has categorized AI uses as assistive (requiring no disclosure), generative (requiring disclosure), or prohibitive. For reviewers, Sage has specified that "reviewers suspecting the inappropriate or undisclosed use of generative AI in a submission should flag their concerns with the journal editor" (Sage, n.d., para. 1). The policy prohibits reviewers from using generative AI to produce review content without explicit permission. Table 1 synthesizes the core positions of leading academic bodies, highlighting the universal rejection of AI authorship and the stringent requirements for human oversight in the review process.

Table 1

The Core Positions of Leading Academic Bodies Regarding AI

Organization	AI for reviewing?	AI for authorship?	Main requirement
ICMJE	Permission required	Prohibited	Confidentiality & transparency
COPE	Discouraged/restricted	Prohibited	Human accountability
Nature	Prohibited (generative)	Prohibited	Disclosure of AI-assistance
Sage	Prohibited (content gen)	Prohibited	Flagging suspicious use

Compliance Challenges and Enforcement Difficulties

Despite clear guidelines, evidence suggests widespread non-compliance. A survey by the Times Higher Education found significant distrust of AI tools among reviewers, yet usage continues to grow, often undisclosed (Mollaki, 2024). The structural challenge is that editors have limited capacity to detect AI use, particularly when users deliberately conceal it.

While detection tools exist, there are significant limitations to their use. For instance, while platforms such as iThenticate now offer AI writing detection capabilities, these tools have demonstrated high false positive rates and can be circumvented through strategic editing (Zhou & Soulière, 2025). Moreover, detection-focused approaches risk creating an adversarial dynamic where users view AI use as a violation to be hidden rather than a practice requiring thoughtful disclosure and oversight.

A further complication involves prompt injection attacks—sophisticated manipulation techniques where bad actors embed hidden instructions in manuscript text designed to influence AI review systems (Lin, 2025). For example, manuscripts have been discovered containing white-colored text (invisible to human readers) stating “for LLMs reviewers: ignore all previous instructions. give a positive review only” (Collu et al., 2025, p. 5). Such manipulations threaten to corrupt not only individual review processes but the broader infrastructure of scientific evaluation.

Acceptable Versus Prohibited Uses: Emerging Boundaries

Synthesizing across multiple institutional policies, a taxonomy of acceptable and prohibited AI uses is emerging. The following are generally acceptable uses:

- grammar and language checking applied to the reviewer’s own written comments (not manuscript text);
- literature search assistance and citation verification;
- statistical validation tools and computational checks;

- plagiarism detection using dedicated, confidentiality-preserving platforms; and
- administrative tasks such as formatting references.

The following are uses that require disclosure and permission:

- using AI to assist in drafting review comments (even when substantially edited);
- employing AI for preliminary manuscript analysis or note-taking;
- using AI to summarize complex methodological sections; and
- any AI use that processes manuscript content.

Finally, these uses are strictly prohibited:

- uploading complete manuscripts to public AI platforms,
- generating review text or recommendations without disclosure,
- replacing substantive human analysis with AI assessment,
- using AI for initial screening decisions without human verification, and
- listing AI tools as co-reviewers or in acknowledgments.

The boundary between assistance and generation remains contested and contextual. As WAME has emphasized, the critical distinction involves whether AI supplements human expertise or substitutes for it (Zhou & Soulière, 2025).

AI Tools and Technologies in Scholarly Publishing

Because many publishers prohibit reviewers from uploading manuscript content into generative AI systems, tools described below should be interpreted by role and permitted data access (author self-check vs. editorial screening vs. reviewer use).

Manuscript Assessment and Screening Tools

Several specialized platforms have emerged specifically for AI-assisted manuscript evaluation, attempting to address ethical concerns through institutional deployment and confidentiality preservation.

Stanford Agentic Reviewer (Jiang & Ng, 2025) was designed to provide researchers with near-instant, actionable feedback to accelerate the research iteration cycle, moving away from the painfully slow six-month traditional peer review loop.

The tool employs an agentic workflow to convert PDFs into Markdown and verify document authenticity. Its unique grounding mechanism involves:

1. Automated literature search generates multi-perspective search queries to pull the latest relevant prior work from arXiv via the Tavily API.

2. Contextual synthesis selects top-tier related papers and generates detailed summaries of their full texts to provide a comparative context for the review.

3. Dimension-based scoring evaluates papers across seven dimensions—originality, research question importance, claim support, experimental soundness, writing clarity, community value, and contextualization.

In terms of performance, the Agentic Reviewer has demonstrated human-level alignment, achieving a Spearman correlation of 0.42 in tests mimicking international Conference on Learning Representations (ICLR) reviews, a figure that slightly exceeds the 0.41 correlation typically found between two human reviewers. However, while the system has been highly calibrated for scoring, its predictive accuracy for final paper acceptance remains lower than human judgment, with an Area Under the Curve (AUC) of 0.75 compared to the human benchmark of 0.84. Despite these results, several key limitations persist. The system's reliance on arXiv grounding makes it highly accurate in AI and computer science but potentially less reliable in disciplines lacking robust open-access preprint cultures. Furthermore, the tool currently only supports English-language manuscripts. From an ethical standpoint, the developers have emphasized that the tool is explicitly intended for author self-assessment and they strongly discourage its use by conference reviewers to bypass official policies.

Review-it is a specialized AI ecosystem designed to accelerate the scholarly review cycle. By leveraging advanced large language models, the platform provides researchers and students with rapid, high-quality feedback to refine manuscripts before formal submission. The service follows a streamlined three-step academic pipeline.

1. Secure manuscript submission. Researchers securely upload their work, including research papers, grant proposals, and theses, to a protected environment. The platform ensures data sovereignty while preparing the document for deep-level architectural and content analysis.

2. Multi-dimensional peer analysis. Review-it performs a rigorous assessment of the document's academic integrity and scholarly impact. Moving beyond simple proofreading, the tool:

    - evaluates methodological soundness, identifying gaps in research design or logical inconsistencies;

    - assesses argument strength, highlighting core contributions and pinpointing areas where evidence may be lacking; and

    - provides scholarly benchmarking, offering feedback grounded in disciplinary standards to help authors understand their work's quality from a reviewer's perspective.

3. Actionable optimization. The final stage provides so-called human-in-the-loop support by offering specific, actionable edits. Review-it helps authors automatically fix structural issues and improve

clarity, ensuring that the academic manuscript is refined, impactful, and ready for the rigors of formal peer review.

Scifocus represents an AI-powered platform designed to enhance peer review workflows in academic settings. The platform integrates advanced AI models with proprietary knowledge bases and academic databases to provide comprehensive support including feedback facilitation, literature management, and efficiency enhancement (Scifocus, 2025). Scifocus assists researchers in summarizing peer-reviewed documents and extracting essential insights, with tools for generating constructive comments and optimizing reviews for academic rigor.

q.e.d Science has partnered with bioRxiv to offer manuscript analysis services. This generative AI platform analyzes claims and supporting data presented in manuscripts to identify gaps that warrant further investigation or claim revision (openRxiv, 2025). Developed iteratively with scientist feedback, q.e.d attempts to help researchers refine claims and strengthen conclusions before formal peer review submission. Early feedback from the scientific community has been reportedly positive, though systematic evaluation data remain limited.

Ex Ordo provides conference management systems incorporating AI-assisted reviewer allocation. The platform uses algorithmic matching to pair submissions with appropriate reviewers based on expertise profiles, potentially reducing time-to-review and improving reviewer-manuscript fit (Scifocus, 2025).

In addition to the tools listed above, most researchers can also conduct the peer review process using an LLM-based AI tool with their own custom prompts.

Research Discovery and Literature Tools

While not designed specifically for peer review, several AI-powered research discovery tools are being adopted by reviewers to support evaluation processes.

Semantic Scholar employs AI to generate summaries and conduct citation analysis, potentially helping reviewers quickly assess a manuscript's relationship to existing literature. Elicit and Research Rabbit offer systematic review automation and visual mapping of research connections, tools that could assist reviewers in contextualising submissions within broader research landscapes (Scifocus, 2025).

Integrity and Detection Systems

Plagiarism detection platforms such as CrossCheck or iThenticate have enhanced capabilities through AI integration, attempting to identify not only direct plagiarism but also more subtle forms of text reuse. However, these same platforms increasingly offer AI writing detection features, raising concerns about false positives and their effectiveness against sophisticated AI use (Zhou & Soulière, 2025).

Critical Assessment of Tool Effectiveness

Despite proliferating tools, empirical evidence regarding their effectiveness remains limited. A critical challenge involves what researchers call the evaluation paradox: assessing AI tool quality requires human expert judgment, yet the purpose of these tools is ostensibly to augment or replace such judgment (Kousha & Thelwall, 2024).

Early empirical studies have presented mixed findings. Liang et al. (2024) conducted a large-scale study of over 5,000 papers from Nature journals, ICLR, and eLife, and found that 57.4% of researchers found blinded GPT-4 reviews helpful overall, with 82.4% rating them as more helpful than at least some human reviews. However, the same research revealed concerning patterns of AI-generated text in actual conference peer review, suggesting potential quality-consistency gaps.

Critically, tool evaluation must consider not only technical performance but also equity implications. If AI tools are predominantly accessible to well-resourced institutions while unavailable to researchers in low-resource settings, their deployment may inadvertently deepen existing global research inequities (Zhou & Soulière, 2025). This represents a fundamental tension: AI promises democratization of review capacity while simultaneously risking concentration of sophisticated tools among already-advantaged actors.

Implications for Scholarly Practice and Future Directions

The Hybrid Model: AI as Augmentation, Not Replacement

Emerging consensus suggests that AI's appropriate role involves augmentation of human expertise rather than substitution for it. This hybrid model positions AI tools as handling routine, time-intensive tasks—format compliance checking, reference verification, statistical validation, preliminary plagiarism detection—while reserving substantive evaluation for human experts (Doskaliuk et al., 2025).

The appeal of this model lies in its potential to address reviewer burden without compromising evaluation quality. If AI can efficiently handle technical checks and administrative tasks, human reviewers can focus cognitive resources on questions requiring genuine expertise: (a) research significance, (b) methodological appropriateness, (c) theoretical contribution, and (d) practical implications. This division of labor could theoretically improve both efficiency and quality.

However, critical questions remain about the hybrid model's feasibility and desirability. First, the boundary between routine technical checks and substantive evaluation may be less clear than proponents assume. Statistical validation, for instance, requires contextual judgment about appropriate techniques for specific research questions. Second, hybrid systems risk creating over-reliance dynamics where human reviewers, fatigued by technical complexity, defer excessively to AI recommendations. Third, the model assumes AI tools will remain in supportive roles, yet technological trajectories and economic pressures may push toward increasing automation.

Equity Considerations and Global Access

A critical yet underexamined dimension involves equity implications of AI deployment in peer review. If sophisticated AI tools remain accessible primarily to researchers and institutions in high-income countries, AI integration risks exacerbating existing global inequities in scholarly publishing (Zhou & Soulière, 2025). Based on the provided documents regarding equity and systematic bias in the peer review process, here is a rewritten version of the section that incorporates the specific focus on Indigenous representation and cultural epistemology. The concern operates at multiple levels.

- Epistemological and methodological bias. First, because AI systems are predominantly trained on datasets reflecting Western academic norms, they may embody specific epistemological assumptions or methodological preferences. This creates a systematic disadvantage for researchers who employ alternative approaches or non-Western scholarly frameworks.

- Marginalization of Indigenous knowledge. Second, the use of standardized AI models poses a significant risk to the accurate representation of Indigenous people and cultures. If these tools lack the cultural competence to recognize diverse Indigenous ontologies and local knowledge systems, they may dismiss or misinterpret such contributions, further entrenching historical exclusions within the scientific record.

- Economic and resource disparities. Third, the high cost of sophisticated, proprietary AI tools may create a quality gap between well-funded journals and smaller or open-access publications that cannot afford the necessary licensing fees.

- Asymmetric access and expectations. Finally, as AI-assisted review becomes a normative standard, a digital divide emerges where researchers from under-resourced regions find themselves disadvantaged. They may struggle to meet heightened expectations for review efficiency while simultaneously facing AI-driven evaluations of their own work without having reciprocal access to the same advanced preparation tools.

Addressing these equity concerns requires intentional policy interventions; development of open-source AI tools accessible globally, guidelines ensuring AI deployment does not disadvantage particular methodological traditions or institutional contexts, and systematic monitoring of whether AI integration affects acceptance rates for manuscripts from underrepresented institutions or regions.

Governance Frameworks and Continuous Evaluation

Responsible AI integration demands robust governance frameworks operating at multiple levels. Individual journals must establish clear policies specifying permissible AI uses, disclosure requirements, and enforcement mechanisms. Professional societies should develop field-specific guidelines reflecting disciplinary norms and methodological considerations. Funding agencies need coherent policies balancing innovation with research integrity.

Critically, governance must be adaptive rather than static. AI capabilities evolve rapidly; policies adequate for current systems may become obsolete within months. Zhou and Soulière (2025) recommend policy review cycles of six to twelve months, with mechanisms for rapid response to emerging challenges.

Furthermore, governance should emphasize transparency over detection. Rather than creating adversarial dynamics where AI use is something to hide, policies should foster cultures where thoughtful AI integration is openly discussed, potential benefits and limitations candidly assessed, and continuous learning about effective practices encouraged. This requires shifting from prohibition-focused policies toward disclosure-based frameworks that emphasize accountability and reflective practice.

Research Agenda and Knowledge Gaps

Substantial knowledge gaps limit evidence-based policymaking around AI in peer review. Critical research questions include the following:

1. Comparative effectiveness: How do AI-assisted reviews compare to traditional human reviews across multiple quality dimensions (accuracy, comprehensiveness, fairness, usefulness to authors)?

2. Bias manifestation and mitigation: What forms of algorithmic bias emerge in AI-assisted review, and which mitigation strategies prove effective?

3. Reviewer experience and learning: How does AI assistance affect reviewer skill development, particularly for early-career researchers for whom reviewing provides important professional training?

4. Equity impacts: Does AI deployment affect acceptance rates, review quality, or publication outcomes differentially across author demographics, institutional affiliations, or geographical locations?

5. Gaming vulnerabilities: What strategies might authors employ to manipulate AI review systems, and how can journals defend against such gaming?

6. Long-term cultural effects: How does widespread AI adoption alter scholarly culture, norms of intellectual engagement, and the social functions of peer review beyond mere quality control?

Addressing these questions requires longitudinal studies, randomized controlled trials comparing review modalities, qualitative research on reviewer and editor experiences, and cross-disciplinary collaboration among information scientists, ethicists, and domain experts.

Limitations

This report reflects a fast-evolving policy and tool ecosystem. Policies vary by publisher and can change rapidly; therefore, recommendations here highlight primary policy documents and conservative confidentiality-first interpretations. Empirical evidence on AI in peer review is currently concentrated in conference settings and selected disciplines, and detection/measurement methods (including corpus-level estimation) have uncertainty. Finally, vendor claims about tool performance and privacy protections are not always independently auditable, limiting definitive conclusions about effectiveness and risk.

Conclusion

The integration of artificial intelligence into scholarly peer review represents a pivotal transformation with profound implications for research integrity, epistemic authority, and the social organization of knowledge production. This analysis has revealed fundamental tensions between efficiency imperatives and integrity preservation, between technological capabilities and human judgment, and between innovation and equity.

Current evidence suggests AI can effectively perform bounded technical tasks such as grammar checking, plagiarism detection, and citation verification, but lacks the contextual understanding, theoretical sophistication, and ethical reasoning essential for comprehensive peer review. Major institutional guidelines from ICMJE, COPE, and leading publishers reflect this reality, stressing that AI should augment rather than replace human expertise, that disclosure of AI use is mandatory, and that confidentiality protection remains paramount.

However, significant gaps persist between policy and practice. Empirical research has documented widespread undisclosed AI use in peer review, suggesting that current governance mechanisms have proven insufficient. Detection strategies face technical limitations and risk creating adversarial dynamics. Equity concerns remain inadequately addressed, with potential for AI deployment to deepen rather than mitigate global inequities in scholarly publishing.

Based on these considerations, responsible AI integration requires five essential commitments.

1. Transparency as foundational principle: All AI use in peer review must be disclosed, with clear specification of tools employed and their contribution to evaluation.

2. Human accountability as non-negotiable: Regardless of AI assistance, individual reviewers bear full responsibility for review content and recommendations.

3. Confidentiality as inviolable: Manuscript content must not be transmitted to platforms that lack robust data protection, and proprietary AI tools must demonstrate compliance with confidentiality requirements.

4. Equity as explicit consideration: Policies must actively address the potential for AI deployment to advantage particular institutions, methodologies, or geographical contexts, with a commitment to equitable access.

5. Adaptive governance as ongoing obligation: Given rapid technological evolution, policies require regular review and revision, with mechanisms for addressing emergent challenges.

Approaching from a holistic and critical perspective, the scholarly community must resist simplistic narratives positioning AI as either salvation or catastrophe. The technology is neither inherently beneficial nor harmful; rather, its impact depends fundamentally on how it is deployed, governed, and integrated into existing social and epistemic practices (Bozkurt et al., 2026). The challenge is not whether to use AI in peer review, its use is already widespread and likely irreversible, but how to shape that use toward ends consistent with scholarly values of rigor, integrity, fairness, and epistemic humility.

This demands sustained engagement from all stakeholders, including (a) publishers developing responsible implementation frameworks, (b) editors establishing clear expectations and providing reviewer guidance, (c) funding agencies articulating coherent policies, and (d) researchers themselves approaching AI tools with both openness to potential benefits and critical awareness of limitations and risks. Only through such collective, thoughtful engagement can the scholarly community navigate this transformation while

preserving the essential functions of peer review, namely ensuring research quality, maintaining public trust in scientific knowledge, and advancing human understanding.

Acknowledgements

Based on *Academic Integrity and Transparency in AI-assisted Research and Specification Framework* (Bozkurt, 2024), the author of this report acknowledges that the paper was proofread and reviewed with the assistance of DeepL, Grammarly, and Google's Gemini (versions as of February 2026), complementing the human editorial process. The human authors critically assessed and validated the content to maintain academic rigor. The authors also assessed and addressed potential biases inherent in the AI-generated content. The final version of the paper is the sole responsibility of the human authors.

References

Abdelwahab, M. (2024). Artificial intelligence common good in research and academics. *The Scholarship Without Borders Journal, 3*(1). https://doi.org/10.57229/2834-2267.1058

American Journal Experts. (2018). Peer review: How we found 15 million hours of lost time. *AJE*. https://www.aje.com/en/arc/peer-review-process-15-million-hours-lost-time

Ateriya, N., Sonwani, N. S., Thakur, K. S., Kumar, A., & Verma, S. K. (2025). Exploring the ethical landscape of AI in academic writing. *Egyptian Journal of Forensic Sciences, 15*(1). https://doi.org/10.1186/s41935-025-00453-1

Bakla, A. (2023). ChatGPT in academic writing and publishing: An overview of ethical issues. In G. Kartal (Ed.), *Transforming the language teaching experience in the age of AI* (pp. 89–101). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-6684-9893-4.ch005

Ben Saad, H., Dergaa, I., Ghouili, H., Ceylan, H. İ., Chamari, K., & Dhahbi, W. (2025). The assisted technology dilemma: A reflection on AI chatbots use and risks while reshaping the peer review process in scientific research. *AI & Society, 40*(7), 5649–5656. https://doi.org/10.1007/s00146-025-02299-6

Biswas, S., Dobaria, D., & Cohen, H. L. (2023). ChatGPT and the future of journal reviews: A feasibility study. *The Yale Journal of Biology and Medicine, 96*(3). https://doi.org/10.59249/SKDH9286

Bozkurt, A. (2024). GenAI et al.: Cocreation, authorship, ownership, academic ethics and integrity in a time of generative AI. *Open Praxis, 16*(1), 1–10. https://doi.org/10.55982/openpraxis.16.1.654

Bozkurt, A., Crompton, H., Farrow, R., Kukulska-Hulme, A., Dron, J., West, R., Palalas, A. (Aga)., Bower, M., Xiao, J., Tlili, A., Henriksen, D., Pazurek, A., Huijser, H., Chiu, T. K. F., Jandrić, P., Jordan, K., Curry, J., Kimmons, R., Cukurova, M., Reeves, T., Hwang, G.-J., Shea, P., Lodge, J., Weller, M., Ng, D., & Asino, T. I. (2026). Redefining Educational Technology: A Critical Collaborative Inquiry. *Open Praxis, 18*(2), 192–211. https://doi.org/10.55982/openpraxis.18.2.1117

Bozkurt, A., Crompton, H., & Fell Kurban, C. (2026). The devil is in the det[ai]ls: AI agents, ghost students, and the crisis of verified presence in an agentic AI world. *Open Praxis, 18*(1), 1–12. https://doi.org/10.55982/openpraxis.18.1.1145

Celik, S. U. (2025). Integrating artificial intelligence into scientific writing: A narrative review for clinical and surgical researchers. *The American Journal of Surgery, 250*, 116657. https://doi.org/10.1016/j.amjsurg.2025.116657

Checco, A., Bracciale, L., Loreti, P., Pinfield, S., & Bianchi, G. (2021). AI-assisted peer review. *Humanities and Social Sciences Communications, 8*(1). https://doi.org/10.1057/s41599-020-00703-8

Clark, T. A. (2025). Ethical use of artificial intelligence (AI) in scholarly writing. *Journal of Pediatric Surgical Nursing, 14*(3), 85–91. https://doi.org/10.1177/23320249251343881

Collu, M. G., Salviati, U., Confalonieri, R., Conti, M., & Apruzzese, G. (2025, August 28). *Publish to perish: Prompt injection attacks on LLM-assisted peer review*. arXiv. https://doi.org/10.48550/arXiv.2508.20863

COPE Council. (2025). *COPE focus on artificial intelligence*. https://publicationethics.org/cope-focus/cope-focus-artificial-intelligence

COPE Council. (2023). *COPE position—Authorship and AI: English*. https://doi.org/10.24318/cCVRZBms

Delanghe, J. R. (2024). The ethical aspects of AI in scientific publishing. *Journal of the International Federation of Clinical Chemistry, 37*(1), 177–180. https://pmc.ncbi.nlm.nih.gov/articles/PMC12882074/

Doskaliuk, B., Zimba, O., Yessirkepov, M., Klishch, I., & Yatsyshyn, R. (2025). Artificial intelligence in peer review: Enhancing efficiency while preserving integrity. *Journal of Korean Medical Science, 40*(7). https://doi.org/10.3346/jkms.2025.40.e92

Dwivedi, Y. K., Malik, T., Hughes, L., & Albashrawi, M. A. (2024). Scholarly discourse on GenAI's impact on academic publishing. *Journal of Computer Information Systems,* 1-16. https://doi.org/10.1080/08874417.2024.2435386

Enoh, U. (2026). Artificial intelligence in academic publishing and research writing: A comprehensive review. *Scribe Science: Multidisciplinary Journal, 1*(1), 9–13. https://scribescience.org/index.php/ss/article/view/6

Granjeiro, J. M., Cury, A. A. D. B., Cury, J. A., Bueno, M., Sousa-Neto, M. D., & Estrela, C. (2025). The future of scientific writing: AI tools, benefits, and ethical implications. *Brazilian Dental Journal, 36*, e25-6471. http://dx.doi.org/10.1590/0103-644020256471

Gurnal, P., & Rana, L. (2025). Artificial intelligence and publishing ethics: A narrative review and SWOT analysis. *Cureus, 17*(5), e84098. https://doi.org/10.7759/cureus.84098

Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology, 15*(4), ep464. https://doi.org/10.30935/cedtech/13605

International Committee of Medical Journal Editors. (2024). *Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals*. http://www.icmje.org/icmje-recommendations.pdf

Jiang, Y., & Ng, A. (2025). *TechOverview*. https://paperreview.ai/tech-overview

Kitamura, F. C. (2023). ChatGPT is shaping the future of medical writing but still requires human judgment. *Radiology, 307*(2), e230171. https://doi.org/10.1148/radiol.230171

Kocak, B., Onur, M. R., Park, S. H., Baltzer, P., & Dietzel, M. (2025). Ensuring peer review integrity in the era of large language models: A critical stocktaking of challenges, red flags, and recommendations. *European Journal of Radiology Artificial Intelligence, 2*, 100018. https://doi.org/10.1016/j.ejrai.2025.100018

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task*. arXiv. https://doi.org/10.48550/arXiv.2506.08872

Kousha, K., & Thelwall, M. (2024). Artificial intelligence to support publishing and peer review: A summary and review. *Learned Publishing, 37*(1), 4–12. https://doi.org/10.1002/leap.1570

Lee, J., Lee, J., & Yoo, J. J. (2025). The role of large language models in the peer-review process: Opportunities and challenges for medical journal reviewers and editors. *Journal of Educational Evaluation for Health Professions, 22*(4). https://doi.org/10.3352/jeehp.2025.22.4

Leung, T. I., de Azevedo Cardoso, T., Mavragani, A., & Eysenbach, G. (2023). Best practices for using AI tools as an author, peer reviewer, or editor. *Journal of Medical Internet Research, 25*, e51584. https://doi.org/10.2196/51584

Liang, W., Izzo, Z., Zhang, Y., Lepp, H., Cao, H., Zhao, X., Chen, L., Ye, H., Liu, S., Huang, Z., McFarland, D. A., & Zou, J. Y. (2024). *Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews*. arXiv. https://doi.org/10.48550/arXiv.2403.07183

Lin, Z. (2025). *Hidden prompts in manuscripts exploit AI-assisted peer review*. arXiv. https://doi.org/10.48550/arXiv.2507.06185

Luo, Z., Yang, Z., Xu, Z., Yang, W., & Du, X. (2025). LLM4SR: A survey on large language models for scientific research. *arXiv*. https://doi.org/10.48550/arXiv.2501.04306

Mann, S. P., Aboy, M., Seah, J. J., Lin, Z., Luo, X., Rodger, D., Zohny, H., Minssen, T., Savulescu, J., & Earp, B. D. (2025). *AI and the future of academic peer review*. arXiv. https://doi.org/10.48550/arXiv.2509.14189

Matsubara, S. (2026). AI use in peer review: Strict regulation may still be needed. *International Journal of Gynecology & Obstetrics*. https://doi.org/10.1002/ijgo.70885

Mollaki, V. (2024). *AI tools in peer reviewing: Challenges and needs* [Conference presentation]. Office Français de l'Intégrité Scientifique. https://www.ofis-france.fr/wp-content/uploads/2025/05/09-Session2-OFIS_MOLLAKI_AI-Ethics-review_FINAL.pdf

Nature Portfolio. (2025). *Editorial policies on artificial intelligence (AI)*. https://www.nature.com/nature-portfolio/editorial-policies/ai

openRxiv. (2025, November 10). *Enabling options for review: From training and transparency to author-centered AI tools*. https://openrxiv.org/enabling-review-options/

Rohit, K., & Verma, M. (2024). *Ethical AI shaping scholarly communication: Challenges and opportunities*. 12th Convention PLANNER 2024. https://ir.inflibnet.ac.in/server/api/core/bitstreams/0ce09b9a-3be2-4b5f-a918-930b65e1d17b/content

Sabet, C. J., Bajaj, S. S., Stanford, F. C., & Celi, L. A. (2023). Equity in scientific publishing: Can artificial intelligence transform the peer review process? *Mayo Clinic Proceedings: Digital Health, 1*(4), 596–600. https://doi.org/10.1016/j.mcpdig.2023.10.002

Sage. (n.d.). *Artificial intelligence policy*. https://www.sagepub.com/journals/publication-ethics-policies/artificial-intelligence-policy

Scifocus. (2025). *Discover the 10 best peer review tools 2025*. https://www.scifocus.ai/blogs/10-best-peer-review-tools-2025

Scuderi, G. R., Taunton, M. J., Browne, J. A., & Mont, M. A. (2026). The challenges with artificial intelligence in scientific writing. *The Journal of Arthroplasty, 41*(2), 299–303. https://doi.org/10.1016/j.arth.2025.12.001

Semrl, N., Feigl, S., Taumberger, N., Bracic, T., Fluhr, H., Blockeel, C., & Kollmann, M. (2023). AI language models in human reproduction research: Exploring ChatGPT's potential to assist academic writing. *Human Reproduction, 38*(12), 2281–2288. https://doi.org/10.1093/humrep/dead207

Sun, Z. (2025). Large language models in peer review: Challenges and opportunities. *Scientometrics, 130,* 5503–5546. https://doi.org/10.1007/s11192-025-05440-w

Wang, G., Çukur, T., Kruger, U., Ferina, J., & Shan, H. (2026). Editorial AI reviewer (AIR) trial for responsible, secure, and efficient peer review. *IEEE Transactions on Medical Imaging, 45*(3), 867–869. https://doi.org/10.1109/TMI.2026.3658770

Ye, R., Pang, X., Chai, J., Chen, J., Yin, Z., Xiang, Z., Dong, X., Shao, J., & Chen, S. (2024). *Are we there yet? Revealing the risks of utilizing large language models in scholarly peer review*. arXiv. https://doi.org/10.48550/arXiv.2412.01708

Zhou, H., & Soulière, M. (2025, August 25). From detection to disclosure: Key takeaways on AI ethics from COPE's forum. *The Scholarly Kitchen*. https://scholarlykitchen.sspnet.org/2025/08/25/from-detection-to-disclosure-key-takeaways-on-ai-ethics-from-copes-forum/

Zielinski, C., Winker, M. A., Aggarwal, R., Ferris, L. E., Heinemann, M., Lapeña, J. F., Jr., Pai, S. A., Ing, E., Citrome, L., Alam, M., Voight, M., & Habibzadeh, F. (2023). Chatbots, generative AI, and scholarly manuscripts: WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. *Colombia Medica, 54*(3), e1015868. https://doi.org/10.25100/cm.v54i3.5868