Jurnal Penelitian dan Pengkajian Ilmu Pendidikan: e-Saintika
https://journal-center.litpam.com/index.php/e-Saintika/index

July 2026 Vol. 10, No. 2
e-ISSN: 2615-6881
pp. 758-784

Artificial Intelligence in Scientific Publishing: A Narrative Review of Writing, Peer Review, Ethics, Equity, and Governance

Lalu Nurul Yaqin

Malay Language and Linguistics and Literature Program, Faculty of Arts and Social Sciences, University Brunei Darussalam, Gadong, Brunei Darussalam

Corresponding Author e-mail: lalu.yaqin@ubd.edu.bn

Received: February 2026; Revised: June 2026; Published: July 2026

Abstract

Artificial intelligence (AI) is increasingly reshaping scientific publishing by supporting manuscript preparation, language editing, editorial screening, and peer review while introducing new ethical, methodological, and governance challenges. This narrative review examines the emerging role of large language models (LLMs), including ChatGPT, across scientific writing and editing, editorial workflows, disciplinary adoption, research integrity, publishing equity, and institutional policy. A structured search of the Scopus database identified 152 records, of which 22 studies met the eligibility criteria and formed the primary synthesis. The reviewed evidence indicates that AI can improve readability, linguistic clarity, translation, and the efficiency of structured editorial tasks; however, these benefits do not consistently translate into greater technical accuracy, analytical depth, or scholarly judgement. Adoption also varies across disciplines and publishing contexts. Major concerns include citation hallucination, factual inaccuracies, authorship ambiguity, bias, confidentiality, inadequate disclosure, and over-reliance on automated outputs. At the same time, AI-assisted language and translation tools may reduce barriers faced by non-native English-speaking and under-resourced researchers, although unequal access to advanced systems may reinforce existing disparities. Current journal and institutional policies increasingly converge on disclosure, prohibition of AI authorship, reference verification, and human accountability, although implementation remains inconsistent. Overall, the evidence supports the use of AI as an assistive rather than autonomous scholarly tool. Responsible integration requires transparent disclosure, rigorous verification, meaningful human oversight, equitable access, and adaptive governance.

Keywords: Artificial Intelligence; Large Language Models; Scientific Publishing; Research Integrity; Peer Review; Publishing Ethics

How to Cite: Yaqin, L. N. (2026). Artificial Intelligence in Scientific Publishing: A Narrative Review of Writing, Peer Review, Ethics, Equity, and Governance. *Jurnal Penelitian Dan Pengkajian Ilmu Pendidikan: E-Saintika, 10*(2), 758-784. https://doi.org/10.36312/e-saintika.v10i2.4583

10.36312/e-saintika.v10i2.4583

INTRODUCTION

The advent of powerful large language models (LLMs) has sparked profound discussions about the future of scientific writing and publishing. The public release of ChatGPT in late 2022, a conversational AI developed by OpenAI, marked a turning point. Its rapid uptake introduced new possibilities for AI-assisted scholarly writing while simultaneously raising questions about authorship, accountability, and the integrity of scientific communication. Soon after its release, researchers began experimenting with AI-assisted drafting in scholarly communication. In one early case, authors even credited ChatGPT as a co-author on a journal article, only to have that decision reversed after widespread debate in media and editorial circles (Hill-Yardin et al., 2023). This early controversy illustrated a broader challenge that

continues to shape the use of generative AI in academia: technological capability has advanced more rapidly than established norms governing its responsible use.

Generative AI offers substantial practical benefits for manuscript preparation, including rapid drafting, language polishing, and literature summarisation. For example, text produced by ChatGPT can be "fairly accurate, logical, [and] grammatically correct," providing a coherent introduction on complex topics within minutes when used as a drafting aid (Hill-Yardin et al., 2023). However, linguistic fluency does not necessarily correspond to scientific depth or reliability. Early adopters noted that AI-written prose, while fluent, often lacks substance and originality; it tends to be "largely shallow, bland, dry, and generic, lacking a distinct 'voice'" (Hill-Yardin et al., 2023). More troublingly, the same AI output that reads smoothly can harbour hidden flaws, such as fabricated citations that appear plausible but are entirely made up (Hill-Yardin et al., 2023). These contrasting capabilities reveal a central tension in AI-assisted scientific publishing: tools that improve efficiency and readability may simultaneously introduce risks to accuracy, intellectual depth, transparency, and scholarly accountability.

Since 2023, a growing body of research has explored AI's impact on scholarly communication. However, this rapidly expanding literature addresses multiple dimensions of AI-assisted publishing that are often considered separately, including manuscript preparation, editorial work, peer review, disciplinary adoption, research integrity, and publishing equity. An integrated assessment is therefore needed to clarify where AI provides demonstrable support, where its limitations remain consequential, and how these benefits and risks vary across scholarly contexts. This narrative review draws on recent empirical studies and relevant scholarly literature to synthesise these interconnected developments. We examine AI's use in scientific writing, editing, and abstract generation; its integration into editorial workflows and peer review; and cross-disciplinary differences in uptake. We also examine the complex ethical terrain—including issues of transparency, "citation hallucination," authorship credit, and bias—and their implications for scientific integrity and global publishing equity, particularly for non-native English-speaking and under-resourced researchers.

Beyond the technological capabilities of AI, its integration into scientific publishing is increasingly shaped by editorial and institutional governance. The review therefore considers evolving policy responses by journals and institutions aimed at governing AI use, including field-specific debates in health education and clinical publishing (O'Connor, 2023) and early case-based reports that explicitly documented ChatGPT assistance in manuscript preparation (Manohar & Prasad, 2023). These developments are important because decisions about disclosure, authorship, confidentiality, and acceptable forms of AI assistance increasingly determine how AI can be incorporated into scholarly workflows without weakening human responsibility.

Accordingly, this review aims to synthesise evidence on AI-assisted scientific writing, editorial practice, peer review, disciplinary adoption, ethics, equity, and governance within a unified analytical framework. Specifically, it examines how generative AI is reshaping scientific publishing, identifies the benefits and limitations reported across recent studies, and evaluates the safeguards required for its responsible integration. Two summary tables distil key findings and policy variations to date. By connecting evidence on technological performance with ethical,

disciplinary, and equity considerations, the review seeks to move beyond a simple benefits-versus-risks debate. Instead, it provides a balanced account of the conditions under which AI may enhance efficiency and inclusivity while preserving scientific rigour, transparency, accountability, and trust.

METHOD

Review Design and Scope

This article is a narrative review of recent literature on the use of artificial intelligence (AI), particularly large language models (LLMs), in scientific publishing. The review aimed to synthesise current evidence across five interconnected areas: (1) AI in scientific writing and editing, (2) AI in editorial work and peer review, (3) cross-disciplinary differences in adoption, (4) ethical and transparency challenges, and (5) implications for non-native English-speaking researchers and global publishing equity. A narrative review approach was adopted to enable the integration of empirical findings, policy-oriented analyses, and scholarly perspectives across a rapidly evolving and multidisciplinary field.

Search Strategy

To identify relevant literature, the authors conducted a structured Boolean search in the Scopus database. The search combined AI-related terms ("artificial intelligence," "large language models," and "ChatGPT") with terms related to scholarly communication and publishing ("scientific writing," "academic publishing," "peer review," "editorial policy," "citation hallucination," "authorship," and "ethic*"). The search focused on publications addressing AI-assisted scholarly communication. Publication-year eligibility was subsequently applied during screening to prioritise literature published from 2023 onward, reflecting the rapid expansion of research following the public release of ChatGPT in late 2022.

The search identified 152 records in Scopus. Before screening, no records were removed as duplicates (n = 0), through automation tools (n = 0), or for other reasons (n = 0). Accordingly, all 152 identified records proceeded to screening.

Screening and Eligibility Criteria

During screening, 15 records published before 2023 were excluded, leaving 137 reports sought for retrieval. All 137 reports were successfully retrieved (reports not retrieved, n = 0) and proceeded to eligibility assessment. The complete identification, screening, eligibility, and inclusion process is summarised in Figure 1.

Eligibility was further restricted to journal articles published in English. After these limits were applied, 137 English-language articles remained available for detailed assessment. The review considered peer-reviewed journal articles, empirical studies, policy-focused analyses, and relevant scholarly commentaries addressing AI in manuscript preparation, editorial workflows, peer review, publishing ethics, or academic equity.

Studies were considered eligible when they provided direct empirical findings, substantive scholarly discussion, or policy analysis relevant to the role of AI in scientific publishing. Sources were excluded when they focused primarily on general AI development without clear relevance to scholarly publishing, duplicated findings already represented in the selected literature, or lacked sufficient relevance to the review objectives. During full-text eligibility assessment, 115 reports that did not

adequately address the predefined thematic scope were classified as out of scope and excluded.

Following full-text assessment, 22 studies met the eligibility criteria and constituted the primary evidence base of the review. The number of included reports therefore also totalled 22. These studies formed the basis of the main results presented in the review and are subsequently summarised according to their AI application, disciplinary context, and principal findings.

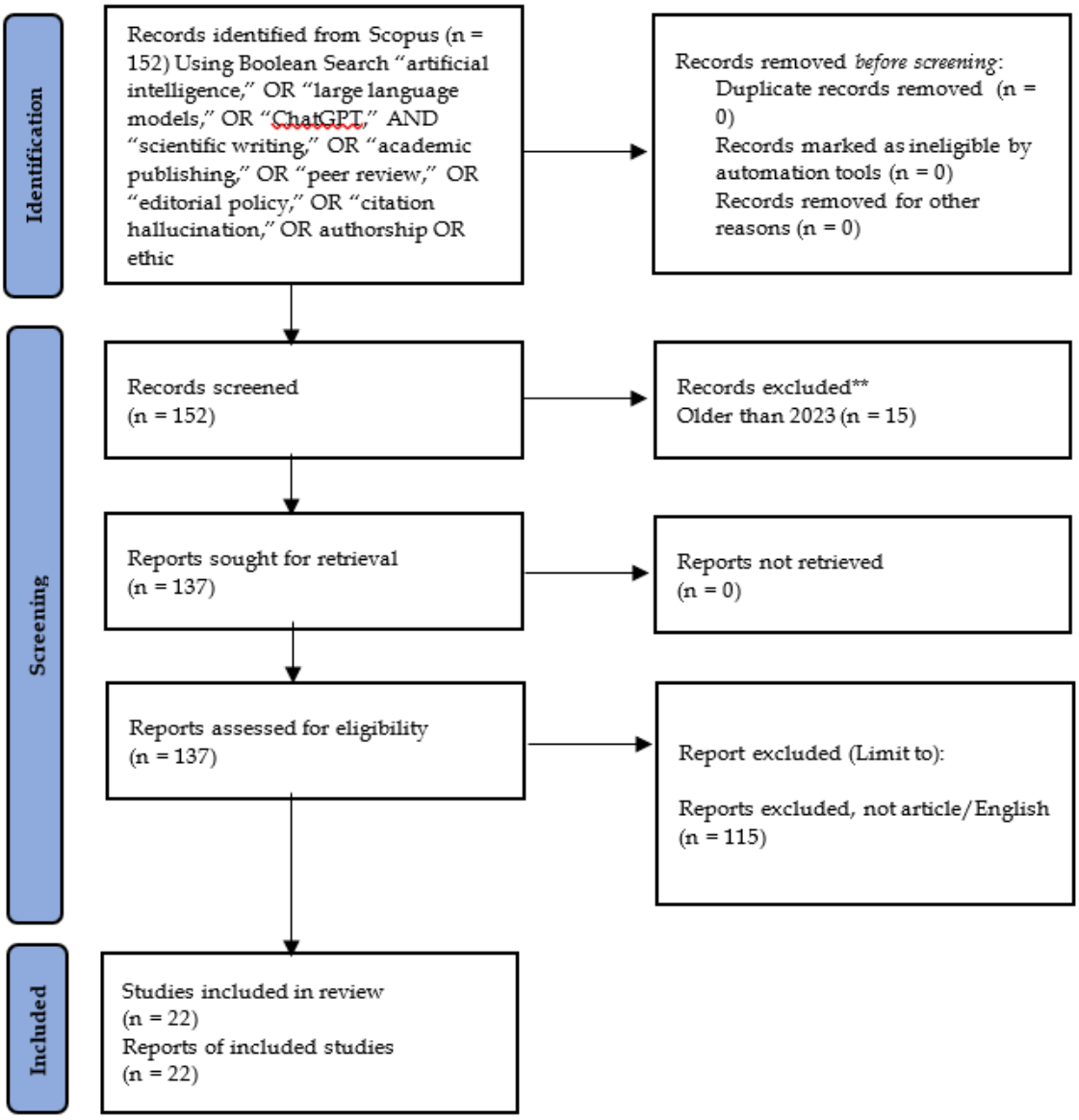

Figure 1. PRISMA-style flow diagram of the database search, screening, eligibility assessment, and study inclusion process.

Data Organisation and Narrative Synthesis

The 22 included studies were read in full and organised thematically. Rather than statistically pooling heterogeneous outcomes, the review employed comparative reading and narrative synthesis to identify recurrent patterns across the evidence base. Particular attention was given to reported benefits of AI-assisted scholarly work, limitations in technical and analytical performance, ethical and transparency risks, disciplinary variation, implications for publishing equity, and evolving policy responses.

The synthesis proceeded by comparing findings both within and across thematic domains to identify areas of convergence, divergence, and emerging practice. The resulting evidence was organised around the major analytical themes of the review, including scientific writing and editing, editorial work and peer review, cross-disciplinary adoption, ethical and transparency challenges, and implications for non-native English-speaking researchers and global equity. The principal characteristics and findings of the included studies are presented in Table 1, whereas broader policy responses discussed in relation to the findings are synthesised separately in Table 2.

Methodological Scope and Considerations

Because this study is a narrative review rather than a systematic review, it does not aim to provide exhaustive coverage of all publications on the topic or to conduct a quantitative meta-analysis. Instead, it offers a critical and integrative synthesis of recent and influential literature to capture emerging trends, areas of convergence and disagreement, and evolving debates in a rapidly developing field.

This approach is particularly appropriate for AI in scientific publishing because the available evidence spans heterogeneous disciplines, research designs, applications, and normative questions that cannot readily be reduced to a single quantitative outcome. Accordingly, the review is intended to provide a transparent and focused synthesis of the available evidence while preserving contextual differences among studies and highlighting issues that require continued empirical and policy attention.

RESULTS AND DISCUSSION

Characteristics of the Included Studies

A total of 22 studies met the eligibility criteria and constituted the primary evidence base of this review. The included studies were published between 2023 and 2026 and represented a broad range of disciplinary and publishing contexts, including medicine, biomedical publishing, pharmacy education, accounting, information science, orthopaedics, surgery, radiology, neuroscience, and cross-disciplinary scholarly communication. The studies examined AI applications across manuscript preparation, translation, abstract generation, peer review, editorial screening, AI-text detection, citation integrity, authorship, disclosure practices, and journal policy.

As summarised in Table 1, the evidence base comprised both direct evaluations of AI performance and observational or bibliometric analyses of how AI is being adopted and governed in scientific publishing. Across the included studies, four major patterns emerged: (1) AI frequently improved linguistic clarity and structured writing tasks but showed limitations in technical and analytical depth; (2) AI demonstrated potential for supporting editorial screening and peer review, although performance varied by task and model; (3) AI-assisted scholarly writing and disclosure practices differed substantially across disciplines and journals; and (4) concerns regarding fabricated references, AI-generated text detection, authorship, transparency, and human accountability remained prominent.

Figure 2 provides a conceptual evidence landscape of these interconnected applications, showing how writing, evaluation, integrity, and governance functions collectively define the emerging role of AI in scientific publishing.

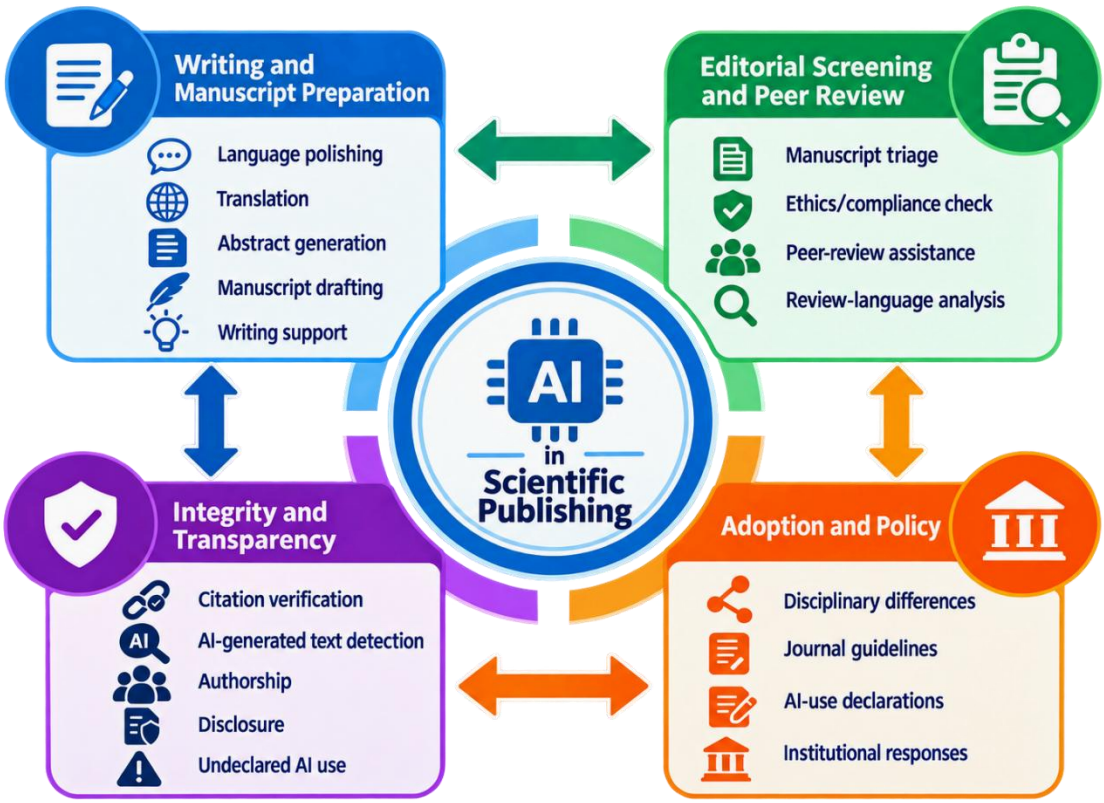

Figure 2. Conceptual landscape of artificial intelligence applications identified across the included studies in scientific publishing. The figure groups the evidence into four interconnected domains: AI-assisted writing and manuscript preparation, editorial screening and peer review, research integrity and transparency, and disciplinary adoption and publishing policy.

AI-Assisted Scientific Writing and Manuscript Preparation

Eight included studies directly examined AI-assisted writing, translation, abstract generation, manuscript development, or writing-related learning. Collectively, these studies showed that generative AI can improve readability, linguistic fluency, and structural organisation, while performance in technical accuracy, analytical depth, and reference reliability remains inconsistent.

de Albuquerque and Gomes Dos Santos (2026) asked ChatGPT-4.0 to rewrite abstracts from qualitative accounting papers and had experienced academics evaluate the outputs. AI-rewritten abstracts were frequently preferred for readability and perceived completeness, although they did not consistently outperform the original human-written abstracts in overall quality or trustworthiness. Similarly, Kim et al. (2025), comparing author-written and AI-generated abstracts from 120 conference papers, found that different prompt-engineering approaches produced limited semantic variation and that experts sometimes preferred AI-generated abstracts when the author-written abstracts did not adequately reflect the content of the papers.

Nathani et al. (2025) compared a full scientific manuscript written by an experienced author with a manuscript generated by ChatGPT-4.0 from the same data and outline. The AI-generated manuscript received higher scores for readability (9.0 vs. 7.2), but lower scores for technical accuracy (6.3 vs. 9.3) and analytical depth (5.5 vs. 7.5). Reviewers also identified limitations in critical analysis and were able to distinguish the AI-generated manuscript from the human-written version. Sheridan

et al. (2025) similarly evaluated AI-generated discussion and conclusion sections in orthopaedic publishing. Although AI-generated sections scored somewhat lower than human-written versions, five of six blinded reviewers considered them potentially acceptable for publication after major revision.

Kadi and Aslaner (2024) assessed ChatGPT for medical case-report writing and peer review. The system was able to generate complete case-report text, but the outputs contained reference inaccuracies and received mixed professional evaluations, with 33.3% of reviewers recommending rejection. Hill-Yardin et al. (2023) likewise reported that ChatGPT could rapidly generate text that was logical, grammatically correct, and accessible, while also identifying recurrent limitations including fabricated references, generic prose, limited source transparency, and insufficient high-level critical reasoning.

AI-assisted translation showed comparatively strong linguistic performance. Pinto et al. (2026) evaluated several generative AI translators, including ChatGPT-4 and DeepSeek, in oral and maxillofacial surgery. DeepSeek achieved the highest reported accuracy and clarity score (mean ≈92.9), while other systems also improved grammar and readability to varying degrees. These findings indicate substantial variation in translation performance across generative AI systems.

Writing-related adoption was also evident in educational settings. Iwasawa et al. (2026) compared pharmacy students' engagement with ChatGPT between 2023 and 2024. Following a shift from prohibition to conditional institutional use, the later cohort demonstrated greater AI awareness and adoption, although knowledge of appropriate use remained limited. Approximately 62% of participants supported the incorporation of formal AI education.

AI in Editorial Screening and Peer Review

Several included studies evaluated whether generative AI could support structured editorial and peer-review tasks. The results generally showed stronger performance for checklist-based, structured, or compliance-oriented assessments than for tasks requiring complex evaluative judgement.

Doskaliuk et al. (2026) assessed ChatGPT-4.0 as an automated tool for identifying missing ethics and transparency statements in biomedical publications. The model demonstrated high sensitivity and recall for omissions involving institutional review board approval, informed consent, and data-sharing statements. However, precision varied across disclosure categories and was particularly low for funding statements (precision = 0.16). The study also reported a higher frequency of missing disclosures in Q4 journals than in Q1 journals.

George et al. (2025) compared two AI chatbots with human reviewers using a STROBE-based completeness metric. One AI system, Claude, produced the most complete reviews, achieving approximately 76% completeness compared with approximately 45% for human reviewers. Performance nevertheless varied between AI systems and according to manuscript quality.

Comparable variability was observed in other peer-review studies. Marrella et al. (2025) examined 11 published surgical articles and found that ChatGPT-generated reviews achieved higher ARCADIA quality scores (4.8–4.9) than human reviews (2.8–3.2), although agreement with decisions made by high-impact journals remained limited at approximately 29%–32%. Saad et al. (2024), evaluating 21 peer-reviewed articles, found moderate similarity between ChatGPT-generated and human peer

reviews, with mean similarity scores of 3.6/5 for ChatGPT-3.5 and 3.76/5 for ChatGPT-4.0.

Kadi and Aslaner (2024) additionally found that ChatGPT's peer-review performance was weaker than its manuscript-generation performance and that the system failed to identify some major inconsistencies. Verharen (2023), using more than 500 peer-review reports from 200 neuroscience papers, demonstrated a different application of AI: ChatGPT was able to analyse reviewer language and identify substantial variability in scoring as well as patterns suggestive of gender-related differences in review tone and favourability.

Taken descriptively, these studies indicate that AI performance in editorial and peer-review applications is task dependent. Higher performance was generally reported for structured completeness assessment and pattern identification, whereas results were less consistent for evaluative decisions and detection of substantive manuscript problems.

Citation Integrity, AI-Generated Text Detection, and Authorship

A further group of included studies examined integrity-related consequences of AI-assisted scholarly communication, including reference reliability, detection of AI-generated text, undeclared AI assistance, and AI authorship.

Yao et al. (2025) evaluated how different AI systems handled retracted literature. ChatGPT-4 retrieved approximately 80% of the retracted articles examined and correctly identified approximately 62% as retracted, while Grok 3 flagged 67% and DeepSeek identified very few. When models failed to retrieve the relevant articles, fabricated citations were generated in 26% of missing cases for ChatGPT, 63% for Grok, and 88% for DeepSeek.

Cheng et al. (2025) compared three AI-text detection tools with human judgement. All three automated detectors differentiated human from AI-generated text at a statistically significant level ($p < .001$), although agreement between tools varied substantially (ICC = 0.57–0.95). Human evaluators achieved approximately 19% accuracy, indicating poor performance in identifying AI-generated passages.

Evidence of AI-assisted writing in published literature was also reported. Callanan et al. (2025), analysing 3,374 orthopaedic manuscripts using ZeroGPT, estimated that 16.7% showed AI involvement above the study's defined threshold, with journal-level estimates ranging from 5.6% to 38.3%. Strzelecki (2025), using the SPAR4SLR approach, identified published papers across multiple disciplines that appeared to contain undeclared ChatGPT-generated content; several of these papers had already entered the citation record.

Authorship-related concerns were documented by Nazarovets and Teixeira da Silva (2025). Their bibliometric analysis identified 14 papers in which ChatGPT had been credited as an author, co-author, or group author despite existing authorship guidance. Together, these studies demonstrate that reference integrity, disclosure, detection, and authorship constitute measurable components of AI's emerging presence in the scholarly record.

Cross-Disciplinary Adoption, Disclosure, and Journal Policy

The included studies also revealed substantial variation in AI adoption and governance across disciplines, geographic contexts, and publishing functions.

Ziyang (2025) analysed 7,953 declarations of AI use across scientific fields and found that reported AI adoption increased from 62.6% in October 2023 to 78.2% in

March 2025. ChatGPT remained the most frequently declared tool, while patterns of use differed by discipline and by function, including writing and translation.

Raman (2025), examining 1,226 publications containing acknowledgements of ChatGPT use, found that AI-use acknowledgements were concentrated in the United States, China, and India and occurred particularly frequently in biomedical, clinical, and computing-related fields. The study identified acknowledgement statements as an emerging form of transparency in AI-assisted scholarship.

Journal policies likewise varied by field. Simsek et al. (2024), reviewing 112 MEDLINE-indexed imaging journals, reported that 69 journals (61.6%) had an AI-use policy and 40 specifically addressed AI-generated figures. Journals with formal AI policies also tended to have higher citation and impact metrics. Goyanes et al. (2025), analysing author guidelines across eight social science disciplines, found that 73.8% of leading journals addressed AI use in drafting, 61.3% required some form of transparency, and only 12% explicitly addressed privacy or data protection. Acceptance of AI-assisted practices also varied across disciplines, with communication and sociology showing greater openness than economics.

These findings show that AI adoption is not uniform across scholarly publishing. Differences were evident not only between disciplines but also across specific functions, including manuscript drafting, translation, disclosure, figure generation, and editorial policy.

Summary of the Primary Evidence

Across the 22 included studies, the results show a consistent distinction between linguistic or structured-task performance and higher-order scholarly performance. AI frequently improved readability, translation, organisation, checklist completeness, and other structured tasks, but results were more variable for technical accuracy, analytical depth, reference integrity, substantive peer-review judgement, and authorship-related accountability. At the same time, studies of adoption and policy demonstrated rapid but uneven integration of AI across disciplines, journals, and geographic contexts. These empirical patterns constitute the primary results of the review and provide the basis for the subsequent Discussion. Table 1 summarises all 22 included studies, including their AI use cases, disciplinary or publishing contexts, and principal reported outcomes.

Table 1. Overview of the 22 included studies on AI applications in scientific writing, reviewing, integrity, adoption, and publishing policy

Author (Year)	AI Use Case	Domain/Context	Key Findings & Outcomes
Pinto et al. (2026)	AI translation for non-native authors (ChatGPT-4, DeepSeek, etc.)	Medical writing (OMFS surgery)	DeepSeek achieved the best accuracy/clarity (mean ≈92.9), improving grammar and readability. Gemini ranked second but sometimes refused revisions, stating that the text was "already perfect."
de Albuquerque & Gomes	Abstract rewriting with ChatGPT-4	Business (Accounting)	Expert readers generally preferred AI-rewritten abstracts for readability and completeness. However, AI

Author (Year)	AI Use Case	Domain/Context	Key Findings & Outcomes
Dos Santos (2026)			versions did not consistently exceed human originals in perceived trustworthiness or overall quality.
Nathani et al. (2025)	Full manuscript writing: AI vs. human	Medical research	AI paper was more readable (9.0 vs. 7.2) but weaker in technical accuracy (6.3 vs. 9.3) and depth (5.5 vs. 7.5). Reviewers recognised AI output and noted limited critical analysis.
Iwasawa et al. (2026)	AI exposure in student writing and learning	Education (Pharmacy)	Policy shifted from banning to conditional use (2023–2024), increasing awareness and adoption. The 2024 cohort showed higher AI knowledge, yet appropriate-use literacy remained limited. About 62% supported formal AI education.
George et al. (2025)	AI vs. human performance in peer review	Public health (Publishing)	Using a STROBE-based completeness metric, one AI (Claude) produced the most complete reviews (≈76%) compared with humans (≈45%). Performance varied across AI systems and manuscript quality.
Doskaliuk et al. (2026)	Automated ethical compliance checking (ChatGPT-4.0)	Biomedical publishing	ChatGPT-4 showed high sensitivity/recall for missing ethics statements (IRB, consent, data sharing). Precision varied and was particularly low for funding (precision 0.16). Q4 journals had more omissions than Q1.
Cheng et al. (2025)	AI-generated text detection vs. human judgement	Healthcare journals	Three detectors differentiated human from AI-generated text ($p < .001$), but outputs differed across tools (ICC 0.57–0.95). Humans were approximately 19% accurate.
Yao et al. (2025)	AI handling of retracted references (ChatGPT-4 vs. others)	Bibliometric analysis (stem cell literature)	ChatGPT-4 retrieved approximately 80% of retracted articles and flagged approximately 62%. Grok 3 flagged 67%; DeepSeek almost none. When unaware

Author (Year)	AI Use Case	Domain/Context	Key Findings & Outcomes
			of retraction, models frequently fabricated citations (ChatGPT: 26%; Grok: 63%; DeepSeek: 88% of missing cases).
Kim et al. (2025)	Automatic abstract generation with LLM AI	Research publications / information science	Comparing author-written and AI-generated abstracts from 120 conference papers, different prompt-engineering approaches did not produce major semantic differences. Experts often preferred AI-generated abstracts when author abstracts diverged from paper content.
Ziyang (2025)	Analysis of AI-use declarations in academic writing	Cross-disciplinary scholarly publishing	Based on 7,953 AI-use declarations, AI adoption rose from 62.6% to 78.2% between October 2023 and March 2025. ChatGPT dominated usage across fields, while adoption differed by discipline and function.
Simsek et al. (2024)	Comparative analysis of journal AI-use policies	Imaging / radiology journals	Among 112 MEDLINE-indexed imaging journals, 69 (61.6%) had an AI-use policy and 40 specifically addressed AI-generated figures. Journals with AI policies tended to have higher citation and impact metrics.
Marrella et al. (2025)	AI-generated vs. human peer review	Surgery / hand surgery publishing	Across 11 published articles, ChatGPT-generated reviews received higher ARCADIA quality scores (4.8–4.9) than human reviews (2.8–3.2). Concordance with high-impact journal decisions remained limited (29%–32%).
Sheridan et al. (2025)	AI generation of discussion and conclusion sections	Orthopaedic journal publishing	Reviewer-blinded assessment showed AI-generated discussion/conclusion sections scored somewhat lower than human-written sections, but 5 of 6 reviewers recommended acceptance after major revision.

Author (Year)	AI Use Case	Domain/Context	Key Findings & Outcomes
Callanan et al. (2025)	Detection and prevalence of AI-based writing assistance	Orthopaedic literature	Using ZeroGPT on 3,374 recent orthopaedic manuscripts, the study estimated that 16.7% showed significant AI involvement above its predefined threshold. Rates ranged from 5.6% to 38.3% across journals.
Goyanes et al. (2025)	Content analysis of author guidelines on AI	Eight social science disciplines	Among leading journals, 73.8% mentioned AI for drafting text, 61.3% required transparency, but only 12% addressed privacy/data protection. Acceptance varied by discipline.
Raman (2025)	Analysis of ChatGPT-use acknowledgements	Cross-disciplinary and cross-geographic scholarly literature	Reviewing 1,226 publications, acknowledgement patterns were concentrated in the United States, China, and India, particularly in biomedical/clinical and computing fields.
Saad et al. (2024)	ChatGPT-assisted peer review	Medical/orthopaedic publishing	In an observational study of 21 peer-reviewed articles, ChatGPT-3.5 and 4.0 produced reviews with moderate similarity to human reviewers (mean 3.6/5 and 3.76/5, respectively).
Kadi & Aslaner (2024)	AI-generated case writing and AI-based peer review	Medical article writing / emergency medicine	ChatGPT generated case reports, but the reports contained referencing inaccuracies and received mixed reviewer ratings; 33.3% of professionals recommended rejection. Peer-review performance was weaker than text-generation performance.
Hill-Yardin et al. (2023)	Evaluation of ChatGPT-generated scientific text and manuscript drafting	Scientific publishing / neuroscience / academic writing	ChatGPT generated text that was logical, grammatically correct, and rapidly produced, but limitations included fabricated references, generic writing, limited source transparency, weak high-level critical

Author (Year)	AI Use Case	Domain/Context	Key Findings & Outcomes
			thinking, and difficulties with specialised knowledge.
Verharen (2023)	AI analysis of peer-review language and bias	Neuroscience / meta-science of peer review	Using more than 500 peer-review reports from 200 neuroscience papers, ChatGPT identified variability in reviewer scoring and patterns suggesting gender-related differences in review tone and favourability.
Nazarovets & Teixeira da Silva (2025)	Bibliometric analysis of AI listed as author	Academic authorship ethics / scholarly publishing	A database review identified 14 papers in which ChatGPT was credited as author, co-author, or group author despite existing guidance from COPE and ICMJE.
Strzelecki (2025)	Identification of undeclared AI-generated text in published papers	Cross-disciplinary scholarly publishing	Using the SPAR4SLR approach, the study identified papers in premier journals that appeared to contain undeclared ChatGPT-generated content across multiple disciplines; several had already begun receiving citations.

Discussion

Principal Findings and Overall Interpretation

The findings of this review indicate that the role of artificial intelligence in scientific publishing is best understood not as a simple replacement of human scholarly work, but as a redistribution of tasks between automated systems and human expertise. Across the 22 included studies, AI performed particularly well in linguistic, organisational, and other structured tasks, including abstract rewriting, translation, manuscript formatting, compliance screening, and checklist-based review. However, its performance was less consistent when tasks required technical accuracy, critical interpretation, contextual reasoning, reference verification, or accountability. This distinction provides an important framework for interpreting the rapidly expanding use of generative AI across scientific publishing.

The contrast between linguistic fluency and scholarly depth was particularly evident in studies of scientific writing. de Albuquerque and Gomes Dos Santos (2026) found that AI-rewritten abstracts were often preferred for readability and completeness, while Nathani et al. (2025) demonstrated that an AI-generated manuscript could achieve higher readability scores despite substantially lower technical accuracy and analytical depth. Kim et al. (2025) similarly showed that AI-generated abstracts could be perceived favourably when human-written abstracts inadequately represented the source article. Together, these findings suggest that improvements in surface-level communication should not be interpreted as

equivalent improvements in scientific reasoning. Fluency may enhance accessibility, but it can also create an impression of authority that exceeds the evidentiary or analytical quality of the underlying content.

This interpretation is consistent with the earlier observations of Hill-Yardin et al. (2023), who described ChatGPT-generated scientific prose as rapid, logical, and grammatically correct while simultaneously warning about generic argumentation, fabricated references, and limited critical reasoning. Sheridan et al. (2025) further showed that AI-generated discussion and conclusion sections could approach publishable quality after substantial human revision, while Kadi and Aslaner (2024) found that apparently coherent AI-generated medical text could still contain reference inaccuracies and substantive weaknesses. The central issue, therefore, is not whether generative AI can produce academically styled text, but whether the resulting text preserves the evidentiary, interpretive, and disciplinary standards expected of scientific scholarship.

AI-Assisted Writing: From Linguistic Support to Scholarly Reasoning

The strongest and most immediately defensible application of generative AI appears to be linguistic and editorial assistance rather than autonomous intellectual production. Pinto et al. (2026) demonstrated substantial improvements in translation quality for non-native English-speaking authors, supporting the use of AI for grammar, clarity, and language conversion. These benefits are particularly relevant because linguistic proficiency and scientific expertise are not equivalent, yet language quality can influence how research is perceived during editorial and peer-review processes.

At the same time, AI-supported writing raises a distinction between language mediation and intellectual delegation. Language polishing can help researchers communicate ideas they have already developed, whereas delegating literature synthesis, conceptual framing, or interpretation may transfer elements of scholarly reasoning to systems that do not possess disciplinary accountability. Iwasawa et al. (2026) provide an important educational perspective: increased access and permissive institutional policy were accompanied by greater student adoption, but knowledge of appropriate AI use remained incomplete. Linardon et al. (2025) similarly reported concerns that excessive reliance on AI could reduce opportunities for critical thinking.

This distinction has implications beyond student learning. Scientific writing is itself part of the reasoning process: constructing an argument, reconciling conflicting evidence, and deciding which interpretations are defensible are intellectual activities rather than merely linguistic ones. Consequently, responsible AI-assisted writing should preserve the human author's active role in these processes. The findings of Nathani et al. (2025), Sheridan et al. (2025), and Kadi and Aslaner (2024) reinforce this interpretation by showing that polished AI-generated prose can coexist with limited analytical depth, reference inaccuracies, or the need for substantial revision.

The increasing visibility of AI-assisted writing in published literature also complicates reliance on detection as a governance mechanism. Callanan et al. (2025) reported measurable levels of probable AI involvement in orthopaedic manuscripts, while Strzelecki (2025) documented apparent undeclared AI-generated content across several disciplines. However, Cheng et al. (2025) found substantial variability among automated AI-text detectors and very poor human accuracy in identifying AI-generated passages. These findings weaken the case for treating detection scores as

definitive evidence of inappropriate AI use. Disclosure, provenance, and accountable authorship are therefore likely to provide a more scientifically defensible basis for governance than stylistic detection alone.

AI in Editorial Screening and Peer Review: Augmentation Rather Than Replacement

The results also demonstrate meaningful potential for AI in editorial workflows. Doskaliuk et al. (2026) found that ChatGPT-4.0 achieved high sensitivity for identifying missing ethics and transparency statements, although variable precision produced false-positive alerts. This profile is well suited to editorial pre-screening: AI can identify items requiring attention, while editors determine whether a genuine reporting deficiency exists. Such an arrangement exploits the scalability of automated screening without transferring final responsibility to the model.

Evidence from peer-review studies suggests a similar pattern. George et al. (2025) reported that one LLM achieved substantially greater STROBE-based completeness than human reviewers. Marrella et al. (2025) also found higher structured quality scores for AI-generated reviews, whereas Saad et al. (2024) reported moderate similarity between ChatGPT-generated and human reviews. These findings indicate that AI may be particularly effective at systematically checking predefined criteria and generating consistently organised feedback.

However, checklist completeness and editorial judgement represent different dimensions of peer-review quality. A model may identify whether reporting elements are present without adequately determining whether a research question is important, an interpretation is biologically or theoretically plausible, or a finding constitutes a meaningful contribution to the field. Kadi and Aslaner (2024) demonstrated this limitation when AI-assisted peer review failed to identify important inconsistencies, while Sheridan et al. (2025) showed that AI-generated scholarly interpretation still required substantial revision. Hill-Yardin et al. (2023) likewise emphasised limitations in high-level critical reasoning.

These findings support a hybrid division of scholarly labour in which AI assists with repetitive, structured, or compliance-oriented tasks and human reviewers retain responsibility for novelty, methodological interpretation, disciplinary relevance, and final evaluative judgement. Importantly, this should not be interpreted as evidence that AI is inherently objective. The greater consistency observed by George et al. (2025) may reduce some forms of reviewer variability, but algorithmic systems can reproduce biases embedded in their training data and prompts. Figure 3 illustrates this task-dependent division, positioning AI primarily in bounded and verifiable activities while preserving human control over interpretive and accountable decisions.

AI may nevertheless offer a complementary role in examining the peer-review process itself. Verharen (2023) demonstrated that ChatGPT could analyse large numbers of review reports and detect patterns associated with variation in reviewer scoring and gender-related differences in tone. Such meta-review applications may be particularly valuable because they use AI to audit patterns across reviews rather than to replace the intellectual judgement of individual reviewers.

Confidentiality remains a fundamental constraint. Eggmann et al. (2025) highlighted the risks associated with transmitting unpublished manuscripts to external AI systems and proposed stronger safeguards, including restricted or locally hosted systems where appropriate. The distinction between using AI on publicly

available information and exposing unpublished manuscripts to third-party systems is therefore crucial: the latter introduces not only questions of accuracy but also obligations concerning confidentiality, intellectual property, and author trust.

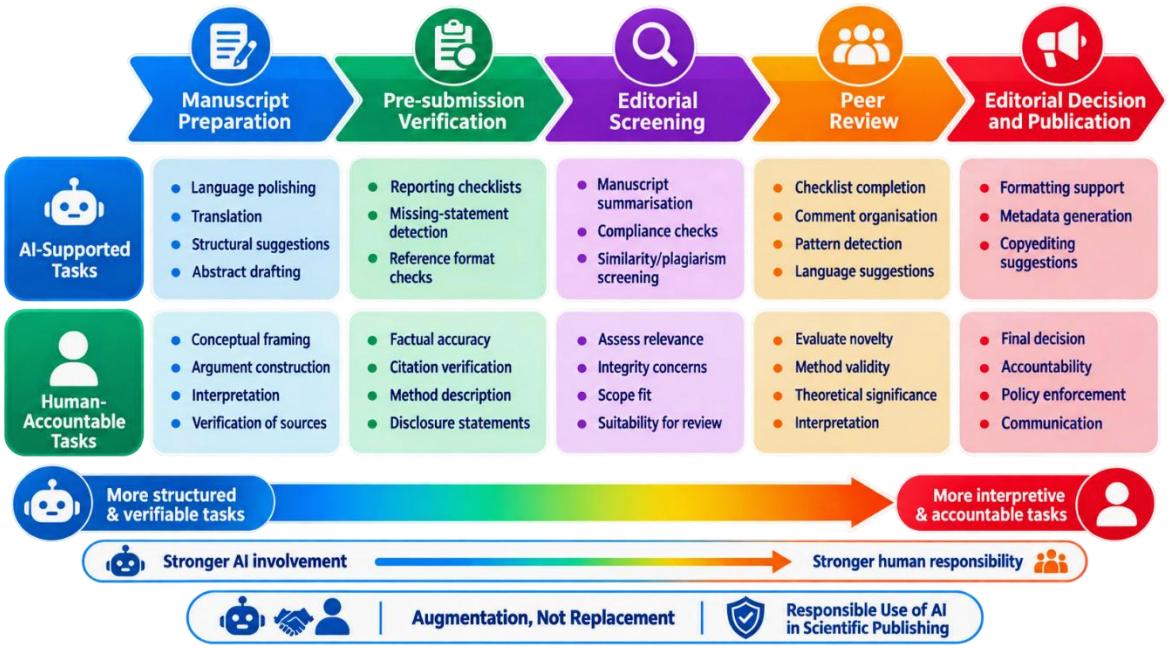

Figure 3. Proposed human–AI division of scholarly labour across the scientific publishing workflow. AI is positioned primarily as an assistive technology for structured, repetitive, and verifiable tasks, whereas humans retain responsibility for scientific reasoning, interpretation, accountability, and final decision-making (Doskaliuk et al., 2026; George et al., 2025; Hill-Yardin et al., 2023; Kadi & Aslaner, 2024; Marrella et al., 2025; Nathani et al., 2025; Saad et al., 2024; Sheridan et al., 2025; Verharen, 2023; Eggmann et al., 2025).

Citation Integrity, Transparency, and Human Accountability

Among the risks identified in this review, citation integrity represents one of the clearest boundaries for responsible use. LLMs generate text probabilistically rather than retrieving verified bibliographic records by default, allowing plausible but nonexistent references or inaccurate metadata to be produced alongside otherwise fluent prose. Hill-Yardin et al. (2023) identified fabricated references as an early and prominent weakness of ChatGPT, while Jain et al. (2025) characterised reference hallucination as a direct threat to citation integrity.

Yao et al. (2025) provide further empirical support for this concern. Their comparison of AI systems showed substantial variability in the identification of retracted literature and documented fabricated references when models were unable to retrieve relevant sources. The implication is not simply that AI-generated references can be incorrect, but that the linguistic plausibility of those references may conceal their unreliability from users who do not independently verify them.

This problem becomes more consequential when AI-assisted text enters the published record without transparent disclosure. Strzelecki (2025) identified apparently undeclared AI-generated passages in published papers, some of which had already received citations. Once unreliable AI-generated material becomes part of the scholarly record, subsequent authors may unknowingly propagate its claims or references, transforming an individual verification failure into a broader problem of evidence provenance.

Accordingly, verification must remain a human responsibility. Mijatović et al. (2026) emphasise that LLMs cannot replace human reasoning, interpretation, or accountability. Current international guidance reinforces this principle. COPE states that AI systems cannot qualify as authors because they cannot assume responsibility for submitted work, while human authors remain responsible for content produced with AI assistance (COPE Council, 2024). Current ICMJE guidance likewise emphasises human responsibility for accuracy, attribution, disclosure, and verification of AI-assisted content and cautions against uploading confidential manuscripts to systems where confidentiality cannot be assured (International Committee of Medical Journal Editors [ICMJE], 2026). These principles shift the ethical focus away from whether AI was used at all and toward whether its use remained transparent, verifiable, and accountable.

Authorship is therefore more than a matter of attribution. Nazarovets and Teixeira da Silva (2025) identified published papers in which ChatGPT had been credited as an author despite existing ethical guidance. Tao and Shen (2025) likewise documented the broader scholarly debate surrounding AI authorship. The persistence of such cases illustrates a gap between policy availability and policy implementation. Because authorship entails responsibility for accuracy, integrity, intellectual contribution, and responses to critique, assigning authorship to a non-accountable system undermines the accountability structure on which scientific publication depends.

The broader transparency problem includes plagiarism, copyright, and data confidentiality. Mijatović et al. (2026) emphasise these risks, particularly when users submit unpublished or proprietary material to external systems. Eggmann et al. (2025) extend the concern specifically to peer review, where confidentiality is an explicit professional obligation. The WAME recommendations similarly emphasise transparency, authorship responsibility, and protection of confidential scholarly material when generative AI is used (Zielinski et al., 2024). Across these sources, a consistent governance principle emerges: responsibility may be assisted technologically, but it cannot be transferred technologically.

Cross-Disciplinary Variation in Adoption and Practice

The results demonstrate that AI adoption is not uniform across scholarly fields. Ziyang (2025) showed substantial differences in AI-use declarations across disciplines and functions, while Raman (2025) identified geographic and disciplinary concentrations in ChatGPT acknowledgements. These differences suggest that AI adoption is shaped not only by technological availability but also by disciplinary norms, writing practices, editorial cultures, resource environments, and perceived risks.

Evidence from biomedical publishing illustrates rapid but uneven adoption. Saner et al. (2026) reported increasing indicators of ChatGPT-assisted writing in biomedical literature, although prevalence varied by publisher and geography. Carlson et al. (2026) found widespread formal policy adoption among neurosurgical journals but relatively limited numbers of articles formally declaring AI assistance. Simsek et al. (2024) likewise demonstrated that AI policies were more common among some imaging journals than others and were associated with journal-level characteristics.

These findings highlight an important distinction between policy adoption and behavioural adoption. A journal may establish an AI policy without necessarily producing high rates of disclosure, while researchers may use AI in environments where explicit policies remain absent or underdeveloped. Consequently, declaration frequencies should not automatically be interpreted as direct measures of true AI use. They are simultaneously indicators of adoption, policy awareness, and willingness to disclose.

The contrast across fields is also evident in social sciences and humanities. Tao and Shen (2025) identified substantial critical and ethical framing in social-science discourse on ChatGPT, while Goyanes et al. (2025) found disciplinary differences in journal guidelines, particularly concerning acceptable uses and privacy protections. Trojanowski and Barmentloo (2025) identified generative AI as an emerging theme in marketing scholarship but observed that ethical and social dimensions remained comparatively underdeveloped.

These patterns caution against universal governance models that ignore disciplinary context. A model appropriate for language editing in a technical field may not map directly onto disciplines in which authorial interpretation, rhetorical positioning, or theoretical argument constitute central components of scholarly contribution. At the same time, common minimum standards—particularly disclosure, accountability, confidentiality, and reference verification—remain applicable across fields.

AI, Linguistic Inclusion, and Global Publishing Equity

One of the most consequential potential benefits of generative AI is its capacity to reduce linguistic barriers in international scientific publishing. Pinto et al. (2026) demonstrated that AI-assisted translation can substantially improve clarity and linguistic quality in technical writing. For researchers who possess scientific expertise but work outside dominant English-speaking environments, such tools may separate the evaluation of scientific contribution from disadvantages arising primarily from linguistic fluency.

This interpretation is supported by broader evidence. Prakash et al. (2025) documented convergence in characteristics of academic English across countries and suggested that digital and AI-assisted writing technologies may contribute to changing patterns of scholarly communication. Sharma (2026) conceptualises generative AI as a form of academic infrastructure that could extend proofreading, translation, drafting support, and related services to researchers who lack access to well-resourced institutional environments.

However, the equity effect of AI is conditional rather than automatic. Advanced models may require subscription fees, reliable connectivity, technical literacy, or institutional support. If higher-performing systems remain concentrated among well-resourced researchers, AI could reproduce existing inequalities rather than eliminate them. Raman (2025) similarly showed that documented AI-assisted publishing remains geographically concentrated, indicating that access and adoption are uneven in practice.

A second equity concern is epistemic homogenisation. Models trained predominantly on widely available and dominant-language materials may favour conventional rhetorical forms, mainstream perspectives, or widely represented knowledge traditions. Hill-Yardin et al. (2023) noted the tendency of ChatGPT to

produce generic and relatively uncontroversial prose, while Tao and Shen (2025) highlighted disciplinary concerns regarding loss of nuance. Thus, linguistic democratisation should not come at the cost of intellectual or cultural standardisation. Responsible use should improve communicative accessibility while preserving disciplinary diversity, authorial interpretation, and plural forms of scholarly reasoning.

Sharma's (2026) argument that resistance to AI may sometimes reproduce forms of academic gatekeeping is particularly relevant in this context. Rather than treating all AI assistance as equivalent, publishing policies should distinguish between uses that enhance linguistic accessibility and uses that delegate substantive intellectual responsibility. Such differentiation may provide a more equitable approach than either unrestricted adoption or blanket prohibition.

From Heterogeneous Policies to Responsible Governance

The evidence reviewed in this manuscript shows rapid development of AI-related policies but substantial variation in their implementation. Table 2 synthesises representative editorial and institutional responses and demonstrates both convergence and continuing divergence across governance settings. The most consistent principles are that AI should not be listed as an author, human users remain responsible for submitted content, AI assistance should be disclosed when required, and confidential manuscript material should not be exposed to insecure external systems.

Carlson et al. (2026) found that all neurosurgical journals examined had introduced disclosure requirements, although the required location and level of detail varied. Simsek et al. (2024) found similarly heterogeneous policy adoption among imaging journals. Eggmann et al. (2025) demonstrated even greater variation in policies governing AI-assisted peer review, ranging from prohibition to conditional use under confidentiality safeguards. Goyanes et al. (2025) showed that privacy and data-protection provisions were considerably less common than general statements about AI-assisted drafting.

Institutional policy is also evolving. Iwasawa et al. (2026) documented a transition from blanket prohibition toward conditional and educationally supported use. Linardon et al. (2025) similarly described field-specific guidance favouring supportive applications while emphasising verification and continued intellectual engagement. Together, these developments suggest that governance is moving from an early binary question—whether AI should be permitted—to a more mature question of which uses are acceptable, under what safeguards, and with what level of disclosure.

This direction is consistent with broader international guidance. COPE explicitly excludes AI systems from authorship because they cannot assume responsibility for submitted work (COPE Council, 2024). The ICMJE requires transparency regarding AI use, maintains human responsibility for accuracy and attribution, and warns reviewers and editors about confidentiality risks (ICMJE, 2026). WAME similarly recommends transparent reporting of generative AI use and emphasises accountability and confidentiality in scholarly publication (Zielinski et al., 2024). The convergence of these independent frameworks strengthens the case for governance based on function, risk, and accountability rather than on the mere presence or absence of AI assistance.

Table 2. Editorial and institutional policy responses on AI usage in scientific publishing

Journal/Organisation & Context	Policy Stance on AI in Writing/Review	Key Policy Provisions
COPE (Publication Ethics) – multidisciplinary, global guideline	AI cannot be listed as an author; AI use must be disclosed.	AI cannot satisfy accountability requirements for authorship. Human authors remain responsible for AI-assisted content and should transparently report relevant AI use.
Neurosurgery journals (Carlson et al., 2026) – 9 top journals (h-index ≥100)	All require disclosure of AI-assisted writing; AI authorship prohibited.	All nine journals required authors to indicate whether and how AI was used. Disclosure location and granularity varied across journals.
Dental journals (Eggmann et al., 2025) – top 50 by Impact Factor	Mixed: many prohibit AI in peer review or lack explicit policy; a minority permit limited use under specified conditions.	Major concerns include confidentiality, factual reliability, and reviewer accountability. Proposed safeguards include disclosure to editors, permission where required, and secure or locally hosted tools.
International Journal of Eating Disorders (Linardon et al., 2025) – field-specific guidance	Supportive uses are permitted with human oversight; independent AI judgement is discouraged.	Guidance emphasises verification, continued intellectual engagement, and disclosure where relevant. Survey findings indicated substantial use of AI for writing support but lower rates of disclosure.
University policy (Pharmacy school) (Iwasawa et al., 2026) – academic integrity and training	Shift from blanket prohibition in 2023 to conditional use in 2024.	AI assistance must be used within specified guidelines, with emphasis on responsible use, declaration, protection of sensitive information, and AI literacy.

As Table 2 illustrates, policy convergence is strongest at the level of fundamental principles but weaker at the level of implementation. Journals differ in where disclosure should appear, which forms of assistance require declaration, whether reviewers may use AI, and what technical safeguards constitute adequate protection of confidential material. Such heterogeneity may create uncertainty for authors who publish across multiple disciplines or publishers.

A practical next step would therefore be greater harmonisation of minimum reporting standards while retaining discipline-specific flexibility. At minimum, reporting frameworks could distinguish the AI tool used, the purpose of use, the stage of the research or publication process in which it was applied, and the human verification undertaken. For peer review, confidentiality and editorial permission require additional consideration.

Figure 4 integrates the principal governance conditions identified across the review, showing how scientific integrity, transparency, human accountability, confidentiality, and equity must operate together across the scholarly ecosystem.

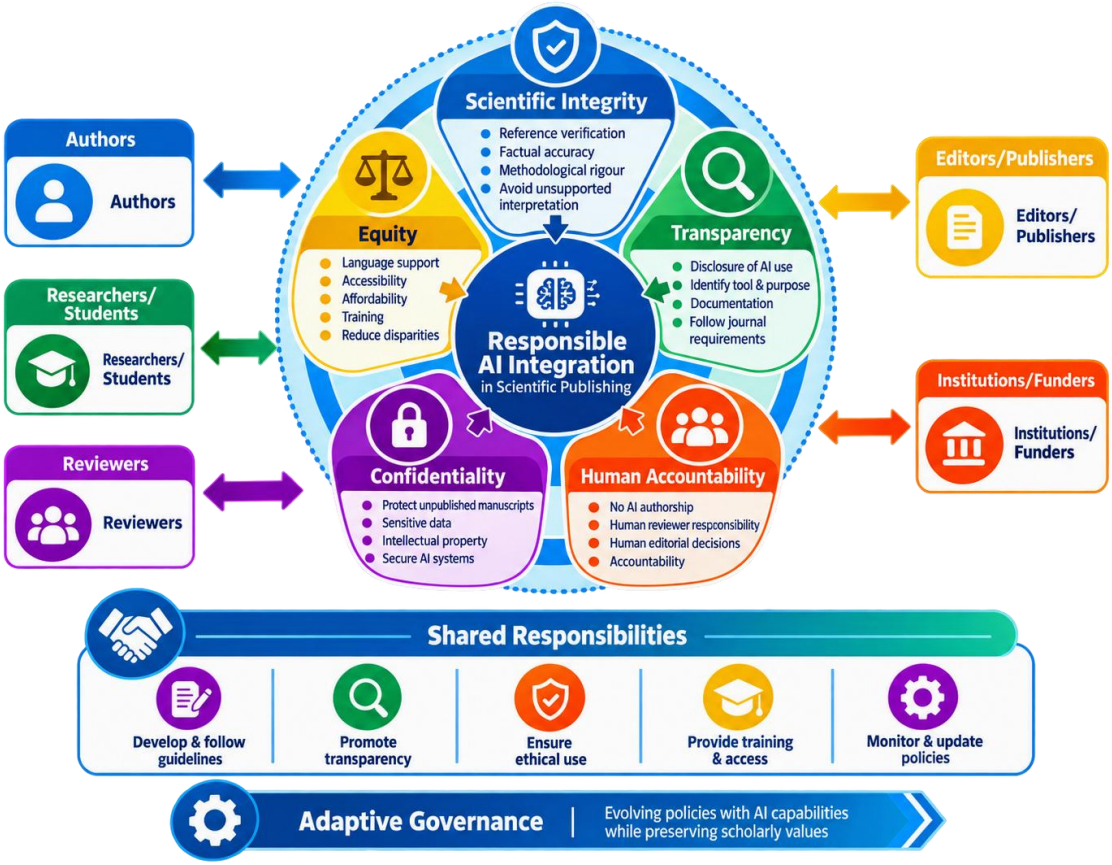

Figure 4. Integrated framework for responsible and equitable use of artificial intelligence in scientific publishing. Responsible integration requires the simultaneous protection of scientific integrity, transparency, human accountability, confidentiality, and equitable access across authorship, peer review, editorial practice, and institutional governance based on synthesized results of Carlson et al. (2026), Eggmann et al. (2025), Goyanes et al. (2025), Hill-Yardin et al. (2023), Iwasawa et al. (2026), Mijatović et al. (2026), Pinto et al. (2026), Raman (2025), Sharma (2026), and international guidance from COPE, ICMJE, and WAME.

Implications for Scientific Publishing and Future Research

Taken together, the findings support a model of responsible augmentation rather than autonomous scholarly production. The benefits of AI are most convincing when the task is transparent, bounded, and verifiable—for example, language refinement, translation, structured compliance checking, or assistance with organising reviewer feedback. Risk increases as AI use moves toward tasks that depend on originality, interpretation, evidentiary judgement, or responsibility for scientific claims.

Several areas require further investigation. First, comparative studies should evaluate AI-assisted peer review using measures that extend beyond checklist completeness to include methodological reasoning, novelty assessment, citation appropriateness, and domain-specific interpretation. Second, studies of AI-assisted writing should distinguish linguistic improvement from changes in scientific accuracy and conceptual depth. Third, longitudinal research is needed to determine whether sustained AI use affects the development of critical-writing and reasoning skills among early-career researchers. Fourth, equity research should examine whether

access to high-performing AI systems reduces or reinforces existing geographic and institutional disparities. Finally, policy research should evaluate whether disclosure requirements are understandable, consistently implemented, and capable of adapting to rapidly changing AI capabilities.

Because generative AI systems and publishing policies are evolving rapidly, the objective of governance should not be to establish static rules for a fixed technology. Rather, scientific publishing requires adaptive standards centred on enduring scholarly principles: accuracy, transparency, confidentiality, accountability, intellectual responsibility, and equitable participation. Under such a framework, AI can contribute meaningfully to scientific communication without being treated as an autonomous scientific actor.

Strengths and Limitations of the Review

This review offers an integrative perspective by connecting evidence on scientific writing, editorial workflows, peer review, disciplinary adoption, ethics, global equity, and governance rather than examining these domains in isolation. The inclusion of studies from multiple disciplines also enables comparison of emerging practices across diverse publishing environments.

However, several limitations should be acknowledged. First, as a narrative review, the study was designed to provide a focused and critical synthesis rather than exhaustive coverage or quantitative meta-analysis. Second, the evidence base is highly heterogeneous with respect to AI models, prompts, disciplinary settings, study designs, and evaluation criteria, limiting direct comparison of performance estimates across studies. Third, AI technology and journal policies are changing rapidly, meaning that specific model capabilities and governance arrangements may evolve after publication. Finally, some available evidence relies on indirect indicators of AI use, including linguistic markers, automated detection systems, and author declarations, each of which has methodological limitations. These considerations reinforce the need to interpret prevalence estimates and cross-study comparisons cautiously.

CONCLUSIONS

Artificial intelligence is reshaping scientific publishing by creating both substantial opportunities and significant challenges. Across the reviewed evidence, AI demonstrated its greatest value in supportive and structured tasks, including language polishing, translation, abstract and draft generation, editorial screening, and checklist-based review. These applications can improve efficiency, readability, and access to scholarly communication, particularly for non-native English-speaking researchers and authors with limited access to professional language support.

However, improvements in fluency and organisation do not consistently translate into stronger scientific reasoning. Generative AI may produce technically inaccurate statements, fabricated or unreliable citations, shallow interpretation, and ambiguous authorship or disclosure practices. Its performance is therefore most reliable when outputs can be independently verified and when the task does not require autonomous scholarly judgement. Authors, reviewers, and editors must remain responsible for verifying claims, checking references, protecting confidential information, and ensuring methodological and interpretive rigour.

The review also demonstrates that AI adoption and governance remain uneven across disciplines, journals, institutions, and geographic contexts. While policies

increasingly converge around transparency, human accountability, prohibition of AI authorship, and responsible disclosure, implementation remains heterogeneous. This variability reflects differences in disciplinary norms, editorial resources, perceived risks, and access to advanced AI systems.

Overall, the evidence supports a model of responsible augmentation rather than replacement. AI should be treated as an assistive scholarly technology that can enhance selected components of scientific communication while remaining subject to meaningful human oversight. Its responsible integration depends on transparency, verification, confidentiality, accountability, and equitable access. Scientific judgement, interpretation, and responsibility must remain fundamentally human.

RECOMMENDATIONS

Based on the findings of this review, several priorities should guide the responsible integration of AI into scientific publishing.

First, authors should use generative AI primarily for clearly defined and verifiable tasks, such as language refinement, translation, structural assistance, and preliminary checking. AI-generated factual statements, interpretations, and references should be independently verified before submission, and human authors should retain full intellectual and ethical responsibility for the final manuscript.

Second, journals and publishers should develop clearer and more harmonised disclosure standards. These standards should specify when AI use must be reported, what information should be disclosed, and where the disclosure should appear. At minimum, reporting should identify the AI tool, the purpose for which it was used, and the extent of human verification.

Third, AI use in peer review and editorial assessment should remain assistive rather than autonomous. AI may be valuable for structured screening, reporting checklists, and identification of potential omissions, but final decisions concerning novelty, methodological validity, interpretation, and scientific significance should remain with qualified human editors and reviewers. Confidential manuscripts should not be uploaded to external AI systems unless adequate data-protection and confidentiality safeguards are in place.

Fourth, institutions and journals should strengthen AI literacy among researchers, reviewers, editors, and students. Training should extend beyond technical tool use to include reference verification, disclosure requirements, confidentiality, bias, authorship responsibility, and the distinction between linguistic assistance and delegation of scholarly reasoning.

Fifth, efforts to integrate AI into publishing should explicitly address equity. AI-assisted language and translation tools may reduce barriers for non-native English-speaking and under-resourced researchers, but unequal access to high-performing systems could reinforce existing disparities. Institutions, publishers, and funders should therefore consider equitable access, training, and infrastructure as part of responsible AI governance.

Finally, future research should evaluate AI performance using measures that extend beyond readability and checklist completeness. Priority areas include technical accuracy, analytical depth, citation integrity, domain-specific reasoning, long-term effects on critical-writing skills, confidentiality-preserving AI systems, and the effectiveness of disclosure and governance policies across disciplines. Longitudinal

and cross-disciplinary studies will be particularly important as both AI capabilities and publishing practices continue to evolve.

Acknowledgment
The authors gratefully acknowledge all individuals and institutions who contributed to the completion of this study through academic discussion, technical assistance, editorial support, and constructive feedback during the preparation and revision of the manuscript. The authors also acknowledge the use of ChatGPT (OpenAI) as an assistive tool for language refinement, structural organisation, readability improvement, and presentation of the manuscript. All AI-assisted content was critically reviewed, revised, and verified by the authors, who retained full responsibility for the scientific accuracy, interpretation of the literature, citation integrity, and final content of the manuscript.

Funding Information
This research received no external funding.

Conflict of Interest Statement
The authors declare no conflict of interest.

Ethical Approval
Not applicable.

Data Availability
Not applicable.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Lalu Nurul Yaqin	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

REFERENCES

Callanan, T., Marquez, J., Pisani, C., Schmitt, P., Pietro, J., Chen, M., Milner, J., Daher, M., Katz, L., Liu, J., & Daniels, A. H. (2025). Evaluating artificial intelligence-based writing assistance among published orthopaedic studies: Detection and trends for future interpretation. *Journal of Bone and Joint Surgery, 107*(16), 1887–1893. https://doi.org/10.2106/JBJS.24.01462

Carlson, B., Laffaye, T., Gray, L., Bathini, A. R., Dubey, A., & Patra, D. P. (2026). The use of artificial intelligence in neurosurgical manuscript writing: Journal specific policies and their implementation. *Journal of Clinical Neuroscience, 144*, 111819. https://doi.org/10.1016/j.jocn.2025.111819

Cheng, A., Lin, Y., Reedy, G., Joseph, C., Wirkowski, S., Mallette, V., Nagesh, V., Krieser, D., & Calhoun, A. (2025). Ability of AI detection tools and humans to accurately identify different forms of AI-generated written content. *Advances in Simulation, 10*(1). https://doi.org/10.1186/s41077-025-00396-6

COPE Council. (2024). *COPE position – Authorship and AI*. Committee on Publication Ethics. https://doi.org/10.24318/cCVRZBms

de Albuquerque, F., & Gomes Dos Santos, P. (2026). Who moved my text? Assessing whether ChatGPT can write better abstracts than humans. *Cogent Education, 13*(1). https://doi.org/10.1080/2331186X.2026.2616858

Doskaliuk, B., Seiil, B., & Qumar, A. (2026). ChatGPT-4.0 as a tool for automated review of ethics and transparency in biomedical literature. *Journal of Korean Medical Science, 41*(2), e24. https://doi.org/10.3346/jkms.2026.41.e24

Eggmann, F., Hildebrand, H., & Bornstein, M. M. (2025). Who reviewed this? Toward responsible integration of large language models for peer review of scientific articles in dental medicine. *Swiss Dental Journal, 135*(3), 1–15. https://doi.org/10.61872/sdj-2025-03-01

George, R., Ismail, N. M., Kukkamalla, M. A., Abas, A. L., Soe, H. H. K., Donald, P. M., & Ismail, A. R. H. (2025). Comparative evaluation of manuscript review accuracy: Human reviewers vs. AI chatbots. *Malaysian Journal of Medicine and Health Sciences, 21*, 26–32. https://doi.org/10.47836/mjmhs.21.s13.5

Goyanes, M., Lopezosa, C., & Piñeiro-Naval, V. (2025). The use of artificial intelligence (AI) in research: A review of author guidelines in leading journals across eight social science disciplines. *Scientometrics, 130*(7), 3725–3741. https://doi.org/10.1007/s11192-025-05377-0

Hill-Yardin, E. L., Hutchinson, M. R., Laycock, R., & Spencer, S. J. (2023). A Chat(GPT) about the future of scientific publishing. *Brain, Behavior, and Immunity, 110*, 152–154. https://doi.org/10.1016/j.bbi.2023.02.022

International Committee of Medical Journal Editors. (2026). *Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals*. https://www.icmje.org/recommendations/

Iwasawa, M., Kobayashi, M., & Otori, K. (2026). Tracking changes in pharmacy students' engagement with ChatGPT: A comparative cross-sectional study (2023–2024). *Currents in Pharmacy Teaching and Learning, 18*(3), 102555. https://doi.org/10.1016/j.cptl.2025.102555

Jain, A., Nimonkar, P., & Jadhav, P. (2025). Citation integrity in the age of AI: Evaluating the risks of reference hallucination in maxillofacial literature. *Journal of Cranio-Maxillofacial Surgery, 53*(10), 1871–1872. https://doi.org/10.1016/j.jcms.2025.08.004

Kadi, G., & Aslaner, M. A. (2024). Exploring ChatGPT's abilities in medical article writing and peer review. *Croatian Medical Journal, 65*(2), 93–100. https://doi.org/10.3325/cmj.2024.65.93

Kim, Y., Lee, J., & Yang, S. (2025). Exploring LLM AI in automatic generation of abstracts for research publications. *Proceedings of the Association for Information Science and Technology, 62*(1), 967–971. https://doi.org/10.1002/pra2.1323

Linardon, J., Thomas, J. J., Crow, S. J., Ghaderi, A., Hilbert, A., Klump, K. L., Wade, T. D., Walsh, B. T., & Weissman, R. (2025). Conducting eating disorder research in the era of generative AI: Researcher perspectives and guidelines from the International Journal of Eating Disorders. *International Journal of Eating Disorders, 58*(12), 2306–2316. https://doi.org/10.1002/eat.24543

Manohar, N., & Prasad, S. S. (2023). Use of ChatGPT in academic publishing: A rare case of seronegative systemic lupus erythematosus in a patient with HIV infection. *Cureus, 15*(2), e34616. https://doi.org/10.7759/cureus.34616

Marrella, D., Jiang, S., Ipaktchi, K., & Liverneaux, P. (2025). Comparing AI-generated and human peer reviews: A study on 11 articles. *Hand Surgery and Rehabilitation, 44*(4), 102225. https://doi.org/10.1016/j.hansur.2025.102225

Mijatović, A., Žuljević, M. F., Uršič, L., & Marušić, A. (2026). Responsible use of large language models in manuscript preparation. *Current Protocols, 6*(1). https://doi.org/10.1002/cpz1.70300

Nathani, K. R., Nathani, A.-M., Delawan, M., Safdar, A., & Bydon, M. (2025). Can artificial intelligence write science? A comparative analysis of human-written and artificial intelligence-generated scientific writings. *Journal of Neurosurgery: Spine, 43*(6), 767–772. https://doi.org/10.3171/2025.4.SPINE25519

Nazarovets, S., & Teixeira da Silva, J. A. (2025). ChatGPT as an "author": Bibliometric analysis to assess the validity of authorship. *Accountability in Research, 32*(7), 1148–1158. https://doi.org/10.1080/08989621.2024.2345713

O'Connor, S. (2023). Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Education in Practice, 66*, 103537. https://doi.org/10.1016/j.nepr.2022.103537

Pinto, M. M., Araújo, E. E. R., Rahhal, Z., Stanbouly, D., & Grillo, R. (2026). AI-assisted translation in oral and maxillofacial surgery: A comparative evaluation. *Journal of Stomatology, Oral and Maxillofacial Surgery, 127*(3), 102717. https://doi.org/10.1016/j.jormas.2026.102717

Prakash, A., Aggarwal, S., Varghese, J. J., & Varghese, J. J. (2025). Writing without borders: AI and cross-cultural convergence in academic writing quality. *Humanities and Social Sciences Communications, 12*(1). https://doi.org/10.1057/s41599-025-05484-6

Raman, R. (2025). Transparency in research: An analysis of ChatGPT usage acknowledgment by authors across disciplines and geographies. *Accountability in Research, 32*(3), 277–298. https://doi.org/10.1080/08989621.2023.2273377

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 and Metabolic Syndrome: Clinical Research and Reviews, 18*(2), 102946. https://doi.org/10.1016/j.dsx.2024.102946

Saner, F. H., Schumann, R., Alchibi, L., Kareem, S. A., Marquez, K. A. H., Saner, Y. M., Abufarhaneh, E., Bröering, D. C., & Raptis, D. A. (2026). The evolution of ChatGPT-generated text in scientific medical publications. *Saudi Journal of Anaesthesia, 20*(1), 63–67. https://doi.org/10.4103/sja.sja_526_25

Sharma, D. (2026). Artificial intelligence as an equaliser: How ChatGPT empowers academics in the Global South. *Tropical Doctor, 56*(1), 6–8. https://doi.org/10.1177/00494755251369621

Sheridan, G. A., Howard, L. C., Neufeld, M. E., Doyle, T. R., Hughes, A. J., Sculco, P. K., Beverland, D. E., Garbuz, D. S., & Masri, B. A. (2025). Can artificial intelligence generate scientific discussion that passes peer review for publication in a high-impact orthopaedic journal? *Irish Journal of Medical Science, 194*(4), 1191–1198. https://doi.org/10.1007/s11845-025-03971-y

Simsek, O., Manteghinejad, A., & Vossough, A. (2024). A comparative review of imaging journal policies for use of AI in manuscript generation. *Academic Radiology, 31*(12), 5232–5236. https://doi.org/10.1016/j.acra.2024.05.006

Strzelecki, A. (2025). 'As of my last knowledge update': How is content generated by ChatGPT infiltrating scientific papers published in premier journals? *Learned Publishing, 38*(1). https://doi.org/10.1002/leap.1650

Tao, Y., & Shen, Q. (2025). Academic discourse on ChatGPT in social sciences: A topic modeling and sentiment analysis of research article abstracts. *PLOS ONE, 20*(10), e0334331. https://doi.org/10.1371/journal.pone.0334331

Trojanowski, M., & Barmentloo, E. (2025). Convergence of marketing technologies and artificial intelligence: A bibliometric review (1987–2025). *Engineering Management in Production and Services, 17*(4), 29–51. https://doi.org/10.2478/emj-2025-0025

Verharen, J. P. H. (2023). ChatGPT identifies gender disparities in scientific peer review. *eLife, 12*, e90230. https://doi.org/10.7554/eLife.90230

Yao, L., Gu, T., Li, X., Jiao, Y., Li, M., Graff, J. C., Li, Y., & Gu, W. (2025). AI misuse of retracted literature: A comparative study of ChatGPT4o, DeepSeek, and Grok 3 in stem cell research. *The Science of Nature, 112*, 85. https://doi.org/10.1007/s00114-025-02036-5

Zielinski, C., Winker, M. A., Aggarwal, R., Ferris, L. E., Heinemann, M., Lapeña, J. F., Pai, S. A., Ing, E., Citrome, L., Alam, M., Voight, M., & Habibzadeh, F. (2024). Chatbots, generative AI, and scholarly manuscripts: WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. *Current Medical Research and Opinion, 40*(1), 11–13. https://doi.org/10.1080/03007995.2023.2286102

Ziyang, X. (2025). Disciplinary diversity in academic AI adoption: A comparative analysis of AI tool usage declarations across scientific fields. *Proceedings of the Association for Information Science and Technology, 62*(1), 788–798. https://doi.org/10.1002/pra2.1297