v1: 17 July 2026

Review Article

The Role of Artificial Intelligence in the Lifecycle of Scientific Manuscripts: Authoring, Reviewing, and Editorial Selection

Preprinted: 18 January 2026
Peer-approved: 17 July 2026

Qeios, Vol. 8 (2026)
ISSN: 2632-3834

Jose L. Domingo[1]

1. School of Medicine, Universitat Rovira i Virgili, Spain

The integration of Large Language Models (LLMs) and Artificial Intelligence (AI) into scientific publishing is accelerating, driven by systemic crises in peer review and academic economic pressures. This manuscript, presented as a critical commentary and policy analysis, provides a three-part examination: a) AI as a co-authoring tool, balancing its democratizing potential against risks like citation fabrication, b) AI's proficiency in technical review versus its inability to assess novelty, and c) the risk of amplified bias in AI-driven editorial decisions. In the context of peer review, LLMs are being increasingly used for tasks such as preliminary technical verification and language editing. However, their adoption raises critical concerns. Recent evidence confirms that citation hallucination remains a persistent threat, although emerging mitigation strategies such as Retrieval-Augmented Generation (RAG) and verified knowledge integration are being developed to address this challenge. Furthermore, 2025 surveys indicate that over 50% of researchers utilize AI during peer review, often violating existing policies. This shift occurs within an exploitative model that relies on unpaid labor while charging substantial Article Processing Charges (APCs). To address these challenges, this paper proposes a framework for sustainable, human-centered integration of AI. A hybrid model is proposed in which AI is restricted to technical verification and efficiency, while judgments on scientific merit, ethics, and paradigm-shifting research are reserved for appropriately valued human experts. The economic feasibility of compensating reviewers, the allocation of verification responsibilities within editorial workflows, and the need for clear legal accountability frameworks are examined. Maintaining scientific credibility requires both the ethical integration of AI and a fundamental reform of the economic structures governing research communication.

Corresponding author: Jose L. Domingo, joseluis.domingo@urv.cat

 doi.org/10.32388/9AOB4V.3

1. Introduction

The world of academic publishing is currently facing a major crisis: a massive increase in new journals, many papers, a growing shortage of willing peer reviewers, and a business model that profits from researchers' hard work without giving much back[1][2][3][4]. By early 2026, this unsustainable system will be simultaneously challenged and complicated by the rapid integration of Artificial Intelligence (AI). AI encompasses a broad range of computational tools, with Large Language Models (LLMs) like ChatGPT, Claude, Deep Seek, Grok, Gemini, and many more, being a prominent subset focused on language processing. These tools are now ubiquitous in labs and editorial offices[5]. A 2025 survey of 1,600 academics reported that over one-half used AI tools during the peer-review process[6], frequently relying on LLMs for tasks such as grammar checking, content summarization, and preliminary technical screening. This adoption creates a paradox. While AI promises to alleviate burdens like linguistic inequality and technical error detection, it introduces existential threats to data integrity through fabricated citations and the potential automation of scientific conservatism[7][8].

This manuscript presents a holistic examination of the role of AI across the publishing triad: authorship, review, and editorial decision. It expands upon previous critiques of publishing economics, which have highlighted the ethical contradiction of charging authors article processing charges (APCs) while reviewers work for free. Here, the potential of AI to address logistical crises is juxtaposed with its inherent limitations and dangers. The point is the following: can AI tools truly solve a peer-review crisis that is, at its core, an economic and systemic failure? In this paper, recent empirical studies are analyzed, the complementary and conflicting capabilities of human versus AI reviewers are compared, and a governance framework that prioritizes scientific integrity over mere efficiency is proposed. The analysis draws upon the latest guidelines from the Committee on Publication Ethics[9], which emphasize transparency, human accountability, and mandatory disclosure of AI use across all stages of the publication process. Unlike studies that focus narrowly on AI's technical capabilities in manuscript screening or language editing[10], this analysis situates AI adoption within the broader political economy of scholarly publishing, arguing that technological fixes cannot resolve a crisis rooted in labor exploitation and profit-driven incentives. The main argument is that unless the financial model of publishing is fundamentally changed, using AI could further frustrate scientists and damage the credibility of research. This manuscript is presented as a critical commentary and policy analysis rather than a systematic review, aimed at researchers, editors, publishers, and policymakers engaged with AI in scholarly communication. It advances a clearly normative argument for specific policy measures and ethical safeguards while remaining anchored in available empirical evidence.

1.1. Scope, literature selection, and methodological approach

This analysis is presented as a critical commentary and policy synthesis rather than a systematic review or empirical study. The literature foundation draws on a targeted search of peer-reviewed publications, preprints, and policy documents issued between 2023 and early 2026, with particular emphasis on empirical studies evaluating AI/LLM performance in manuscript authoring, reviewing, and editorial selection. Primary databases consulted include PubMed, Google Scholar,

arXiv, and Web of Science. Inclusion criteria prioritized: (a) original empirical studies or large-scale surveys on AI adoption in scientific publishing; (b) technical evaluations of LLM performance metrics (hallucination rates, accuracy of citation generation, detection of plagiarism); (c) policy guidance from recognized bodies (Committee on Publication Ethics, UNESCO, OECD); and (d) economic analyses of peer review sustainability and article processing charge structures. Gray literature (preprints, working papers, and reports from research institutions) was included when contributing original data unavailable in peer-reviewed sources. The search strategy prioritized recent publications (2023–2025) to capture the rapidly evolving landscape of AI integration, but foundational literature on peer review economics and bias in algorithmic decision-making extends to earlier periods. This approach prioritizes breadth and relevance over exhaustive completeness, reflecting the scope of a policy-oriented commentary. The limitations of this methodology are acknowledged: the completeness of the literature search cannot be guaranteed, publication bias toward reporting dramatic failures or successes of AI tools may skew the evidence base, and the rapid evolution of LLM capabilities means that some citations may reflect model versions or performance characteristics that have since been superseded. Where specific studies are cited, their date, model version, and methodological context are noted to permit readers to assess the currency and applicability of findings.

2. The Roots of the Crisis: Unsustainable Economics and the Exploitation of Peer Reviewers

2.1. The broken social contract of peer review

Peer review operates on a fragile gift economy. Authors receive career-advancing credit for publication, while reviewers receive only intangible academic capital. The Global State of Peer Review report[11] revealed that most researchers felt overburdened[12][13], a situation that has intensified in recent years. Reviewer acceptance rates have significantly decreased, and editors now routinely invite numerous colleagues to secure 2-3 reviews, creating a massive hidden workload[14][15][16]. This reviewer fatigue is not merely a logistical problem. It is, rather, a symptom of a system that undervalues critical scholarly labor.

2.2. The gold open access paradox and the rise of the profit-maximizing publisher

The transition to digital and Open Access (OA) publishing was announced as a democratizing force. Instead, it has meant a lucrative new business model. As noted in the analyses of this economic model, authors who want to see their paper published as Open Access must pay high amounts (APCs > 3,000–5,000 Euros). In contrast, the reviewers of that paper do not receive a single euro/dollar for their work. This creates a fundamental ethical breach. If publication is a paid service, why is its quality control unpaid labor? This model has been perfected by so-called "predatory" journals but has also been eagerly adopted by legacy publishers[17].

Modern science is increasingly treated like a commercial product, where the goal is to move papers through the system as quickly as possible rather than carefully maintaining the quality of knowledge. This profit-driven atmosphere is the real reason AI is being introduced to the peer-review process. Supporters claim AI

will solve the shortage of human reviewers, but this solution hides a deeper problem: the reliance of the system on the free labor and generosity of experts. By using AI to handle the massive influx of papers, we might simply be treating the symptoms of the volume crisis while allowing the actual disease, the exploitation of researchers, to become even more deeply rooted.

3. AI in Manuscript Preparation

3.1. Democratization and enhanced efficiency

AI writing assistants offer tangible benefits, particularly in promoting linguistic equity. For non-native English speakers, these tools can reduce the linguistic tax, improving readability, grammar, and adherence to academic conventions, which may positively influence editorial first impressions[18]. More than just editing, AI can help structure complex arguments, draft technical methodology sections from structured data, and assist in initial literature synthesis, potentially increasing researcher productivity[19].

3.2. The hallucination epidemic and the erosion of verifiability

The most severe and well-documented threat from the use of AI to write manuscripts is citation hallucination; that is to say, the generation of plausible but entirely fictitious references. While this is not an occasional or random error, it arises from the statistical, pattern-matching nature of LLMs, which generate text based on probabilistic predictions rather than verified factual retrieval.

Critical developments in mitigation strategies warrant detailed examination, as several emerging approaches show measurable promise in reducing—though not eliminating—hallucination rates. Retrieval-Augmented Generation (RAG) systems, which ground LLM responses in verified external knowledge bases rather than relying solely on training data, have demonstrated meaningful reductions in fabrication rates when properly implemented (e.g., reducing citation errors by 30–50% in controlled settings: [20][21]). Model fine-tuning on curated, verified scientific literature, such as training on CrossRef or PubMed Central metadata, has similarly shown incremental improvements in reference accuracy for domain-specific applications. However, these strategies face fundamental limitations: source conflicts within knowledge bases can generate conflicting information, data poisoning (where false information infiltrates training or retrieval sources) remains a persistent vulnerability, and RAG systems require continuous curation and updating to remain accurate across rapidly expanding scientific literature. Furthermore, human-in-the-loop verification workflows, in which AI-generated citations are automatically cross-checked against multiple authoritative databases (CrossRef, PubMed, Scopus) before acceptance, represent a practical intermediate approach, reducing reviewer workload while maintaining accuracy verification. Studies of such workflows indicate error detection rates exceeding 95% when properly configured[22]. Nevertheless, none of these mitigation strategies can substitute for genuine understanding of scientific literature or remove the fundamental responsibility from authors to verify their own citations. Despite advances in RAG and fine-tuning, current rates of fabrication remain unacceptably high for scientific work. These emerging tools offer partial shields, improving accuracy in specific, well-curated domains, but they cannot replace true comprehension,

leaving the fundamental risk of AI-generated misinformation substantially unresolved.

Recent (2023–2025) evidence reveals the alarming scope of this problem across various models and disciplines (see Table 1, legend, for evidence quality indicators). Various studies have consistently demonstrated that while newer models show incremental improvements, fabrication and error rates remain unacceptably high for scientific work. Walters and Wilder[8] documented that while GPT-3.5's 55% fabrication rate improved to GPT-4's 18%, subsequent testing of GPT-4o (2025) showed it fabricated approximately 20% of citations while introducing errors in 45% of real references. Among fabricated citations that included DOIs, 64% linked to real but completely unrelated papers, making detection extremely difficult without careful verification[23][24]. Cheng et al.[25], evaluating ChatGPT-4o performance in medical writing, found that only 7% of ChatGPT-generated DOIs were fully accurate, with 38% being incorrect or fabricated. Similarly, in a study conducted by Kim et al.[26] using four LLMs (including GPT-4, Claude, DeepSeek, and Gemini) in a comparative analysis across geographic regions, it was found that they frequently fabricated DOIs in academic citations, which was especially notable for lower-income countries (e.g., India and Bangladesh). Error rates even exceeded 80%, while hallucinations were found to be most common in recent publications, varying by model, prompt engineering approach, and domain of study, highlighting geographic and economic biases. A 2025 cross-model study by Cabezas-Clavijo and Sidorenko-Bautista[27], comparing eight LLMs (ChatGPT, Copilot, Claude, Gemini, Perplexity, Grok, DeepSeek, and others), revealed that only 26.5% of AI-generated references were entirely correct, with nearly 40% being erroneous or completely fabricated, highlighting the persistent unreliability across different language models. AI errors increase significantly in specialized fields, where finding an expert to check the facts is most difficult[22].

Study	Year	AI Model(s) Tested	Sample Size / Methodology	Key Finding: Fabrication / Error Rate	Additional Notes
Walters and Wilder[8]	2023	GPT-3.5, GPT-4	Comparative analysis	GPT-3.5: 55% fabrication; GPT-4: 18% fabrication	Early evidence of improvement in newer models, but persistent hallucination risk
Linardon et al.[23]	2025	ChatGPT (GPT-4o)	Topic-specific mental health citations	20% fabrication; 45% contains errors in real references	Hallucination rates spike for niche topics where expert verification is difficult
Cabezas-Clavijo and Sidorenko-Bautista[27]	2025	Various LLMs (ChatGPT, Claude, Copilot, Gemini, Perplexity, Grok, DeepSeek, others)	400 references across 5 academic disciplines	26.5% entirely correct; 39.8% erroneous or completely fabricated; 33.8% real but with partial errors	Cross-model comparison; Grok and DeepSeek performed best; Copilot, Perplexity, and Claude had the highest hallucination rates
Cheng et al. [25]	2025	ChatGPT	DOI-specific analysis	38% of generated references contained incorrect or fabricated DOIs; only 7% had fully accurate DOIs	DOI fabrication is particularly deceptive as it appears authentic
Kim et al. [26]	2025	4 LLMs (unspecified)	Geographic bias analysis (India, Bangladesh, other low-income countries)	Error rates exceeded 80% for some models in certain regions; geographic and economic biases were evident	Hallucinations are most common in recent publications; bias varies by model and geographic context
Buchanan et al.[28]	2024	ChatGPT	Detection of fabricated citations in economics	Fabricated citations were attributable to real researchers in credible-looking journal templates	Demonstrates the deceptive realism of hallucinations; difficult to detect without verification

Table 1. Summary of empirical evidence on AI citation fabrication rates and error patterns (2023–2025)

The reliability of published research is under serious threat as fake references become increasingly common. Buchanan et al.[28] showed that these artificial citations are deceptively realistic, often attributing fake findings to actual researchers within credible-looking journal templates. While emerging strategies such as RAG systems, model fine-tuning on verified datasets, and automated citation verification workflows offer partial improvements in reducing hallucinations in specific, well-curated domains (with documented accuracy improvements of 30–50% in controlled laboratory settings), they face inherent limitations in scalability, source reliability, and adaptability to emerging scientific literature. These tools cannot, at present, substitute for a true understanding of scientific literature or remove the fundamental obligation of authors to verify citations. This places the primary responsibility for fraud detection on authors, who must rigorously verify all references for both existence and fit with the cited content. While reviewers and editors share responsibility for identifying suspicious citations, the fundamental obligation rests with the author. Submission processes may reasonably include an explicit author declaration confirming that all references have been manually verified for accuracy and relevance, a task that is becoming increasingly demanding in an age of data saturation. Although the adoption of RAG tools and verification workflows offers a partial shield, improving accuracy in curated environments, it cannot replace true comprehension, leaving the fundamental risk of AI-generated misinformation substantially unresolved.

4. Human Reviewers vs. AI Chatbot Reviewers

4.1. AI as a technical auditor: strengths in verification

Where AI shows genuine promise is in augmenting the technical aspect of peer review. Automated systems can perform with superhuman consistency in specific, rule-based tasks:

a. Statistical and Mathematical Check: AI systems demonstrate high accuracy in rule-based technical verification tasks: in biomedical research, studies using automated statistical verification tools show detection of p-value inconsistencies and mathematical errors at rates exceeding 90% accuracy for standard reporting frameworks ([7] evaluated on manuscripts submitted to the Journal of Korean Medical Science, 2024–2025). However, performance varies considerably by discipline, with such rule-based approaches proving less effective in fields employing non-standard statistical methods or qualitative designs.
b. Image Manipulation Detection: AI algorithms can screen for duplicated, spliced, or otherwise manipulated images more thoroughly than the human eye.
c. Plagiarism and Text Recycling: LLMs can cross-check text against massive databases more efficiently than standard software.

d. Compliance Checking: Ensuring adherence to reporting guidelines (e.g., CONSORT, PRISMA) and ethical statements.

These capabilities address known weaknesses in human review, where statistical errors are frequently undetected[29]. AI can act as a first-line technical auditor, freeing human reviewers from tedious verification work. The efficiency of AI-based technical auditing differs markedly among academic fields. In biomedical research, standardized reporting frameworks such as CONSORT and PRISMA establish clear guidelines that AI systems can assess with relative consistency. By contrast, in disciplines like the theoretical sciences or the humanities, where methods are interpretive and context-dependent, rule-based AI evaluation proves less effective. On the other hand, computational fields can gain significant advantages from automated code analysis and reproducibility checks, while clinical research demands rigorous statistical verification. These disciplinary differences highlight the need for AI auditing tools designed in accordance with the methodological norms of each field, rather than relying on a universal, one-size-fits-all approach.

4.2. The irreplaceable human roles: judgment, novelty, and context

This is where the comparison becomes stark and highlights the non-fungible value of the human reviewer. The core responsibilities of peer review extend far beyond technical checking:

a. Judging Novelty and Significance: Can the work change thinking in the field? Because contemporary AI systems are trained predominantly on existing scientific literature, they tend to reproduce established patterns of thought, which may inadvertently discourage submissions that challenge prevailing paradigms or pursue high-risk, high-impact avenues of inquiry[30]. They lack the intuition to recognize a brilliant flaw or a new, revolutionary idea.
b. Contextual and Ethical Reasoning: Does the work fit into the broader landscape? Are its ethical implications sound? Human reviewers draw on deep domain knowledge, an understanding of societal impact, and professional ethics.
c. Constructive Dialogue: Peer review is an academic dialogue. A human reviewer can suggest alternative experiments, propose collaborative connections, or engage in a dialectic that improves the science. AI generates feedback based on patterns, not understanding or creativity.

As has been emphasized, the reviewer is "the basic point of the chain of quality," a fact that is unquestionable. Replacing human judgment with computer-generated scores would change peer review from a careful evaluation into a routine filtering process. This shift could strengthen existing prejudices and prevent new ideas from emerging. The fear is not that AI will be wrong but that it will be consistently and invisibly biased toward the conventional.

4.3. The hybrid review model

A sustainable future might lie in a hybrid model that benefits from the strengths of both:

- Role of AI: Technical pre-screening. Perform initial checks for statistics, plagiarism, image integrity, and formatting. Flag potential issues for human attention. Summarize manuscript content for editor triage.

- Human's Role: Evaluate novelty, significance, creativity, and broader impact. Interpret AI-generated flags with contextual wisdom. Make final recommendations on acceptance/rejection. Engage in constructive dialogue with authors.

This model respects humans as the referees of science while using AI as a powerful, precise tool. Figure 1 illustrates this complementary framework, showing how AI functions as a technical auditor, whereas humans retain authority over scientific judgment across the entire manuscript lifecycle. It is worth noting that within emerging hybrid models, the traditional, uncompensated system for human reviewers, motivated by the desire to stay informed and contribute to the scholarly community, still remains a viable and valued basis for many researchers. Accordingly, the debate on compensation represents only one among several possible avenues for reform. A key operational issue is determining who should implement and interpret AI-based technical audits: the editor during the initial manuscript screening or the peer reviewer during the formal evaluation stage. Each option has distinct advantages. When editors employ AI tools at the triage stage, fundamental technical problems can be identified early, preventing unnecessary reviewer workload. In contrast, reviewer-led AI assessments may provide deeper, discipline-specific insights into any flagged concerns. The approach recommended here places primary responsibility for AI-driven technical checks at the editorial pre-screening phase, enabling editors to spot and prioritize potential problems efficiently. Reviewers would then receive concise AI-generated technical reports as supporting material while maintaining full independence in their expert scientific judgment.

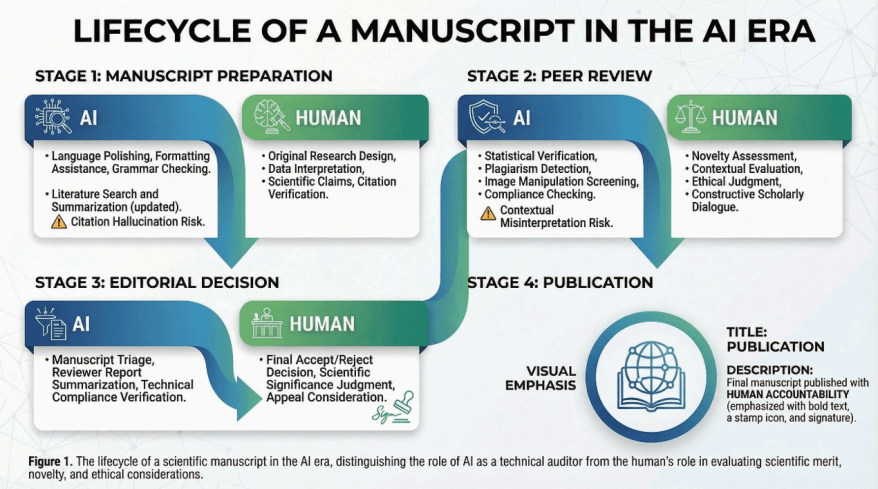

Figure 1. The lifecycle of a scientific manuscript in the AI era, distinguishing the role of AI as a technical auditor from the human's role in evaluating scientific merit, novelty, and ethical considerations.

4.4. Comparative analysis: a structured overview

To synthesize the arguments developed in Sections 4.1–4.3, Table 2 provides a structured comparison of the main capabilities, limitations, and implications of human reviewers versus AI (LLM) reviewers across key assessment dimensions in scientific peer review. This overview highlights that, while AI excels in speed

and technical verification, it remains fundamentally inadequate for evaluating novelty, contextual, and ethical issues. This also ensures the delivery of truly constructive academic feedback, which emphasizes the essential role of responsible human judgment.

Assessment Dimension	Human Reviewer	AI (LLM) Reviewer	Implications for Peer Review
Technical Verification	Prone to fatigue, oversight, and variable standards. High expertise but inconsistent.	Exceptional speed, consistency, and recall for defined tasks (statistics, plagiarism detection).	AI is well suited for initial objective technical screening, reducing human workload and error rates.
Novelty and Significance	Core strength. Uses intuition, experience, and field foresight to judge paradigm-shifting potential.	Biased toward incremental science that conforms to patterns in training data.	AI cannot reliably identify high-risk, transformative research; exclusive reliance may homogenize science.
Contextual and Ethical Reasoning	Integrates societal impact, ethical nuance, and interdisciplinary context.	Limited to text-based pattern recognition; lacks genuine ethical understanding or situational awareness.	Human oversight remains essential for evaluating ethical compliance, dual-use concerns, and societal implications.
Constructive Feedback	Provides creative suggestions, proposes new experiments, and engages in scholarly dialogue.	Produces structured but often generic feedback lacking original insight.	The iterative and developmental nature of peer review depends on human intellectual contribution.
Bias	Subject to conscious and unconscious biases related to affiliation, gender, nationality, or academic networks.	Inherits and may amplify systemic biases embedded in training data.	Both systems are bias-prone. However, AI bias is systemic and highly scalable. While AI bias can, in principle, be tested and audited (unlike much unconscious human bias), it remains embedded in training data and may be difficult to detect without systematic evaluation. This makes it potentially more dangerous at scale than individual human reviewer bias.
Economic Model	Relies on unpaid academic labor; sustainability is increasingly compromised by reviewer fatigue.	Operates at near-zero marginal cost once deployed.	Economic incentives favor AI adoption but risk devaluing human expertise and undermining review quality.
Accountability	Professionally and ethically accountable for evaluations and recommendations.	Lacks legal, ethical, or professional accountability;	Responsibility for the integrity of the scientific record must remain with

doi.org/10.32388/9AOB4V.3

Assessment Dimension	Human Reviewer	AI (LLM) Reviewer	Implications for Peer Review
		operates as a non-transparent system.	accountable human reviewers.

Table 2. Comparative capabilities of human reviewers vs. AI (LLM) reviewers in scientific peer review

5. To Pay the Human or Replace with AI

The advancement of capable AI-based peer-review systems places publishers at a decisive economic juncture. They must determine whether to continue relying on uncompensated expert reviewers, introduce mechanisms for fair remuneration, or shift toward inexpensive algorithmic review processes. This represents more than a simple operational shift. It requires deep ethical and philosophical discussions about the fundamental value of human expertise in a digital world. Table 3 presents four distinct scenarios for the future of peer review economics, contrasting the status quo, full AI automation, a fair hybrid model, and community-led reclamation. This structured overview makes explicit how different combinations of AI use and reviewer compensation may shape costs, incentives, and ultimately, the likely outcomes for the quality and credibility of science.

doi.org/10.32388/9AOB4V.3

Scenario	Model	Pros	Cons	Likely Outcome for Science
Status Quo Exploitation	Continued reliance on unpaid human reviewers while maintaining high article processing charges (APCs).	Low direct financial cost for publishers; preserves elements of human judgment.	Increasing reviewer fatigue; ethical concerns; progressive decline in review quality and system sustainability.	Systemic Collapse. Progressive degradation of peer review quality leading to the erosion of journal credibility.
Full AI Automation	Replacement of human reviewers with fully automated AI-based review systems, while APCs remain unchanged.	Near-zero marginal cost per review; extremely rapid editorial turnaround.	Inability to assess true novelty; amplification of algorithmic bias; absence of accountability.	Homogenized stagnation: A technically consistent but incremental and uninspired scientific literature.
Fair Hybrid Model	AI-based tools conduct technical screening, complemented by compensated human reviewers for intellectual judgment.	Sustainable labor model; combines AI efficiency with human expertise; preserves scientific creativity.	Higher operational costs for publishers; requires structural reform of current publishing economics.	Sustainable integrity that maintains scientific quality, supports innovation, and ethically values expert labor.
Community Reclamation	Scientific- or society-owned journals with reviewer compensation funded through transparent, non-profit budgets. AI is used selectively as a technical auditing tool under community governance, supporting (but not replacing) compensated human reviewers; editorial infrastructure is rebuilt to eliminate commercial intermediaries.	Alignment of incentives with academic community values; elimination of profit-driven distortions.	Requires coordinated collective action and development of new editorial infrastructure.	Renewed trust: A scholar-led publishing ecosystem prioritizing quality, openness, and knowledge dissemination.

Table 3. Scenarios for the future of peer review economics

doi.org/10.32388/9AOB4V.3

5.1. The case for compensating human reviewers

The argument for payment is grounded in fairness, sustainability, and quality preservation.

Fairness in a for-profit system: As some analysts have argued, the current model is ethically unjustifiable. If the publication process generates important benefits, its most critical quality-control component deserves financial recognition.

Sustainable labor model: The gift economy is collapsing. Payment, whether direct stipends (e.g., €200–€500 per review, a range informed by the pilot study referenced below, which must be absorbed by publisher profits, not with higher APCs) or review credits redeemable for APCs, creates a sustainable incentive structure. It is important to recognize that determining the appropriate level of compensation demands a cost-benefit analysis tailored to each journal's financial framework, since even relatively small honoraria can meaningfully influence reviewer behavior. A 2025 pilot by the Journal of Clinical Medicine offering modest honoraria (€100–€300 per review) reported increases of approximately 40% in reviewer acceptance rates, with corresponding 25% decreases in median review completion time[7] (study population: 240 invited reviewers across oncology and cardiology). This pilot provides the empirical basis for the compensation range noted above, though broader cost-benefit analyses across journal types, geographic regions, and research disciplines would be needed before universal adoption.

Preserving quality and commitment: Compensation professionalizes the role. Paid reviewers are more likely to accept invitations within their expertise, meet deadlines, and provide thorough, constructive feedback. It formally recognizes review as essential scholarly work, not a secondary activity[14][31].

Counteracting commercialization: Fair compensation is a step toward rebalancing power. It acknowledges that the research community provides the core product, challenging the role of publishers as mere intermediaries.

5.2. The Publisher temptation: "zero-cost"

For profit-driven publishers, the economic appeal of AI is irresistible: a one-time software investment that eliminates the recurring costs (in time and effort) of managing a volunteer reviewer pool. AI promises instant, 24/7 review, reducing time-to-first decision and eliminating the reviewer chase. This maximizes throughput, a key metric in high-volume, APC-driven models. Marketing AI as an unbiased tool (despite evidence to the contrary) could be used to deflect criticism about editorial decisions, creating a tag of technical neutrality. In a competitive market, the first major publisher to fully automate peer review for certain article types could undercut others on cost and speed, forcing widespread adoption regardless of quality concerns[32]. It should be recognized that not all pressures for automation arise from profit motives. Smaller non-profit and society journals often face real resource constraints, making volunteer reviewer recruitment increasingly difficult. For these, AI tools may serve as a necessary adaptation rather than a simple cost-saving measure.

5.3. The risks of replacing human reviewers with AI

The replacement of human reviewers by AI to save costs would be a significant trade-off, sacrificing scientific progress for corporate efficiency. Scientific progress relies on an essential basis of mutual trust within the scientific

community, based on the belief that human experts conduct thorough and impartial evaluations. The adoption of automated peer-review technologies risks undermining this implicit social agreement, as opaque algorithmic decision processes may significantly reduce confidence in the legitimacy and integrity of the scientific record[33][34]. The community trusts that peers have fairly evaluated their work. Replacing them with inscrutable algorithms would fundamentally break this contract, delegitimizing published science. AI systems remain unable to reliably identify true novelty, contextual significance, or ethical complexity, which makes their exclusive use in peer review a direct threat to scientific progress. A fully automated system would create an incremental science filter, publishing only those papers that confirm existing paradigms. Innovative work from unknown labs or on emerging topics would be systematically rejected. Moreover, who is responsible when an AI reviewer misses a fatal flaw, endorses plagiarized work, or rejects a brilliant idea? The publisher would hide behind the black box, leaving authors with no recourse. Replacing human reviewers with AI merely externalizes the cost of quality onto the scientific community, which will suffer from a stagnant, unreliable literature. The cost savings achieved come at the expense of an investment in the integrity of the scientific record itself[35].

6. Algorithmic Bias in Editorial Decisions

Editors are increasingly using AI-driven predictive analytics to select manuscripts, estimating scientific interest based on historical citation data, author prestige, and keyword trends[36]. This poses a profound danger of algorithmic gatekeeping. An AI trained on past successful papers will inherently favor research that resembles past success, systematically disadvantaging interdisciplinary work, studies from less prestigious institutions, and truly novel paradigms. It creates a vicious, self-reinforcing cycle that could homogenize scientific output.

As highlighted in critiques of metric-based assessment, excessive reliance on journal-level indicators (Impact Factor, CiteScore, etc.) already distorts research evaluation. AI-powered interest predictors risk automating and exacerbating this bias, making it more opaque and harder to challenge. The definition of interesting science must remain a dynamic, human social construct, responsive to new challenges and perspectives[37].

7. A Framework for Ethical Integration and Systemic Reform

7.1. Governance, transparency, and mandatory disclosure

Ad hoc policies are insufficient. According to the guidelines from COPE[9], authors must detail how AI was used (e.g., for language polishing, for literature search, for generating figures, etc.), with vague statements being unacceptable. In addition, all final publication decisions must have explicit human approval, while AI output cannot be the sole basis for rejection. Moreover, journals using AI for review or triage must disclose this to authors and provide a mechanism to appeal or request human assessment. Moving toward open review reports can increase accountability for both human and AI-assisted reviews. Major publishers have now established comprehensive AI policies that universally

prohibit AI authorship and mandate transparency regarding AI tool usage, while emphasizing that authors retain full responsibility for all content accuracy and integrity.

A key concern in AI integration is legal liability and accountability. When AI-assisted reviews fail (due to fabricated citations, undetected plagiarism, or biased rejections), current frameworks inadequately assign responsibility among authors, reviewers, editors, publishers, and AI developers. Publisher policies hold authors fully accountable, but this falters as editors increasingly use AI without robust verification. Emerging rules like the EU AI Act deem such systems "high-risk," demanding transparency and oversight. Therefore, publishers need clear frameworks specifying AI use, verification steps, audit trails for human decisions, and author recourse options. Without them, all parties face legal risks and unresolved errors. Global frameworks for AI governance are advancing quickly. In addition to the EU AI Act, several national initiatives (such as the guidance developed by the US AI Safety Institute and the evaluation standards introduced by the UK AI Safety Institute) are beginning to tackle the regulation of high-risk AI uses within research environments. At the international level, the UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) and the OECD's continuing work on AI governance offer key normative references that the academic publishing community should actively consider.

Regarding accountability, emerging legal scholarship supports a tiered liability model. AI developers bear responsibility for ensuring system safety, transparency, and documented limitations in their technical documentation; publishers who deploy AI tools are accountable for verifying that systems are appropriately calibrated to their use case, maintaining audit logs of AI-assisted decisions, and establishing clear protocols for human verification; editors using AI outputs are responsible for confirming that all AI-generated recommendations are reviewed by qualified human experts before implementation; and authors remain fundamentally accountable for the accuracy and integrity of all submitted content, regardless of AI use in preparation. Establishing standardized contractual frameworks between publishers and AI service providers (specifying performance expectations, liability allocation, and transparent reporting of errors) will be essential for clarifying accountability when AI-assisted workflows contribute to publication errors or ethical breaches.

7.2. Addressing the economic problems

If AI integration proceeds without economic reform, it will merely make a profitable but exploitative system more efficient. Incentives must be redesigned. Thus, if APCs are charged, a significant portion must fund reviewer honoraria or systematic rewards (e.g., review credits redeemable for APCs at partner journals) [31]. In turn, the scientific community should promote and progressively adopt society-owned, diamond open-access, and cooperative publishing models. In any case, the formal recognition of peer review in tenure, promotion, and funding decisions is essential.

7.3. Training the next generation of scientists

Academic institutions and research centers must implement compulsory AI literacy programs that provide scientists with a comprehensive understanding of machine learning capabilities, inherent constraints, and ethical vulnerabilities. A key part of this training must be the development of rigorous verification

protocols. Specifically, researchers require the skills to systematically audit AI-generated technical content and validate the authenticity of provided citations. Training programs should specifically address the identification of hallucinated citations, proper citation verification workflows using databases (CrossRef, PubMed, Scopus, etc.), and the responsible use of RAG-enhanced tools while recognizing their limitations. Furthermore, establishing a robust ethical framework is fundamental to distinguishing between legitimate computational assistance and academic dishonesty.

7.4. Synthesis: A governance framework for responsible AI integration

The governance recommendations converge around three pillars: transparency and disclosure, where authors detail AI use in preparation, journals report AI in reviews, and audit trails track decisions; human accountability, mandating explicit human approval for final decisions, prohibiting sole AI-based rejections, and defining liability for failures; and economic sustainability, directing APC revenues toward reviewer compensation, while promoting society-owned cooperative models over commercial intermediaries. Together, these ensure AI augments technical verification while preserving human judgment on merit, novelty, and ethics within an equitable economic structure.

8. Conclusions

The integration of AI into scientific publishing is not a simple question of adoption. It is a complex negotiation at a time of systemic crisis. AI offers powerful tools to enhance technical rigor and manage scale, particularly in addressing the overwhelming burden contributing to reviewer fatigue. However, it simultaneously introduces severe risks to the veracity of the scientific record through hallucinations and to the diversity of scientific thought through algorithmic bias.

The economic crossroads are clear. The path of least resistance for publishers, replacing unpaid human labor with unpaid machine labor, is a direct threat to the progress of science. The only ethically and scientifically defensible path is a model akin to Scenario 3: The Fair Hybrid Model (Table 3). This model, which views AI as a tool for technical tasks while valuing human expertise, represents a strong candidate for reform. However, the solution space remains open. Thus, other models, including those maintaining the traditional uncompensated contribution of referees within a hybrid AI-assisted system, may also offer viable paths forward depending on community values and economic feasibility. This requires viewing AI not as a replacement, but as a tool that handles arduous technical work, thereby making the valuable time of human experts more efficient and sustainable. It must be paired with fair compensation for those experts, reforming the economic model that has brought us to this crisis point. The goal of fully automating research must be rejected, and instead, a human-centered approach, where AI serves only to assist and enhance human work, must be adopted.

In that model, AI handles the technical, the repetitive, and the scalable, while human experts (adequately valued and compensated) retain authority over judgments of significance, novelty, ethics, and creativity. The future of vibrant, credible scientific literature depends on choosing a model that invests in human expertise, not one that seeks to reduce or eliminate costs.

Statements and Declarations

Funding

No specific funding was received for this work.

Potential Competing Interests

No potential competing interests to declare.

Data Availability

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Use of Generative AI

Artificial Intelligence (Skywork AI Smart Slide Tool) was used to assist in the graphic design of Figure 1.

References

1. ^*Horta H, Jung J. (2024). "The Crisis of Peer Review: Part of the Evolution of Science." High Educ Q. **78**(4):e12511. doi:10.1111/hequ.12511.*
2. ^*Routledge D, Pariente N, on behalf of the PLOS Biology Staff Editors. (2025). "On Improving the Sustainability of Peer Review." PLoS Biol. **23**(3):e3003127. doi:10.1371/journal.pbio.3003127.*
3. ^*Clark AD, Myers TC, Steury TD, Krzton A, Yanes J, Barber A, Barry J, Barua S, Eaton K, Gosavi D, Nance R, Pervaiz Z, Ugochukwu C, Hartman P, Stevison LS. (2024). "Does It Pay to Pay? A Comparison of the Benefits of Open-Access Publishing Across Various Sub-Fields in Biology." PeerJ. **12**:e16824. doi:10.7717/peerj.16824.*
4. ^*Clark LM, Wang DK, Adkins BD, Fitzhugh VA, Walker PD, Khan SS, Fadare O, Stephens LD, Coogan AC, Booth GS, Jacobs JW. (2024). "Paying to Publish: A Cross-Sectional Analysis of Article Processing Charges and Journal Characteristics Among 87 Pathology Journals." Acad Pathol. **11**(4):100153. doi:10.1016/j.acpath.2024.100153.*
5. ^*Hosseini M, Eisenman D, Riddle J, Pyle S, Peters A, Cobbe N, Holmes K. (2026). "Guidelines Needed for the Use of AI in the Preparation or Review of IRB, IBC, and IACUC Applications." Account Res. doi:10.1080/08989621.2025.2612564.*
6. ^*Naddaf M. (2026). "More Than Half of Researchers Now Use AI for Peer Review Often Against Guidance." Nature. **649**:273274. doi:10.1038/d41586-025-04066-5.*
7. a, b, c*Doskaliuk B, Zimba O, Yessirkepov M, Klishch I, Yatsyshyn R. (2025). "Artificial Intelligence in Peer Review: Enhancing Efficiency While Preserving Integrity." J Korean Med Sci. **40**(7):e92. doi:10.3346/jkms.2025.40.e92.*
8. a, b, c*Walters WH, Wilder EI. (2023). "Fabrication and Errors in the Bibliographic Citations Generated by ChatGPT." Sci Rep. **13**:14045. doi:10.1038/s41598-023-41032-5.*
9. a, b*COPE Council. (2024). "COPE Position - Authorship and AI - English." Committee on Publication Ethics. doi:10.24318/cCVRZBms.*
10. ^*Carobene A, Padoan A, Cabitza F, Banfi G, Plebani M. (2024). "Rising Adoption of Artificial Intelligence in Scientific Publishing: Evaluating the Role, Risks, and Ethical Implications in Paper Drafting and Review Process." Clin Chem Lab Med. **62**(5):835843. doi:10.1515/cclm-2023-1136.*

11. ^Publons. (2018). "Global State of Peer Review." Clarivate. https://clarivate.com/academia-government/lp/global-state-of-peer-review-report/.
12. ^Breuning M, Backstrom J, Brannon J, Gross BI, Widmeier M. (2015). "Reviewer Fatigue? Why Scholars Decline to Review Their Peers' Work." PS Polit Sci Polit. **48** (4):595600. doi:10.1017/S1049096515000827.
13. ^Fox CW, Albert AYK, Vines TH. (2017). "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. **2**:3. doi:10.1186/s41073-017-0027-x.
14. a, bAczel B, Barwich A-S, Diekman AB, Fishbach A, Goldstone RL, Gomez P, Gundersen OE, von Hippel PT, Holcombe AO, Lewandowsky S, Nozari N, Pestilli F, Ioannidis JPA. (2025). "The Present and Future of Peer Review: Ideas, Interventions, and Evidence." Proc Natl Acad Sci U S A. **122**(5):e2401232121. doi:10.1073/pnas.2401232121.
15. ^Bauchner H, Rivara FP. (2024). "Use of Artificial Intelligence and the Future of Peer Review." Health Aff Sch. **2**(5):qxae058. doi:10.1093/haschl/qxae058.
16. ^Maliki M, Mehmani B. (2024). "Structured Peer Review: Pilot Results From 23 Elsevier Journals." PeerJ. **12**:e17514. doi:10.7717/peerj.17514.
17. ^Domingo JL. (2024). "To Publish Scientific Journals: For Some, the Big Business of the Century." Qeios. doi:10.32388/YTFEF2.2.
18. ^Khalifa AA, Ibrahim MA. (2024). "Artificial Intelligence (AI) and ChatGPT Involvement in Scientific and Medical Writing, a New Concern for Researchers: A Scoping Review." Arab Gulf J Sci Res. **42**(4):17701787. doi:10.1108/AGJSR-09-2023-0423.
19. ^Liang W, Izzo Z, Zhang Y, Lepp H, Cao H, Zhao X, Chen L, Ye H, Liu S, Huang Z, McFarland DA, Zou JY. (2024). "Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews." arXiv. doi:10.48550/arXiv.2403.07183.
20. ^Guu K, Lee K, Tung Z, Pasupat P, Chang M. (2020). "Retrieval Augmented Language Model Pre-Training." In Proceedings of the 37th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 119, pp. 3929 3938). https://proceedings.mlr.press/v119/guu20a.html.
21. ^Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, Kttler H, Lewis M, Yih W-t, Rocktschel T, Riedel S, Kiela D. (2020). "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." Adv Neural Inf Process Syst. **33**:94599474. doi:10.48550/arXiv.2005.11401.
22. a, bChelli M, Descamps J, Lavou V, Trojani C, Azar M, Deckert M, Raynier JL, Clowez G, Boileau P, Ruetsch-Chelli C. (2024). "Hallucination Rates and Reference Accuracy of ChatGPT and Bard for Systematic Reviews: Comparative Analysis." J Med Internet Res. **26**:e53164. doi:10.2196/53164.
23. a, bLinardon J, Jarman HK, McClure Z, Anderson C, Liu C, Messer M. (2025). "Influence of Topic Familiarity and Prompt Specificity on Citation Fabrication in Mental Health Research Using Large Language Models: Experimental Study." JMIR Ment Health. **12**:e80371. doi:10.2196/80371.
24. ^StudyFinds Analysis. (2025). "ChatGPTs Hallucination Problem: Study Finds More Than Half of AIs References Are Fabricated or Contain Errors in Model GPT-4o." StudyFinds. https://studyfinds.com/chatgpts-hallucination-problem-fabricated-references/.
25. a, bCheng A, Calhoun A, Reedy G. (2025). "Artificial Intelligence-Assisted Academic Writing: Recommendations for Ethical Use." Adv Simul. **10**(1):22. doi:10.1186/s41077-025-00350-6.

26. [a], [b] *Kim E, Kipchumba F, Min S. (2025). "Geographic Variation in LLM DOI Fabrication: Cross-Country Analysis of Citation Accuracy Across Four Large Language Models." Publications. **13**(4):49. doi:10.3390/publications13040049.*
27. [a], [b] *Cabezas-Clavijo , Sidorenko-Bautista P. (2025). "Assessing the Performance of 8 AI Chatbots in Bibliographic Reference Retrieval: Grok and DeepSeek Outperform ChatGPT, but None Are Fully Accurate." arXiv. doi:10.48550/arXiv.2505.18059.*
28. [a], [b] *Buchanan J, Hill S, Shapoval O. (2024). "ChatGPT Hallucinates Non-Existent Citations: Evidence From Economics." Am Econ. **69**(1):8087. doi:10.1177/05694345231218454.*
29. [^] *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." J Educ Eval Health Prof. **22**:4. doi:10.3352/jeehp.2025.22.4.*
30. [^] *Maddi A, Sapinho D. (2022). "Article Processing Charges, Altmetrics and Citation Impact: Is There an Economic Rationale?" Scientometrics. **127**(12):73517368. doi:10.1007/s11192-022-04284-y.*
31. [a], [b] *Cheah PY, Piasecki J. (2022). "Should Peer Reviewers Be Paid to Review Academic Papers?" Lancet. **399**(10335):1601. doi:10.1016/S0140-6736(21)02804-X.*
32. [^] *Drozdz JA, Ladomery MR. (2024). "The Peer Review Process: Past, Present, and Future." Br J Biomed Sci. **81**:12054. doi:10.3389/bjbs.2024.12054.*
33. [^] *Kemal O. (2025). "Artificial Intelligence in Peer Review: Ethical Risks and Practical Limits." Turk Arch Otorhinolaryngol. **63**(3):108109. doi:10.4274/tao.2025.2025-8-12.*
34. [^] *Schintler LA, McNeely CL, Witte J. (2023). "A Critical Examination of the Ethics of AI-Mediated Peer Review." arXiv. doi:10.48550/arXiv.2309.12356.*
35. [^] *Resnik DB, Hosseini M. (2025). "The Ethics of Using Artificial Intelligence in Scientific Research: New Guidance Needed for a New Tool." AI Ethics. **5**(2):14991521. doi:10.1007/s43681-024-00493-8.*
36. [^] *Peddi S, Manoharan G. (2026). "Ethical Guidelines for the Thoughtful Implementation of AI in Higher Education." EthAIca. **5**:409. doi:10.56294/ai2026409.*
37. [^] *Park W, Cullinane A, Gandolfi H, Alameh S, Mesci G. (2024). "Innovations, Challenges and Future Directions in Nature of Science Research: Reflections From Early Career Academics." Res Sci Educ. **54**:2748. doi:10.1007/s11165-023-10102-z.*

Declarations

Funding: No specific funding was received for this work.
Potential competing interests: No potential competing interests to declare.