



Artificial Intelligence as an Assistant in Research Evaluation and Peer Review: A Critical Case for Augmentation Without Replacement

Valentine Joseph Owan¹ , Mercy Valentine Owan^{2*} , Eyo Etim Akpandem³ 

¹ Department of Educational Psychology, University of Calabar, NIGERIA

² Ultimate Research Network, NIGERIA

³ Department of Public Administration, University of Calabar, NIGERIA

* Correspondence: mercyvarcy@gmail.com

CITATION: Owan, V. J., Owan, M. V., & Akpandem, E. E. (2026). Artificial Intelligence as an Assistant in Research Evaluation and Peer Review: A Critical Case for Augmentation Without Replacement. *Educational Point*, 3(2), e174. <https://doi.org/10.71176/edup/19655>

ARTICLE INFO

Received: 3 May 2026

Accepted: 22 September 2026

OPEN ACCESS

ABSTRACT

Peer review is important in maintaining the quality of scholarly publishing, but the process faces challenges such as shortage of reviewers, delays, inconsistent judgements, and overlooked errors. These challenges have increased the need to consider the appropriate use of artificial intelligence (AI) in peer review. Completely excluding AI or allowing it to function as an autonomous reviewer may not adequately address these problems. A Human–AI Augmented Review model is therefore proposed, in which AI supports clearly defined tasks under human supervision, while reviewers and editors retain responsibility for interpretation, ethical judgement, and editorial decisions. Peer review involves both routine technical checks and complex scholarly evaluation. AI can assist with repetitive and rule-based tasks such as manuscript screening, consistency checking, reviewer identification, and identification of reporting gaps. Human reviewers are still needed to assess methodology, scholarly contribution, and ethical issues. Effective use of AI requires a clear division of responsibilities, verification points within the review process, and governance measures to address bias, hallucinations, confidentiality, and excessive reliance on automation. Careful and limited use of AI can improve efficiency and reduce avoidable errors while maintaining accountability and trust in peer review.

Keywords: scholarly publishing, research integrity, peer-review process, editorial decision-making, publication ethics

INTRODUCTION

Peer review is the principal mechanism through which the scientific quality of a scholarly article is evaluated before publication. Its credibility, however, depends largely on the voluntary efforts of a relatively small group of qualified reviewers. As submission volumes continue to increase and research becomes more specialised,

this arrangement is coming under increasing pressure. Researchers have reported prolonged review times, heavy reviewer workloads, low agreement between reviewers, and persistent biases associated with gender, language, and institutional prestige (Lin et al., 2023; Mann et al., 2025). These challenges have raised important questions about the ability of contemporary peer review to maintain consistent standards of scholarly evaluation.

The pressures affecting peer review are beyond the increasing number of manuscript submissions. Reviewers are expected to assess theoretical framing, methodological quality, analytical rigour, reporting accuracy, originality, and the significance of research findings. These responsibilities are often performed within limited time while reviewers also attend to other professional duties. Under such conditions, even experienced reviewers may apply uneven scrutiny or overlook important issues (Checco et al., 2021; Lin et al., 2023). In response to these pressures, journals have introduced reviewer training, reporting checklists, and alternative review models. However, these measures have not kept pace with the growing demand for expert review (Mann et al., 2025).

Against this background, AI is becoming increasingly integrated into editorial workflows. Automated tools are already used for plagiarism detection, formatting checks, reviewer recommendation, journal matching, and preliminary manuscript screening. The main purpose of these applications is to improve efficiency rather than replace editorial judgement (Farber, 2024; Kousha & Thelwall, 2024). Researchers have also reported that AI-assisted reviewer selection can reduce recruitment time and broaden reviewer pools. Comparative studies suggest that AI-generated peer reviews can achieve substantial agreement with human assessments and may attain comparable or higher scores on selected review-quality dimensions (Saad et al., 2024; Zancanaro et al., 2026). However, agreement between assessments does not necessarily establish that either review correctly identifies a manuscript's substantive weaknesses or that AI and human reviewers are equivalent across all dimensions of review quality. However, current AI systems still have important limitations in consistency, depth of analysis, and the assessment of originality. However, similarity in review content or evaluation scores should not be interpreted as evidence of equivalent review quality. Agreement may reflect shared evaluative criteria without establishing that either assessment correctly identifies the manuscript's substantive weaknesses. The extent of agreement may also depend on the instructions given to AI systems and the criteria used by individual reviewers.

The growing use of AI has generated different opinions. Some authors envisage highly automated or even autonomous peer-review systems (Lin et al., 2023), whereas others have cautioned that greater reliance on AI could increase bias, weaken scholarly judgement, compromise confidentiality, and encourage superficial evaluation (Kankanhalli, 2024; Seghier, 2025). Neither position fully addresses the realities of contemporary peer review. Complete automation overlooks the interpretive and ethical dimensions of scholarly evaluation, while complete rejection of AI ignores the pressures already affecting editorial practice. Similar concerns have been raised in broader discussions of AI in higher education, where researchers have argued that debates focused only on risks or benefits are not sufficient for responsible adoption. It is therefore important for AI to be integrated in ways that strengthen human expertise while preserving critical judgement and accountability (Owan, 2026).

A more balanced position is to treat AI as supervised analytical support rather than as an autonomous reviewer. Current evidence indicates that AI performs best in well-defined activities such as reviewer selection, manuscript screening, reporting checks, and other routine editorial tasks (Ngo, 2025). By contrast, evidence that AI improves methodological evaluation, assessment of originality, ethical reasoning, or editorial judgement is still limited and inconsistent. For this reason, AI is better suited to supporting peer review than replacing it. On this basis, attention is given to current evidence on AI-assisted peer review, the aspects of peer review that are suitable for AI support, and those that require human judgement. A governance framework is also considered for the integration of AI into editorial workflows. The main concern is therefore how AI can

strengthen peer review while preserving the accountability, fairness, and scholarly judgement that underpin scientific publishing.

WHAT PEER REVIEW QUALITY REALLY MEANS

Peer review quality is not a single concept because peer review involves different activities requiring different forms of judgement. The extent to which AI can be used therefore depends on the particular activity being considered. Although there is no single framework for defining peer-review quality, methodological rigour, originality, reporting quality, fairness, ethical conduct, and useful reviewer feedback are commonly regarded as important aspects of effective peer review (Chong & Lin, 2024; Sizo et al., 2025). These aspects can be grouped into six broad domains according to the type of assessment involved.

The first two domains are technical and methodological assessment. Technical and reporting accuracy involves checking internal consistency, completeness of reporting, agreement between the text and tables, accuracy of sample-size reporting, adequacy of methodological description, and compliance with reporting guidelines (García-Costa et al., 2022). Methodological adequacy involves checking whether the research design, analytical procedures, and conclusions are appropriate for the research question. Such assessment requires disciplinary knowledge and careful consideration of the evidence used to support the conclusions (García-Costa et al., 2022). The next two domains are scientific contribution and research integrity. Scientific contribution refers to the extent to which a manuscript advances knowledge through original ideas, meaningful findings, or theoretical development (Paliya & Jha, 2024). Ethical and integrity evaluation involves research ethics, participant protection, conflict-of-interest disclosure, responsible reporting, and adherence to accepted scholarly standards. These responsibilities go beyond procedural compliance and form an important part of peer review (Kadaifci et al., 2025). The final two domains are fairness and developmental feedback. Fairness and bias management relate to institutional prestige, language proficiency, geographical location, and other factors that may affect review decisions. Such bias may threaten equitable evaluation (Kadaifci et al., 2025). Developmental feedback refers to constructive comments that help authors improve the quality, clarity, and presentation of their work (García-Costa et al., 2022; Sizo et al., 2025).

These six domains indicate that peer review involves both routine checks and higher-level evaluation. While some technical checks can be performed using established reporting standards, other assessments require disciplinary knowledge, ethical reasoning, and professional judgement. Therefore, AI may assist with some aspects of peer review, but human reviewers are still needed to make important evaluative decisions in other aspects and all-round.

RE-PROBLEMATIZING PEER REVIEW

Peer review is often criticised because of its weaknesses, but these are not simply occasional failures within an otherwise dependable system. Delays, inconsistent evaluations, reviewer fatigue, and overlooked errors are recurring problems in a process that operates under considerable structural and cognitive pressure. These problems arise largely from how contemporary peer review is organised rather than from the shortcomings of individual reviewers. This helps to explain why successive reforms have produced only modest improvements and why AI is increasingly being considered to support rather than replace peer review.

Structural Pressures on Peer Review

Reviewer scarcity is often attributed to increasing academic workloads and declining willingness to accept review invitations. Although these factors contribute to the problem, they do not fully explain the difficulty of recruiting reviewers. Peer review depends on the voluntary efforts of a relatively small group of scholars, while the demand for expert evaluation continues to increase as submission volumes rise, journals expand,

interdisciplinary research grows, and methodological specialisation reduces the number of suitable reviewers (Owan & Ogabor, 2026). As a result, reviewing responsibilities become concentrated among a limited number of experienced researchers, increasing fatigue and extending review times.

This pressure represents a structural imbalance rather than a temporary shortage of reviewers. When experienced reviewers decline invitations, editors have to depend on a smaller group of available experts, placing greater demands on those who continue to review. Research on biomedical publishing has reported that a relatively small proportion of scholars complete a disproportionately large share of peer reviews (Kovanis et al., 2016). Other researchers have reported declining acceptance rates for review invitations and increasing recruitment difficulties, although the extent varies across disciplines and publication outlets (Meyerson et al., 2025; Morley et al., 2025). These reports indicate that workload concentration and recruitment difficulties are important constraints on contemporary peer review.

Publication delays are also part of this problem. Slow review cycles delay the dissemination of knowledge, postpone error correction, and may create unequal consequences for researchers, particularly those at the beginning of their careers or working on fixed-term appointments (Mann et al., 2025). In higher education systems where promotion depends heavily on publication output, such delays may have important career implications (Owan & Asuquo, 2022). Editors also spend considerable time on routine activities such as preliminary screening, compliance checks, and reviewer recruitment. This leaves less time for careful evaluation of manuscripts and reviewer reports (Farber, 2024; Regmi & Vaishya, 2025).

Cognitive Limits and Variation in Review Quality

Effective peer review depends on sustained attention, but human attention is limited. As manuscripts become longer, methodologically more complex, and increasingly interdisciplinary, reviewers have to assess theoretical framing, research design, statistical analyses, interpretation, and reporting quality while managing other professional responsibilities. Under these conditions, some aspects of a manuscript may receive more attention than others.

Consequently, inconsistencies requiring systematic cross-checking, such as discrepancies between hypotheses and analyses, tables and text, or reported results and conclusions, may escape detection (Checco et al., 2021; Cox et al., 2025). Heavy workload can also reduce the detail and developmental value of reviewer feedback. It has been reported that increasing workload and limited attentional capacity can reduce the thoroughness of complex evaluations (Almukhtar et al., 2025; Severin & Chataway, 2021). Although experience helps reviewers manage complexity more effectively, it does not remove the limits of human attention (Gegenfurtner et al., 2011; Ghazanfar et al., 2015). Variation in review quality should therefore be understood not only in terms of reviewer expertise but also as a consequence of cognitive demands.

Differences in reviewer recommendations also require careful interpretation. Some disagreement reflects legitimate differences in theoretical perspectives, methodological preferences, and disciplinary traditions. However, reviewer judgements may also be influenced by institutional prestige, language proficiency, reputational signals, and other visible characteristics unrelated to manuscript quality (Kulal et al., 2025). Editors must therefore distinguish genuine scholarly disagreement from inconsistency arising from uneven standards, implicit bias, or differences in reviewer engagement (Kousha & Thelwall, 2024). Confidence in peer review depends on making this distinction as consistently as possible.

Why Adaptation Has Become Necessary

The challenges facing peer review are closely related. Reviewer scarcity increases workload, heavier workloads limit attention, reduced attention contributes to inconsistent evaluations, and inconsistent evaluations increase the demands placed on editors. Similar concerns about the sustainability of scholarly labour have

been raised in research on contemporary publishing systems (Owan & Ogabor, 2026). These pressures help to explain the growing interest in AI-assisted peer review. Current applications focus mainly on reducing routine workload, improving efficiency, and identifying issues that reviewers may overlook while leaving interpretation and evaluative judgement to human reviewers (Carobene et al., 2024; Checco et al., 2021). The important question, therefore, is not whether peer review has limitations, but how these limitations can be managed without weakening the principles that support fair and credible scholarly evaluation.

WHAT AI CAN CONTRIBUTE WITHOUT BECOMING THE JUDGE

The empirical evidence supporting AI use in peer review is not equally strong across all tasks. Stronger evidence exists for reviewer identification, workflow efficiency, plagiarism detection, manuscript screening, and other routine quality-control activities (Farber, 2024; Kousha & Thelwall, 2024). However, evidence relating to methodological evaluation, assessment of originality, improvement of review quality, and editorial decision-making is still limited and mixed (Russo et al., 2025). Another consideration is the role of prompting in AI-assisted evaluation. Reviewers differ in the criteria they prioritise, and these differences may be transferred to AI systems through reviewer-specific instructions. Consequently, observed agreement between AI and human assessments may partly reflect the evaluative framework encoded in the prompt rather than the independent capacity of the AI system to assess scholarly quality. Comparative studies should therefore distinguish agreement with individual reviewers from the accuracy of assessments against independently established criteria. The case for AI augmentation is therefore stronger for systematic and well-defined tasks than for evaluative judgement.

The use of AI in peer review does not necessarily mean replacing human reviewers. Similar Human–AI arrangements have been proposed in educational measurement, where AI supports pattern recognition, consistency checking, and routine analytical activities while important judgements remain under human control (Owan et al., 2023). AI is therefore more useful for activities that are systematic and repetitive, and that may be affected by fatigue and time pressure. This does not suggest that reviewers lack competence. Rather, it recognises the cognitive demands of scholarly review. When AI is used as a support tool rather than a decision-maker, it can address some weaknesses in the review process without taking over disciplinary judgement, ethical reasoning, or interpretation. This position is also consistent with recommendations that AI should strengthen human capability rather than replace independent scholarly judgement (Owan, 2026).

AI as a Systematic Checker; Not a Verdict Machine

AI tools are already used in routine manuscript handling, including plagiarism detection, duplicate-submission checks, and format compliance screening (Kousha & Thelwall, 2024; Regmi & Vaishya, 2025). These activities follow established rules and depend largely on pattern recognition. Early-stage screening can identify procedural problems before reviewers become involved, thereby saving reviewer time and reducing unnecessary cognitive demands. At this stage, AI does not determine the scholarly merit of a manuscript. It identifies departures from established criteria, while editors retain responsibility for interpretation and decision-making.

AI as a Reviewer-Finding Assistant

Identifying suitable reviewers is one of the demanding and time-consuming aspects of editorial work. Editors traditionally depend on their knowledge of the field, previous reviewers, professional networks, and publication records. Although these approaches are useful, they may limit the pool of potential reviewers. AI-assisted systems can assist by searching publication databases, analysing research profiles, and identifying individuals

whose expertise matches the subject of a submitted manuscript (Checco et al., 2021; Farber, 2024). However, reviewer selection cannot be reduced to algorithmic matching. Editors must still consider conflicts of interest, reviewer workload, disciplinary representation, and the suitability of each candidate. AI can broaden the search and improve efficiency, but responsibility for reviewer selection remains with the editor.

AI as a Clarity and Coherence Tool

Poorly written manuscripts may require reviewers to spend valuable time interpreting unclear arguments before evaluating the research itself. Language problems, weak organisation, and ambiguous expression can increase cognitive demands and divert attention from methodological or theoretical issues. Large language models can improve grammar, identify ambiguous wording, summarise lengthy passages, and suggest clearer organisation (Salih et al., 2025). These functions can make manuscripts easier to read, particularly for authors writing in an additional language. By reducing avoidable language barriers, reviewers can concentrate more on the scientific quality of the work.

AI as an Early Warning System for Reporting Gaps

Every weakness in a manuscript does not result from poor research. Some arise from incomplete reporting, missing methodological information, or inconsistencies introduced during drafting and revision. AI can identify such omissions by comparing information across different parts of a manuscript and identifying departures from expected reporting patterns (Checco et al., 2021; Cox et al., 2025). Such alerts should not be taken as evidence of poor-quality research. Rather, they identify areas that may require closer examination, leaving reviewers to decide whether the reported deficiencies affect the validity or interpretation of the work.

AI as a Consistency Support Tool

Maintaining consistency across a manuscript becomes increasingly difficult as research reports become longer and more analytically complex. Discrepancies between the abstract, methods, results, tables, and conclusions may go unnoticed, particularly when reviewers work under competing time demands. AI can compare different sections of a manuscript and identify possible contradictions, missing links, or claims that are not supported by the reported evidence (Carobene et al., 2024; Checco et al., 2021). Such comparisons can direct reviewers' attention to sections requiring closer scrutiny. Determining whether these inconsistencies represent important problems, however, continues to depend on disciplinary knowledge and scholarly judgement.

AI as a Tool for Structuring Reviewer Feedback

Effective peer review depends not only on identifying weaknesses but also on communicating them clearly and constructively. AI can assist reviewers by organising comments, summarising major issues, generating draft outlines, and arranging observations in line with journal review templates (Drori & Te'eni, 2024; Kankanhalli, 2024). When used carefully, these functions can help reviewers prepare reports that are clearer and easier for authors and editors to follow. Nevertheless, AI-generated text should not be incorporated without careful review. Reviewers remain responsible for checking the accuracy, relevance, and tone of every comment included in their reports.

AI as Decision Support

Reviewer fatigue, limited attention, heavy workload, and increasing cognitive demands can make it difficult to detect reporting inconsistencies and internal errors, even among experienced reviewers (Almukhtar et al., 2025; Carobene et al., 2024; Mann et al., 2025). Although experience helps reviewers manage complex manuscripts, it does not remove the limits of human attention (Gegenfurtner et al., 2011; Ghazanfar et al.,

2015). AI can assist with routine monitoring, consistency checks, reporting gaps, and other pattern-based activities that may be overlooked during manual review (Checco et al., 2021; Salih et al., 2025). Such assistance can allow reviewers to devote more attention to disciplinary expertise, methodological reasoning, theoretical interpretation, and ethical judgement. However, reviewers must still determine whether issues identified by AI are scientifically or methodologically important. Such decisions depend on disciplinary knowledge, professional experience, and careful interpretation (Seghier, 2025). This supportive use of AI is consistent with evidence that AI can strengthen researchers' capacity for independent scholarly work rather than replace human expertise (Akpan & Owan, 2026). AI is therefore better used to support peer review than to replace human judgement. This position is also consistent with recommendations for human-centred workflows in which verification, professional responsibility, and critical evaluation remain central to academic practice (Owan, 2026). Responsibility for evaluation and editorial decisions should therefore remain with human reviewers and editors.

WHY AI CANNOT REPLACE HUMAN PEER REVIEW

The limitations of peer review do not justify replacing human reviewers with AI. Although AI can assist with specific and well-defined tasks, using it as an autonomous reviewer may create technical, ethical, and practical problems. These problems are explained below, with explanations offered on why AI is better used to support human reviewers than to replace them.

Technical Limitations of AI

Large language models can generate fluent and convincing text, but fluency does not guarantee accuracy. They may produce fabricated citations, unsupported criticisms, or incorrect interpretations. Research has reported that AI-generated reviews can combine useful observations with confidently stated errors, particularly when manuscripts are technically complex or structurally unconventional (Carabantes et al., 2023; Cox et al., 2025). Such errors may mislead editors, disadvantage authors, and affect editorial decisions. Peer review also requires judgement about originality, theoretical contribution, methodological soundness, and disciplinary relevance. These judgements depend on disciplinary knowledge, professional experience, and familiarity with ongoing scholarly debates. Although AI can assist with language improvement and structural assessment, it lacks the interpretive understanding and ethical reasoning needed to assess scientific contribution (Farber, 2025). Its developmental functions through mentorship and scholarly dialogue are also limited (Salih et al., 2025).

Ethical and Governance Challenges

AI systems trained on historical publication data may reproduce inequalities associated with language proficiency, institutional prestige, geographical location, and disciplinary visibility (Checco et al., 2021; Salih et al., 2025). Greater efficiency is of limited value if it comes at the expense of fairness. Careful monitoring and appropriate governance are therefore needed to reduce such risks. Responsibility for editorial decisions also cannot be transferred to AI. Although fully autonomous reviewing remains uncommon, some researchers have proposed systems that generate review reports or publication recommendations with minimal human involvement (Lu et al., 2026). Since publication decisions affect careers, reputations, and the integrity of the scholarly record, moral and professional responsibility must remain with human reviewers and editors (Seghier, 2025). Confidentiality is another concern. Many AI applications rely on cloud-based services, creating risks for unpublished manuscripts and sensitive information when appropriate safeguards are absent (Ebadi et al., 2025). Machine-generated comments may also reduce editorial credibility when accepted without careful scrutiny. Editorial policies should therefore specify approved tools, secure infrastructure, and

appropriate procedures for AI use. Editors and reviewers also need sufficient AI literacy to examine AI-generated outputs critically rather than accept them at face value (Owan, 2026).

Human Oversight Remains Essential

Even when AI is used only as a support tool, reviewers may place excessive confidence in its outputs because they appear systematic and reliable. Such reliance may weaken independent critical judgement and reduce opportunities, particularly for early-career reviewers, to develop evaluative expertise through practice (Salih et al., 2025). Automation bias is therefore an important concern in AI-assisted review. Research comparing AI-assisted and human peer review has reported that AI can improve efficiency and reproduce some features of reviewer reports, but its performance is not sufficiently reliable for independent use (Carabantes et al., 2023; Drori & Te'eni, 2024). Similar findings have been reported in clinical and medical publishing, where large language models sometimes produce evaluations resembling those of human reviewers but also introduce systematic differences that may affect editorial decisions (Joachim et al., 2025). Human-AI collaboration is therefore preferable to replacement. AI is most useful for routine and well-defined tasks, while interpretation, ethical judgement, and publication decisions require human oversight (Ebadi et al., 2025; Farber, 2025). This position is consistent with recommendations that AI should complement human expertise (Owan, 2026) and with evidence that its value is greater when it strengthens independent learning and scholarly work rather than replacing human decision-making (Akpan & Owan, 2026).

A HUMAN–AI AUGMENTED REVIEW FRAMEWORK

Much of the debate on AI-assisted peer review treats AI as a single intervention. In practice, responsible augmentation depends on the tasks assigned to AI, where its outputs enter the review process, how such outputs are verified, and who is responsible for editorial decisions. The framework proposed here has three elements: a governing principle that limits AI authority, a task-based division of responsibilities, and a staged review process with human verification. These elements are intended to improve efficiency while preserving accountability and scholarly judgement in peer review.

Core Principle: AI Assists, Humans Decide

The framework is based on a simple principle: AI may assist evaluation, but it cannot assume responsibility for it. Activities such as screening, information synthesis, pattern detection, and attention management are suitable for AI assistance. However, decisions about methodological rigour, scholarly contribution, ethical acceptability, fairness, and publication outcomes require human judgement because they depend on disciplinary expertise, interpretive reasoning, and professional accountability (Kankanhalli, 2024; Regmi & Vaishya, 2025). Although most researchers support augmentation rather than replacement, some have proposed highly automated or autonomous peer-review systems (Lin et al., 2021; Lu et al., 2026). These proposals reinforce the need to separate technical assistance from evaluative authority.

Two practical principles follow from this position. First, AI-generated outputs should be treated as prompts for further review and not as evidence for editorial decisions. A flagged inconsistency, methodological concern, or generated summary should draw attention to an issue, but should not determine its importance. Reviewers and editors must verify AI-assisted observations before using them in an evaluation. Second, responsibility for editorial decisions cannot be transferred to AI. Publication decisions have academic, ethical, and professional consequences. Reviewers and editors therefore remain responsible for evaluative judgements, while AI use should be disclosed in accordance with journal or institutional policy (Seghier, 2025). Thus, AI may inform judgement, but the final judgement belongs to the responsible reviewer or editor.

Division of Labour Across Review Tasks

The principle that AI assists while humans decide should guide everyday editorial practice. Peer review consists of several tasks with different cognitive and ethical demands. Some tasks involve identifying patterns, checking compliance, organising information, and verifying reports, while others require disciplinary expertise, interpretation, and judgement. The former can be assigned to supervised AI assistance, whereas the latter remain human responsibilities. **Table 1** presents this division of labour. AI supports routine screening, consistency checks, reporting verification, and information organisation. Reviewers and editors retain responsibility for interpretation, ethical reasoning, reviewer selection, and publication decisions. The framework therefore assigns AI according to the nature of the task rather than simply its technical capability. This arrangement can reduce routine workload while keeping responsibility for important decisions with reviewers and editors.

A Staged Review Process with Embedded Verification

Defining the roles of AI and human reviewers is only part of responsible augmentation. AI-generated outputs must also be checked before they influence editorial decisions.

Table 1. Division of Labour in Human–AI Augmented Review

Peer Review Activity	AI Support (Under Human Supervision)	Human Responsibility (Cannot Be Delegated)	Supporting Citations
Initial manuscript screening	Plagiarism detection, duplicate-submission checks, format checks, scope checks, and compliance with journal requirements	Review of screening results, interpretation of borderline cases, and application of journal standards	(Kousha & Thelwall, 2024; Regmi & Vaishya, 2025)
Reviewer selection	Identification of potential reviewers, expertise matching, and wider database searches	Selection of reviewers, assessment of conflicts of interest, consideration of reviewer workload and disciplinary balance, and final approval	(Checco et al., 2021; Farber, 2024)
Reporting and consistency checks	Detection of missing information, reporting gaps, inconsistencies, differences between tables and text, mismatches between hypotheses and analyses, and unsupported claims	Verification of flagged issues, assessment of their importance, and interpretation of their implications	(Carobene et al., 2024; Cox et al., 2025; Drori & Te'eni, 2024; Joachim et al., 2025; Regmi & Vaishya, 2025; Seghier, 2025)
Language and reviewer support	Grammar and readability checks, identification of unclear wording, preparation of summaries, draft review outlines, and organisation of reviewer comments	Critical review, revision, preparation of reviewer comments, and responsibility for the final review report	(Kankanhalli, 2024; Salih et al., 2025)
Scholarly evaluation	None	Evaluation of originality, methodological quality, theoretical contribution, disciplinary relevance, ethical acceptability, and publication decisions	(Farber, 2025; Seghier, 2025)
Quality monitoring and governance	Detection of possible AI-generated text, identification of unusual review patterns, and analysis of reviewer trends	Interpretation of audit results, policy development, protection of confidentiality, oversight of fairness, and responsibility for editorial decisions	(Carabantes et al., 2023; Checco et al., 2021; Joachim et al., 2025; Russo et al., 2025; Seghier, 2025)

Although AI systems can produce fluent and well-organised responses, they may also generate errors, fabricated information, biased outputs, and conclusions expressed with unwarranted confidence. Researchers have therefore supported human-in-the-loop approaches rather than fully automated review. For example, GPT-4-assisted reviewing reduced reviewer workload but did not perform consistently across all review situations (Drori & Te'eni, 2024). Similarly, comparisons between large language models and journal reviewers found only moderate agreement and substantial differences in rejection decisions, indicating the need for human verification before AI outputs inform editorial judgement (Joachim et al., 2025). **Table 2** translates these principles into practical verification steps by indicating what should be checked at each stage of AI-assisted review and the risks of failing to do so. The main concern is not the use of AI itself, but the acceptance of its outputs without careful scrutiny. Automation bias, excessive reliance on AI assistance, and inadequate human oversight have therefore been identified as important concerns in AI-assisted peer review (Drori & Te'eni, 2024; Seghier, 2025).

Verification should not be treated as an optional safeguard. It should be part of every stage of the review process. Many of the activities described below are already part of responsible editorial practice. The main purpose of the framework is to bring these activities together by defining the roles of AI and human reviewers, identifying what should be checked, assigning responsibility for editorial decisions, and setting limits on the use of AI. The stages below explain how these principles can be applied throughout the peer-review process while maintaining efficiency and accountability.

Table 2. Potential Governance Risks if Verification Is Not Conducted

Stage of Review	What Should Be Checked	Risk if Not Checked
AI-assisted manuscript screening	Editors should confirm that AI screening results are accurate and relevant before making editorial decisions.	Editorial decisions may rely on inaccurate or misleading AI outputs.
AI-generated alerts	Reviewers and editors should determine whether flagged issues are genuine and important.	False alarms or incorrect suspicions may affect fair judgement.
Reviewer recommendations	Editors should confirm reviewer expertise, suitability, and possible conflicts of interest.	Existing biases related to prestige, professional networks, or visibility may be reinforced.
Language and readability checks	Reviewers should decide whether language problems affect the scientific quality of the manuscript.	Language difficulties may be mistaken for weaknesses in the research itself.
Reporting and statistical checks	Reviewers should determine whether missing information or statistical concerns affect the validity of the study.	Routine checklist compliance or inaccurate AI feedback may influence evaluation inappropriately.
Consistency checks	Reviewers should verify inconsistencies identified across different sections of the manuscript.	Automation bias may encourage uncritical acceptance of AI-generated outputs.
AI-generated summaries and draft reviews	Reviewers should verify summaries, revise AI-assisted drafts, and accept responsibility for the final review.	Review reports may contain inaccuracies and lack intellectual ownership or accountability.
Journal monitoring	Journal editors should regularly review AI-supported decisions and update policies where necessary.	Gradual changes in scoring patterns, editorial decisions, or fairness may go unnoticed. AI-assisted review has already been associated with measurable changes in scoring behaviour and acceptance outcomes, indicating the need for regular monitoring and calibration (Russo et al., 2025).

Stage 1 (Preliminary AI-Supported Manuscript Screening): Editorial review begins with AI-assisted screening of newly submitted manuscripts. The system checks for plagiarism, duplicate publication, journal formatting requirements, reporting checklists, and consistency with the journal's scope before peer review begins (Cox et al., 2025; Kousha & Thelwall, 2024; Regmi & Vaishya, 2025). These checks identify manuscripts that require editorial attention rather than assess their scientific merit. The handling editor should therefore verify every issue identified by AI before deciding whether the manuscript should proceed to external review or be declined at the editorial stage.

Stage 2 (Expert Review Supported by AI): After a manuscript is sent for peer review, reviewers are responsible for assessing originality, methodological rigour, interpretive validity, and scholarly contribution (Farber, 2025; Kankanhalli, 2024). At this stage, AI may assist by summarising lengthy sections, identifying structural inconsistencies, improving language clarity, and drawing attention to areas that require closer examination. However, reviewers remain responsible for the content of their reports and recommendations.

Stage 3 (Verification of AI-Assisted Observations): Human verification is the main safeguard against automation bias. Before AI-generated observations are used, reviewers or editors should confirm that flagged inconsistencies, suggested revisions, and summaries accurately represent the manuscript (Cox et al., 2025; Drori & Te'eni, 2024). Every AI-assisted comment included in a review should be supported by evidence from the manuscript. This means that recommendations are based on verified information rather than AI-generated output.

Stage 4 (Editorial Review and Decision-Making): Editors make publication decisions after considering the reviewers' reports rather than AI-generated recommendations. They consider the available evidence, resolve differences in reviewers' opinions, and determine whether the manuscript meets the journal's scientific, ethical, and editorial standards (Salih et al., 2025; Seghier, 2025). AI may assist in organising reviewer comments or identifying recurring issues. However, editors retain responsibility for every publication decision.

Stage 5 (Monitoring and Periodic Review): Responsible use of AI continues after editorial decisions have been made. Journal editors should periodically review peer-review duration, reviewer agreement, reporting deficiencies, score distributions, and fairness across different groups of authors to identify unintended changes in editorial practice (Farber, 2024; Russo et al., 2025). Where such changes are identified, editorial procedures and policies should be revised to maintain consistency, fairness, and confidence in peer review.

GOVERNANCE, IMPLEMENTATION, AND RESIDUAL RISKS

A central principle in the preceding sections is that AI should support rather than replace human judgement. Putting this principle into practice requires more than dividing tasks between AI and human reviewers. Journal editors also need clear policies on where AI may be used, how AI-generated outputs should be checked, and who is responsible for editorial decisions. Such policies can improve efficiency without weakening transparency, fairness, confidentiality, or trust. Without clear guidance, editorial practice may gradually change in ways that are not immediately recognised as AI use becomes more common.

Transparency, Disclosure, and Confidentiality

Responsible use of AI begins with transparency. Whenever AI assists manuscript preparation, peer review, or editorial assessment, its role should be made clear. Authors, reviewers, editors, and readers need to know where AI was used and where human judgement was applied. Without this information, uncertainty may arise about the source of reviewer comments, the extent of human intellectual input, and the reliability of editorial

decisions (Cox et al., 2025; Ebadi et al., 2025). Journal editors should therefore treat disclosure as part of editorial accountability rather than as an obstacle to innovation. This can be done by requiring brief statements explaining where AI was used and how its outputs were verified. Such statements can also help journal editors monitor changes in reviewer behaviour, scoring patterns, and editorial consistency over time.

Journal editors should also protect confidential information when AI is used in editorial practice. Manuscripts often contain unpublished findings, proprietary methods, sensitive datasets, or commercially valuable information. Uploading such materials to external AI platforms without appropriate safeguards may conflict with journal policies and editorial responsibilities (Salih et al., 2025; Seghier, 2025). Journal editors should therefore identify approved AI platforms, establish rules for data storage and retention, and specify acceptable procedures for handling manuscript information. Reviewers also need guidance on when external AI services may or may not be used, particularly where publicly available systems differ from institutionally approved platforms (Ebadi et al., 2025). Careful management of manuscript information is therefore as important as careful evaluation of AI-generated outputs.

Fairness, Accountability, and Human Oversight

Journal editors should treat fairness as a continuing responsibility throughout AI-assisted peer review. AI systems trained on historical publication records may reproduce existing inequalities associated with institutional prestige, language, geography, or disciplinary visibility unless journal editors take deliberate steps to reduce these risks (Checco et al., 2021; Farber, 2024). Regular reviews of reviewer recommendations, screening outcomes, and editorial decisions can identify emerging disparities before they become established. AI may also assist editors in identifying qualified reviewers beyond their usual professional networks, including early-career researchers and scholars from underrepresented regions. Whether journals realise this opportunity depends largely on editorial policy rather than on AI itself.

Researchers have consistently argued that AI may organise information, identify inconsistencies, and summarise complex material, but reviewers and editors must evaluate originality, assess methodological strengths and weaknesses, resolve differences in reviewer opinion, and make publication decisions (Drori & Te'eni, 2024; Seghier, 2025). Reviewers and editors protect the quality of peer review by maintaining this distinction. AI-generated observations may be fluent and convincing, but convincing language should not be mistaken for sound judgement. Reviewers and editors should therefore examine AI-assisted outputs with the same critical scrutiny applied to other sources of evidence.

Monitoring and Reviewer Competence

Journal editors should regularly assess whether AI-assisted review continues to operate as intended after it becomes part of routine editorial practice. Small changes in reviewer behaviour, scoring patterns, or publication decisions may develop gradually and pass unnoticed unless they are monitored regularly (Russo et al., 2025; Seghier, 2025). Where feasible, monitoring should also distinguish agreement between AI and individual reviewers from agreement with independent quality criteria, while documenting the prompts and assessment frameworks used to generate AI reviews. Journal editors should therefore review indicators such as review duration, reviewer agreement, score distributions, acceptance rates, and fairness across different groups of authors. Where important changes are identified, editorial policies and review procedures should be revised to maintain consistency and fairness. Journal editors should also ensure that reviewers and editors receive appropriate training. They should understand when AI outputs are useful, when they require careful checking, and when they should be disregarded. Training should therefore extend beyond learning how to use AI tools. Reviewers and editors should learn how to recognise common limitations, verify AI-generated outputs, understand the situations in which AI performs well or poorly, and use AI within accepted academic and ethical standards (Drori & Te'eni, 2024; Owan, 2026). Researchers have also reported that AI literacy, perceived

usefulness, and user competence influence whether AI is adopted responsibly and used critically rather than accepted without question (Drori & Te'eni, 2024; Owan et al., 2025). Journal editors should therefore support continuing professional development so that reviewers and editors can continue to exercise independent scholarly judgement when using AI.

Sustainability of Peer Review

Journal editors should also consider the long-term sustainability of peer review when introducing AI into editorial practice. Although AI may reduce routine tasks, greater efficiency does not necessarily reduce reviewer workload. Faster editorial processes may encourage journals to process more submissions without increasing reviewer capacity, thereby redistributing work rather than reducing it (Seghier, 2025). Researchers have expressed similar concerns in discussions of publication persistence, invisible academic labour, and the growing burden of scholarly service within universities (Owan & Ogabor, 2026). Sustained reviewer participation therefore requires greater recognition of peer review as skilled academic work. Formal acknowledgement of reviewer contributions, verified reviewing records, and recognition of editorial service within academic evaluation systems remain important ways of encouraging long-term participation (Carobene et al., 2024). Journal editors should use AI to reduce routine work so that reviewers can devote more time to scholarly evaluation rather than face increasing demands on their time.

Residual System Risks

Journal editors cannot eliminate every risk associated with AI-assisted peer review through governance alone. As AI technologies and editorial practices continue to develop, new challenges are likely to arise. Authors may increasingly adapt manuscripts to satisfy automated screening systems rather than improve scientific quality, while reviewers may rely too heavily on AI-generated comments that resemble expert judgement without equivalent intellectual scrutiny (Ofem et al., 2025; Seghier, 2025). Journal editors should also consider the risks associated with dependence on proprietary AI systems whose internal processes are difficult to examine (Salih et al., 2025). Unequal access to AI technologies may widen differences between reviewers and institutions (Ebadi et al., 2025). Extensive use of AI-assisted language editing may also reduce the variety of writing styles found in reviewer reports. Journal editors should therefore review AI policies regularly because additional challenges are likely to arise as AI technologies and scholarly publishing continue to develop.

CONCLUSION

Peer review remains central to maintaining scholarly quality, but it faces increasing pressure from reviewer shortages, inconsistent judgements, long review times, overlooked errors, and bias. AI can assist with tasks that reviewers may find difficult to perform consistently under limited time and heavy workload, particularly manuscript screening, reviewer identification, and consistency checks. Although some researchers have proposed highly automated or autonomous peer-review systems (Lin et al., 2021; Lu et al., 2026), current evidence supports AI as a supervised assistant rather than an independent reviewer (Joachim et al., 2025; Kousha & Thelwall, 2024). Important limitations include fabricated outputs, weak assessment of originality, reproduction of existing biases, limited transparency, confidentiality concerns, and excessive reliance on AI that may weaken reviewer expertise. These concerns support a Human–AI Augmented Review model in which AI assists reviewers and editors while responsibility for interpretation, ethical judgement, and publication decisions remains with humans. The framework establishes a division of labour, incorporates verification throughout the review process, and sets governance measures for accountability, fairness, and trust. AI can therefore reduce routine workload without transferring evaluative responsibility to the technology. Treating AI as a substitute for reviewers may increase speed but weaken reliability and accountability. Successful implementation depends on careful governance, systematic verification, and continued human judgement throughout the peer-review process (Owan, 2026).

Author Contributions: VJO: Conceptualisation, development of the Human–AI Augmented Review model, literature synthesis, development of the review framework, writing of the original draft, and project administration; MVO: Literature review, synthesis of relevant literature, critical revision of the manuscript, and correspondence; EEA: Critical analysis of the literature, examination of governance and ethical implications, and critical revision of the manuscript. All authors reviewed and approved the final version of the manuscript.

Funding: No funding was received to assist with the preparation of this manuscript.

Declaration of interest: The authors declare no competing or conflicting interests.

Ethical statement: This article is a conceptual review based on existing literature. It did not involve human participants, animal subjects, or the collection of primary data. Ethical approval was therefore not required.

AI statement: The authors used ChatGPT to assist with language refinement and improvement of fluency and clarity. The authors reviewed and verified the final manuscript and take full responsibility for the content of the published work.

Data sharing statement: No original data were collected or generated for this review article. Data sharing is therefore not applicable.

REFERENCES

- Akpan, S. M., & Owan, V. J. (2026). Self-directed research learning as a mediator between artificial intelligence utilisation and research productivity: A hierarchical regression approach. *Education for Today*, 22(4), 268–278. <https://philpapers.org/archive/AKPSRL.pdf>
- Almukhtar, A., Caddick, V., Naik, R., Goble, M., Mylonas, G., Darzi, A., Orihuela-Espina, F., & Leff, D. R. (2025). Objective assessment of cognitive workload in surgery: A systematic review. *Annals of Surgery*, 281(6), 942–951. <https://doi.org/10.1097/SLA.0000000000006370>
- Carabantes, D., González-Geraldo, J. L., & Jover, G. (2023). ChatGPT could be the reviewer of your next scientific paper: Evidence on the limits of AI-assisted academic reviews. *Profesional de la Información*, 32(5), Article e320516. <https://doi.org/10.3145/epi.2023.sep.16>
- 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. *Clinical Chemistry and Laboratory Medicine*, 62(5), 835–843. <https://doi.org/10.1515/cclm-2023-1136>
- Checco, A., Bracciale, L., Loreti, P., Pinfield, S., & Bianchi, G. (2021). AI-assisted peer review. *Humanities and Social Sciences Communications*, 8, Article 25. <https://doi.org/10.1057/s41599-020-00703-8>
- Chong, S. W., & Lin, T. (2024). Feedback practices in journal peer-review: A systematic literature review. *Assessment & Evaluation in Higher Education*, 49(1), 1–12. <https://doi.org/10.1080/02602938.2022.2164757>
- Cox, L. A., Jr., Aven, T., Guikema, S., Haas, C. N., Lambert, J. H., Lowrie, K., Maldonado, G., & Wu, F. (2025). Can AI help authors prepare better risk science manuscripts? *Risk Analysis*, 45, 2471–2493. <https://doi.org/10.1111/risa.70055>
- Drori, I., & Te'eni, D. (2024). Human-in-the-loop AI reviewing: Feasibility, opportunities, and risks. *Journal of the Association for Information Systems*, 25(1), 98–109. <https://doi.org/10.17705/1jais.00867>
- Ebadi, S., Nejadghanbar, H., Salman, A. R., & Khosravi, H. (2025). Exploring the impact of generative AI on peer review: Insights from journal reviewers. *Journal of Academic Ethics*, 23, 1383–1397. <https://doi.org/10.1007/s10805-025-09604-4>
- Farber, S. (2024). Enhancing peer review efficiency: A mixed-methods analysis of artificial intelligence-assisted reviewer selection across academic disciplines. *Learned Publishing*, 37(4), Article e1638. <https://doi.org/10.1002/leap.1638>
- Farber, S. (2025). Comparing human and AI expertise in the academic peer review process: Towards a hybrid approach. *Higher Education Research & Development*, 44(4), 871–885. <https://doi.org/10.1080/07294360.2024.2445575>
- García-Costa, D., Squazzoni, F., Mehmani, B., & Grimaldo, F. (2022). Measuring the developmental function of peer review: A multi-dimensional, cross-disciplinary analysis of peer review reports from 740 academic journals. *PeerJ*, 10, Article e13539. <https://doi.org/10.7717/peerj.13539>

- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23(4), 523–552. <https://doi.org/10.1007/s10648-011-9174-7>
- Ghazanfar, M. A., Cook, M., Tang, B., Tait, I., & Alijani, A. (2015). The effect of divided attention on novices and experts in laparoscopic task performance. *Surgical Endoscopy*, 29(3), 614–619. <https://doi.org/10.1007/s00464-014-3708-2>
- Joachim, M. V., Dodson, T. B., & Laviv, A. (2025). How artificial intelligence differs from humans in peer review. *Journal of Oral and Maxillofacial Surgery*, 83(8), 1040–1050. <https://doi.org/10.1016/j.joms.2025.03.015>
- Kadaifci, C., Isikli, E., & Topcu, Y. I. (2025). Fundamental problems in the peer-review process and stakeholders' perceptions of potential suggestions for improvement. *Learned Publishing*, 38(1), Article e1637. <https://doi.org/10.1002/leap.1637>
- Kankanhalli, A. (2024). Peer review in the age of generative AI. *Journal of the Association for Information Systems*, 25(1), 76–84. <https://doi.org/10.17705/1jais.00865>
- 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>
- Kovanis, M., Porcher, R., Ravaut, P., & Trinquart, L. (2016). The global burden of journal peer review in the biomedical literature: Strong imbalance in the collective enterprise. *PLoS ONE*, 11(11), Article e0166387. <https://doi.org/10.1371/journal.pone.0166387>
- Kulal, A., N, A., Shareena, P., & Dinesh, S. (2025). Unmasking favoritism and bias in academic publishing: An empirical study on editorial practices. *Public Integrity*, 27(6), 714–735. <https://doi.org/10.1080/10999922.2024.2448875>
- Lin, J., Song, J., Zhou, Z., Chen, Y., & Shi, X. (2023). Automated scholarly paper review: Concepts, technologies, and challenges. *Information Fusion*, 98, Article 101830. <https://doi.org/10.1016/j.inffus.2023.101830>
- Lu, C., Lu, C., Lange, R. T., Yamada, Y., Hu, S., Foerster, J., Ha, D., & Clune, J. (2026). Towards end-to-end automation of AI research. *Nature*, 651, 914–919. <https://doi.org/10.1038/s41586-026-10265-5>
- Mann, P. S., 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*, abs/2509.14189. <https://doi.org/10.48550/arxiv.2509.14189>
- Meyerson, L. A., Suzzi-Simmons, A., & Simberloff, D. (2025). Quantifying reviewer declines in scientific publishing: Twenty-one years of data from *Biological Invasions* 2002–2024. *Biological Invasions*, 27, Article 223. <https://doi.org/10.1007/s10530-025-03679-1>
- Morley, C. P., Prunuske, J., Phillips, J. P., Wendling, A. L., Heidelbaugh, J. J., & Grammer, S. (2025). Reviewer engagement trends at a journal: Cause for concern. *PRiMER*, 9, Article 59. <https://doi.org/10.22454/PRiMER.2025.831615>
- Ngo, V. M. (2025). Human–AI collaboration in high-stakes decisions: A meta-analysis of healthcare and public sectors. *Applied Economics Letters*, 1–6. <https://doi.org/10.1080/13504851.2025.2586160>
- Ofem, U. J., Owan, V. J., Iyam, M. A., Udeh, M. I., Anake, P. M., & Ovat, S. V. (2025). Students' perceptions, attitudes and utilisation of ChatGPT for academic dishonesty: Multigroup analyses via PLS–SEM. *Education and Information Technologies*, 30(1), 159–187. <https://doi.org/10.1007/s10639-024-12850-5>
- Owan, V. J. (2026). “We talk about the risks, what about the gains?” Large language models and learning in mathematics, statistics, and computing. *Journal of Education, Innovation and Communication*, 8(2), 62–88. <https://doi.org/10.34097/jeicom-8-2-5>
- Owan, V. J., & Asuquo, M. E. (2022). “Publish or perish,” “publish and perish”: The Nigerian experience. In J. A. Undie, J. B. Babalola, B. A. Bello, & I. N. Nwankwo (Eds.), *Management of higher educational systems* (pp. 986–994). University of Calabar Press. <https://philarchive.org/rec/OWAQOP>
- Owan, V. J., & Ogabor, J. O. (2026). Publish or perish revisited: Persistence, academic labour, and institutional logics in scholarly publishing. *Aminu Kano Academic Scholars Association Multidisciplinary Journal*, 4(4), 1–24. <https://philarchive.org/rec/OWAPOP>
- Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(8), Article em2307. <https://doi.org/10.29333/ejmsste/13428>
- Owan, V. J., Chukwu, C. O., Agama, V. U., Owan, T. J., Ogar, J. O., & Etorti, I. J. (2025). Acceptance and use of artificial intelligence for self-directed research learning among postgraduate students in Nigerian public universities. *Discover Education*, 4(1), Article 329. <https://doi.org/10.1007/s44217-025-00770-6>
- Paliya, N., & Jha, S. S. (2024). Quality indicators for academic journals: What makes a journal trustworthy? *IP Indian Journal of Anatomy and Surgery of Head, Neck and Brain*, 10(4), 83–91. <https://doi.org/10.18231/j.ijashnb.2024.019>

- Regmi, A., & Vaishya, R. (2025). Artificial intelligence in scholarly publishing: Enhancing editorial efficiency while preserving human expertise. *National Board of Examinations Journal of Medical Sciences*, 3(10), 1288–1297. <https://doi.org/10.61770/NBEJMS.2025.v03.i10.012>
- Russo, G., Horta Ribeiro, M., Davidson, T. R., Veselovsky, V., & West, R. (2025). The AI review lottery: Widespread AI-assisted peer reviews boost paper scores and acceptance rates. *Proceedings of the ACM on Human-Computer Interaction*, 9(7), Article CSCW486. <https://doi.org/10.1145/3757667>
- Saad, A., Jenko, N., Ariyaratne, S., Birch, N., Iyengar, K. P., Davies, A. M., Vaishya, R., & Botchu, R. (2024). Exploring the potential of ChatGPT in the peer review process: An observational study. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 18(2), Article 102946. <https://doi.org/10.1016/j.dsx.2024.102946>
- Salih, M. H. B., Gul, T., Bukhari, S. S., Liaqat, I., & Majeed, A. (2025). AI-enhanced scholarly communication: Transforming peer review, knowledge dissemination, and academic publishing workflows in the digital era. *International Journal of Ethical AI Application*, 1(4), 13–19. <https://doi.org/10.64229/4qp06e84>
- Seghier, M. L. (2025). AI-powered peer review needs human supervision. *Journal of Information, Communication and Ethics in Society*, 23(1), 104–116. <https://doi.org/10.1108/JICES-09-2024-0132>
- Severin, A., & Chataway, J. (2021). Overburdening of peer reviewers: A multi-stakeholder perspective on causes and effects. *Learned Publishing*, 34(4), 537–546. <https://doi.org/10.1002/leap.1392>
- Sizo, A., Lino, A., Rocha, Á., & Reis, L. P. (2025). Defining quality in peer review reports: A scoping review. *Knowledge and Information Systems*, 67, 6413–6460. <https://doi.org/10.1007/s10115-025-02435-0>
- Zancanaro, E., Giannopoulos, A., Gimelli, A., & Kresoja, K.-P. (2026). AI as a peer reviewer: A blinded comparative study of LLM-generated and human reviews in a cardiology journal. *European Heart Journal - Imaging Methods and Practice*, 4(1), Article qyag097. <https://doi.org/10.1093/ehjimp/qyag097>