American Journal of Multidisciplinary Research and Innovation (AJMRI)

ISSN: 2158-8155 (ONLINE), 2832-4854 (PRINT)

VOLUME 5 ISSUE 5 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Multidisciplinary Research and Innovation (AJMRI)

Volume 5 Issue 5, Year 2026
ISSN: 2158-8155 (Online), 2832-4854 (Print)
DOI: https://doi.org/10.54536/ajmri.v5i5.8415
https://journals.e-palli.com/home/index.php/ajmri

From Adoption to Accountability in Generative AI Use: A Risk-Based Framework

Richard Mulenga[1*], Mildred Muhyila[2]

Article Information

Received: June 12, 2026

Accepted: August 23, 2026

Published: September 16, 2026

Keywords

Gen-AI Risk-Based Framework, Generative Artificial Intelligence, Professional Ethics, Responsible AI Use

ABSTRACT

This study analyses the challenges and benefits of the increasing use of generative artificial intelligence (Gen-AI) models into academic, professional, and industrial writing practices. The rapid adoption of Gen-AI in academic, professional and industrial writing has significantly enhanced efficiency in ideation, translation, summarization, drafting, and editing, reducing both time and cost. But the adoption has outpaced the development of consistent ethical and governance standards. The use of Gen-AI models introduces substantial ethical and governance challenges including the generation of fabricated evidence, citation inaccuracies, violations of intellectual property, professional misrepresentation, and the erosion of human cognitive and compositional competencies. Existing responses to the Gen-AI challenges in literature remain fragmented, leaving uncertainty about acceptable assistance, disclosure, authorship and accountability regarding Gen-AI use. This study addresses the gaps by proposing an Integrated Risk-Based Framework for Responsible Generative AI Writing (IRF–GenAIW) for the responsible and accountable use of Gen-AI. Findings from the systematic literature review analysis of literature indicate that current institutional approaches are inconsistent and concentrate on issues of plagiarism and academic misconduct, with inadequate consideration of broader applications in professional and industrial sectors. The contribution of this study is that it shifts attention from whether Gen-AI is used to whether it is used responsibly and with integrity. It develops a risk-based Gen-AI framework that guides disclosure, verification, human oversight, authorship, confidentiality and accountability in Gen-AI use across various academic, professional and industrial sectors.

INTRODUCTION

The emergence and proliferation of generative artificial intelligence (Gen-AI) systems have changed both the economics and the process of writing. Large language models (LLMs) are able to generate texts with negligible additional costs and can aid in the processes of ideation, drafting, translation, reorganization, and editing. Thus, the application of such systems is valuable not only in the academic context but across an array of disciplines and industries. The value of Gen-AI extends beyond university education and academic publishing, as professionals in law, consulting, journalism, marketing, engineering, and politics can leverage its capacity to generate, restructure, and summarize writing. Corporate communicators can use such technology to prepare internal briefings, external announcements, and other materials for stakeholders. The emergence and use of Gen-AI is described by Dwivedi *et al.* (2023) as a multidisciplinary revolution affecting research, organizational practice, and policies, while Kasneci *et al.* (2023) point out the presence of educational opportunities of Gen-AI use accompanied by the risks of bias, inaccuracies, misuse, and lack of critical thinking. According to Al-Kfairy *et al.* (2024), Gen-AI can cause significant harm through misinformation, privacy and intellectual property risks, ethical and legal challenges, and security threats. The issue of authorship and intellectual property is particularly pertinent to the use of generative AI in academic, professional and industry communication. Traditional notions of authorship presuppose that the writer has made substantial intellectual contributions, understood the generated material, and takes responsibility for its accuracy and ethical integrity. Beyond the ethics of plagiarism, the use of Gen-AI models for writing involves issues of veracity, authorship, intellectual property, confidentiality, privacy, transparency, fairness, competence, and liability (Bozkurt, 2024; Chetwynd, 2024). From a similar perspective, confidentiality represents a significant concern in the use of generative AI for writing. This consideration is particularly relevant to professionals who may use such technology to add value to their practice, including lawyers, journalists, clinicians, researchers, and managers. When using AI for writing support, professionals risk disclosing confidential information, including research data, organizational strategies, private communications, legal documents, and client details (Kfairy *et al.*, 2024; Chetwynd, 2024).

The use of Gen-AI models raises critical questions about fairness and inclusion in professional communication and research practices. On the one hand, the use of such models can increase accessibility and promote inclusiveness by providing writing support to non-native language speakers and those unable to afford traditional editing and publishing services. On the other hand, such tools may also disadvantage certain populations by providing unequal levels of support due to varying levels of access to technology, digital literacy, and language competencies (UNESCO, 2023; Al-Kfairy *et al.*, 2024).

[1] Department of Economics, ZCAS University, Lusaka, Zambia
[2] School of Law, ZCAS University, Lusaka, Zambia
[*] Corresponding author's e-mail: richardmulenga2@gmail.com

Users of Gen-AI models must be diligent to check, fix and approve the work done with the help of Gen-AI models (Bozkurt, 2024; Chetwynd, 2024). UNESCO (2023) noted that the rise of Gen-AI preceded policy, regulation, and institutional preparedness. Current policies on GenAI writing tend to focus on detecting and deterring plagiarism and dishonesty by students and researchers. However, there are other risks presented by the use of Gen-AI including the generation of false evidence, provision of misleading professional advice, failure to disclose authorship, violation of confidentiality, intellectual property infringement, discrimination, and the de-skilling of human writers with the increased and intense use of Gen-AI tools (UNESCO, 2023; Magesh *et al.*, 2025). Additionally, policies that attempt to combat plagiarism by AI writing do not always succeed, since the text can be altered, reworded, or combined with other material to conceal its origins. Without clearer guidance, institutions are left speculating about what might count as acceptable use of Gen-AI, what should be disclosed, what must be verified, and who bears responsibility (Magesh *et al.*, 2025; Resnik & Hosseini, 2025). The rapid and evolving use of Gen-AI in academic, professional, and industrial publishing far outruns discussion of appropriate, ethical standards and policies. Universities, professional bodies, publishers, businesses, and governments have responded in different ways, banning certain kinds of AI-assisted writing or requiring disclosure, while leaving many questions about acceptable assistance, substitution, and professional misrepresentation open (Sallam, 2023; Bozkurt, 2024; Magesh *et al.*, 2025). It is therefore essential to move toward standardized but risk-sensitive governance policies that distinguish between benign and maleficent uses of generative AI, and establish requirements for disclosure, verification, human authorship, confidentiality, and accountability in Gen-AI use.

The central research problem is not whether or not Gen-AI has been used, but whether such use is done with integrity, which include truthfulness, traceability, intellectual property rights, privacy, fairness, competence, and responsibility. According to Dwivedi *et al.* (2023), Gen-AI can be employed academic settings to provide explanatory feedback, enhance accessibility, and enable learners who are linguistically or educationally disadvantaged to participate in writing-intensive tasks. It can also reduce the cognitive load associated with writing, enabling students and researchers to devote more attention to high-level processing and analytical activities. Yet, ethical concerns surrounding the use of Gen-AI and similar technological advancements remain. These include hallucination, plagiarism, deception, reduced intellectual effort, and undermining of sound writing and publishing practices (Kasneci *et al.*, 2023). UNESCO (2023) asserts that Gen-AI use should be governed by a human-centric approach that promotes autonomy, inclusion, privacy, intellectual development, and accountability. The corresponding key research question focuses not just on whether Gen-AI was used, but on how it was used, with what level of transparency and human verification, and what the potential negative consequences or risks accompany its use.

The aim of the study is to assess ethical, governance and contemporary issues confronting the adoption and use of Gen-AI in academic, professional, and industrial writing and to propose an integrated risk-based framework that promotes responsible and accountable use of Gen-AI.

LITERATURE REVIEW

Theoretical Underpinnings

This study is underpinned by two complementary theories; socio-technical systems theory (Trist & Bamforth, 1951) and the Cognitive offloading theory (Risko & Gilbert, 2016).

Socio-Technical Systems Theory

Socio-Technical Systems Theory (1951) posits that organisational performance is determined by the interplay between technical and social subsystems, encompassing human actors, institutional frameworks, established norms, defined roles, procedural workflows, and policy mechanisms. Optimal functionality necessitates the concurrent optimisation of both dimensions, as reliance on technological components in isolation is insufficient for achieving effective outcomes (Trist & Bamforth, 1951). Science and Technology Studies offer a suitable framework for examining Gen-AI, as issues including fabricated references, authorship attribution, transparency, intellectual property rights, privacy concerns, bias, and accountability are not inherent to the technology alone. Rather, these challenges arise from the complex interplay between AI systems and a range of actors, including users, academic institutions, employers, publishing entities, professional organizations, and regulatory bodies (UNESCO ,2023).

Cognitive Offloading Theory

Cognitive Offloading Theory (2016) elucidates the mechanisms by which individuals employ external actions or artifacts to alleviate the cognitive load associated with specific tasks (Risko & Gilbert, 2016). Generative artificial intelligence facilitates the offloading of various cognitive functions, including drafting, summarization, translation, idea organization, information retrieval, and argument construction. While such offloading may enhance efficiency and accessibility, unmonitored or excessive reliance on Gen-AI systems risks undermining core cognitive competencies particularly analytical reasoning, critical evaluation, knowledge consolidation, and autonomous writing proficiency (Kfairy *et al.*, 2024; Chetwynd, 2024).

MATERIALS AND METHODS

The search strategy, inclusion and exclusion criteria

This study employed a systematic literature review methodology to examine ethical, governance, and current challenges pertaining to the use of generative artificial intelligence (Gen-AI) in academic, professional, and industrial writing contexts. A systematic integrative

literature review approach was adopted as a suitable methodology given the multifaceted nature of the research inquiry, which spans empirical research, theoretical frameworks, professional guidelines, and regulatory documents. The review process integrated systematic literature retrieval with explicit criteria for study inclusion and exclusion, followed by a rigorous thematic synthesis that emphasized critical analysis of the sample literature. Reporting adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) framework, ensuring transparency and replicability throughout the stages of literature identification, screening, eligibility assessment, and data integration (Page *et al.*, 2021), supplemented by a Boolean proximity search strategy designed to capture semantically related terms such as " generative artificial intelligence," "generative AI," " ChatGPT," " large language model," " Academic writing using GenAI," " professional use of AI writing," " Industrial use of GenAI," " Publishing with generative AI" " Accountability in AI Usage," " Disclosure in AI usage in writing" " Human oversight in GenAI use," " Hallucinations of GenAI," " Plagiarism with AI use," and " Integrity in GenAI use". The databases surveyed in this study include Elsevier, Springer Nature, Taylor & Francis, Google Scholar, Wiley, MDPI, JMIR Publications, Oxford University Press, the American Association for the Advancement of Science (AAAS), the National Academy of Sciences, as well as Cureus and Open Praxis publishing platforms. Inclusion criteria required that materials be scholarly outputs both peer-reviewed, grey literature, reports, working papers published in English within the period, from January 2020 to May 2026 and made available under open access terms or without subscription-based access restrictions.

Publications were excluded if they were published outside the predefined time frame, not authored in English, or lacked publicly available full texts. Additionally, exclusion criteria encompassed works that did not engage meaningfully with generative artificial intelligence in the domains of writing, knowledge production, ethical considerations, evidentiary frameworks, or governance structures; those centered solely on technical aspects of model performance; and items classified as editorials, news articles, blog posts, promotional content, abstracts, presentation slides, non-peer-reviewed student research, or redundant entries. Further exclusions applied to studies providing inadequate methodological descriptions, formally retracted publications, and preprints for which peer-reviewed counterparts existed. Foundational sources originating outside the designated review period were retained exclusively to provide theoretical or historical background and were not incorporated into the systematic synthesis of empirical evidence.

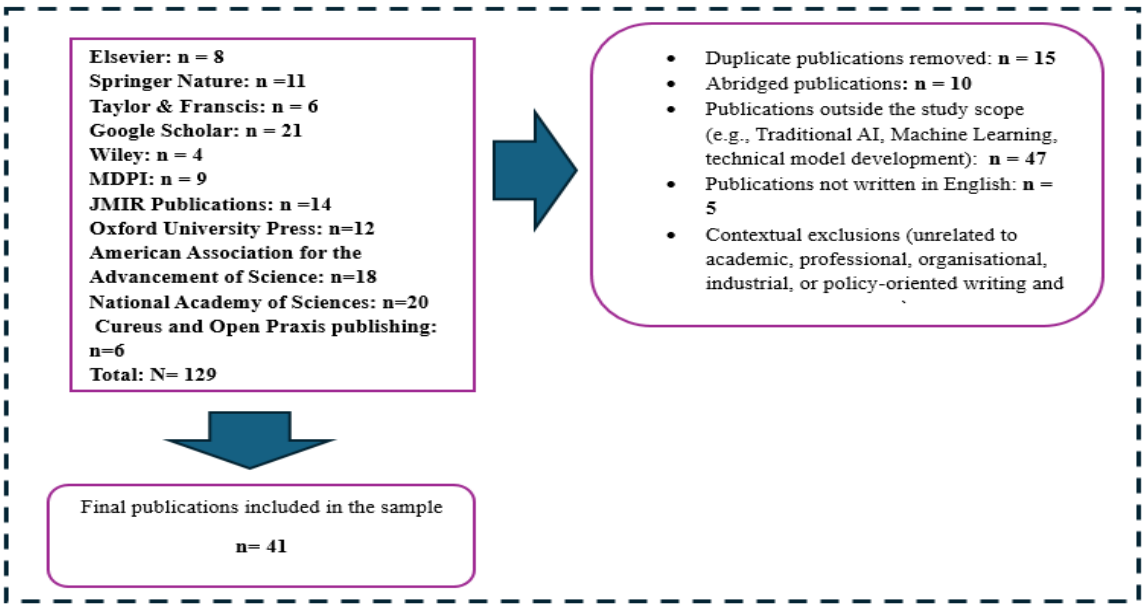

Figure 1: PRISMA 2021 Flow Diagram: Publication Identification and Selection Process

RESULTS AND DISCUSSION

A systematic literature search yielded 129 publications sourced from multiple academic databases, search engines, and publishing platforms. The distribution across publishers was as follows: eight from Elsevier, 11 from Springer Nature, six from Taylor & Francis, 21 from Google Scholar, four from Wiley, nine from MDPI, 14 from JMIR Publications, 12 from Oxford University Press, 18 from the American Association for the Advancement of Science, 20 from the National Academy of Sciences, and six from Cureus and Open Praxis Publishing. Of these, 88 publications were excluded during the screening phase. Exclusions consisted of 15 duplicates, 10 abridged versions, 47 studies falling outside the research scope particularly those centered on conventional artificial intelligence, machine learning, or technical model

development five non-English publications, and 11 works deemed contextually irrelevant due to their lack of engagement with academic, professional, organizational, industrial, or policy-related writing and knowledge production. As a result, 41 publications fulfilled the predefined inclusion criteria and were retained for analysis in the final systematic review. Table 1 reports a summary of geographical coverage, frequency of publications by region and the sectors of dominance.

Table 1: Geographical Coverage & Frequency of Publications

Region	Evidence represented in the review	Number of publications
Europe	Research associated with Germany, the Czech Republic and the United Kingdom, particularly education, detection and publication-integrity studies	12
North America (USA)	Research from the United States, including professional writing, productivity, legal research and education studies.	9
Asia-Pacific	Studies concerning AI-assisted learning, programming and technology use.	14
Middle East	Healthcare, scientific-writing and research-ethics literature.	6

Source: Authors' Elaboration on the Literature Review

The 41 studies incorporated in the systematic review originated from four principal geographic regions and spanned multiple domains. The Asia-Pacific region yielded the highest number of contributions, with 14 publications analyzing artificial intelligence in learning environments, programming applications, and technological integration. Europe contributed 12 studies, predominantly from Germany, the Czech Republic, and the United Kingdom, with a thematic emphasis on educational practices, AI-based content identification, and scholarly publication integrity. North America, represented exclusively by the United States, produced nine publications addressing areas such as professional writing, workplace efficiency, legal research methodologies, and educational innovation. The Middle East contributed six studies, primarily investigating applications in healthcare, scientific authorship, and ethical considerations in research. However, the review does not clearly include studies situated in Africa or Latin America. This limits the applicability of its conclusions to multilingual, lower-resource and institutionally different settings. Collectively, the body of evidence encompasses education, technology, healthcare, legal studies, professional communication, publishing, and research integrity. However, the geographic distribution reveals a notable underrepresentation of Africa, Latin America, and other developing regions.

Empirical Literature Review

Identified Themes from the Empirical Review

Gen-AI as Transformative Tools of Writing and Knowledge Generation

The literature published after the public availability of the first advanced large language models (LLMs) positions Generative Artificial Intelligence (Gen-AI) as both writing facilitator, source of knowledge-production and institutional risk. The early articles on the subject matter were multidisciplinary in nature, with Dwivedi *et al.* (2023) bringing together researchers from information systems, education, management, marketing, computer science and public policy in their systematic literature review. By utilising an expert-opinion methodology, the authors concluded that GenAI offered substantial benefits for students by supporting drafting, summarising, and communication with customers and researchers alike, while posing significant risks in terms of misinformation, bias, privacy, intellectual property and accountability, yet it also fails to produce any concrete metrics on the subject. Kasneci *et al.* (2023) attempt a similar approach in their position paper titled 'Large Language Models in Education', managing to produce another comprehensive assessment while utilising almost the same set of disciplines. Like Dwivedi *et al.* (2023), Kasneci *et al.* (2023) highlight the ability of GenAI to provide assistance in learning, planning, explaining and evaluating lessons, translating languages and providing language-specific feedback. While the authors suggest that AI literacy, human monitoring and redesigned assessments could help address these issues, their recommendations remain too conceptual and do not attempt to empirically prove the effectiveness of specific methods, including disclosure statements, oral exams or limited-scope evaluations.

Evidence on Hallucinations, Fabricated References and False Evidence

A critical issue in AI-assisted publishing is that often, polished language can conceal factual inaccuracies. Gen-AI can create fictious data that appear to be true evidence, such as real academic sources, functioning judicial cases, or viable datasets. Májovský *et al.* (2023) examine this ability by conducting an experiment where ChatGPT was asked to write an entire article about a topic in neurosurgery. Authors find that the model produces a coherent, multi-paragraph text with a conventional journal structure. Nevertheless, the paper contains multiple factual deficient details and references. Similarly, Chelli *et al.* (2024) in their comparison of ChatGPT-3.5, ChatGPT-4, and Bard (2024) employ a systematic approach by asking the models to produce references to existing research papers and then

cross-checking them with databases to see whether these papers actually exist. Their findings confirm that all three models tend to generate non-existent or hallucinated references. The authors find that while hallucination remains a serious problem for reference generation, newer and more sophisticated generative models have fewer issues performing this task. Alkaissi and McFarlane (2023) analyse how references to real scientific articles could be interspersed with fictional ones in ChatGPT responses. The authors conclude that it is relatively easy for non-experts to mistake fake references for real ones. While this article lacks the methodological rigour, it exposes a critical weakness in the approach of relying on common sense to spot AI-generated text. Finally, Sallam (2023) conducts a literature review of 60 published works that discuss ChatGPT's applications, risks, and dangers in healthcare. Out of these 60 works, 58 identify risks posed by using ChatGPT for clinical purposes. The risks include providing users with false or biased information, plagiarising source material, violating copyright laws, breaching patient confidentiality, and failing to meet legal and regulatory requirements. The limitations of this review are imposed by its nature as a literature review, implying that it provides evidence of ChatGPT's risks primarily through other people's perspectives. Magesh et al (2025) evaluate three AI platforms designed to help legal professionals find relevant case law and legislation. Using standard tests and questions, the authors find that in 17-33 percent of cases, the three tested platforms; Lexis+ AI, Westlaw AI-Assisted Research, and Ask Practical Law AI, provided incorrect or unsubstantiated answers. This study provides hard evidence concerning the practical application of generative AI models in a professional law setting. However, this study primarily examines the danger of individual mistakes without describing how these errors propagate through the system.

Authorship, Disclosure and Intellectual Ownership

Bozkurt (2024) employed a critical-conceptual methodology to explore co-creation, authorship, academic integrity, and ownership. The paper explains how GenAI undermines the conventional distinction between author and tool because it produces significant textual elements that cannot be fully explained or admitted by the creator. The study suggests transparency in disclosure and the need to maintain human responsibility but does not empirically identify the threshold of AI-assisted inputs that require acknowledgment. Chetwynd's (2024) applied ethical commentary methodology that concentrates on scientific writing and explores GenAI's role as an ethical writing assistant. The study asserts that Gen-AI can be used to increase efficiency, improve grammar, and aid organization, but the human writer must retain responsibility for veracity, attribution, confidentiality, and the final product. Al-Kfairy *et al.* (2024) utilized an interdisciplinary survey methodology to investigate GenAI's impact on ethics. The work comprehensively reviewed the existing literature on GenAI's privacy, security, bias, misinformation, intellectual property, and responsibility. The paper identifies intellectual property risks posed by GenAI at different stages, including using copyrighted material in training sets, generating texts with copyrighted elements, and the ambiguity of ownership of AI generated content. The work recommends additional regulation, greater transparency, better data governance, and human oversight but focuses primarily on theoretical recommendations. Resnik and Hosseini (2025) applied normative ethical analysis to examine the implications of AI use in science. The paper maintained that ethical principles of truthfulness, objectivity, transparency, responsibility, and respect for privacy still apply to AI use but require additional clarifications.

Limitations and Inequities of AI-Text Detection

Scholarly evidence consistently indicates that detection-based governance mechanisms are inadequate for reliably ascertaining the involvement of GenAI in document creation. Weber-Wulff *et al.* (2023) conducted a comprehensive evaluation of 14 AI-text detection systems comprising 12 publicly available tools and two commercial solutions using a diverse corpus of human-written, AI-generated, translated, and obfuscated texts. Through rigorous accuracy assessment and error-type classification, the study demonstrated that these systems lack both the precision and consistency required for high-stakes academic adjudications. Similarly, Liang *et al.* (2023) examined the efficacy of widely deployed GPT detectors across writing samples produced by native and non-native English speakers. Their analysis revealed a systematic misclassification bias. A significant portion of authentic texts authored by non-native speakers were erroneously flagged as AI-generated, whereas detection accuracy improved markedly on native English compositions. This disparity was attributed to reduced linguistic perplexity and more predictable lexical patterns. Furthermore, Perkins *et al.* (2024) assessed six prominent AI detection tools using a dataset of 805 samples subjected to basic adversarial modifications, such as syntactic restructuring, stylistic variation, and alterations in linguistic complexity. The results indicated a substantial decline in detection efficacy an average reduction of 17.4% following minimal textual manipulation. The authors concluded that current detection frameworks exhibit significant susceptibility to minor adversarial interventions and therefore cannot serve as definitive indicators of academic misconduct.

Bias, Confidentiality and Privacy

Privacy and confidentiality are commonly cited as potential risks in the use of generative artificial intelligence (GenAI), yet empirical examination of these concerns remains limited. Sallam (2023), in a systematic review, reports recurring apprehensions regarding users inputting personal or clinically sensitive information into externally managed platforms. Parallel findings by Al-Kfairy *et al.* (2024) highlight data leakage, unauthorized secondary use of information, and ambiguities related

to model training protocols as significant vulnerabilities. Within academic contexts, such risks may encompass exposure of unpublished research findings, verbatim interview records, assessment instruments, and data permitting participant identification (Weber-Wulff *et al.*, 2023). In professional domains, they may extend to legally privileged communications, patient health records, proprietary business strategies, or classified governmental data (Magesh et al ., 2025; Faheem, 2025). Empirical data are scarce on the types of information professionals disclose within GenAI interfaces. Furthermore, the literature frequently fails to distinguish between distinct system architectures such as public consumer chatbots, enterprise-grade platforms, locally deployed models, and retrieval-augmented systems despite their differing implications for data confidentiality. In contrast, bias has been subjected to more rigorous empirical scrutiny, particularly at the level of model behavior and detection mechanisms (Liang *et al.*, 2023).

Over-reliance and the De-skilling of Human Writers

The potential for repeated use of artificial intelligence to erode human competencies in writing, reasoning, and information verification has become a subject of growing scholarly attention, though empirical support remains limited and preliminary. Zhai *et al.* (2024) undertook a systematic review guided by PRISMA protocols, analyzing 14 studies sourced from ProQuest, IEEE Xplore, ScienceDirect, and Web of Science. The study investigated the cognitive and behavioral consequences of dependence on AI-driven dialogue systems, particularly with respect to decision-making, critical thinking, and analytical reasoning. Findings indicated that uncritical reliance on AI outputs may predispose users to accept generated content without independent verification. Furthermore, excessive dependence was associated with risks such as exposure to AI-generated hallucinations, algorithmic bias, privacy vulnerabilities, and unwarranted trust in system outputs. However, the small sample of 14studies comprised largely conceptual analyses, self-reported user perceptions, and short-term assessments within educational contexts. Longitudinal data demonstrating a causal relationship between sustained Gen-AI usage and de-skilling in writing proficiency remain insufficient.

Conflicting and Convergent Findings

A central issue in the existing literature concerns the distinction between performance and underlying capability. Empirical evidence indicates that Gen- AI can significantly enhance task efficiency and output quality in specific professional domains. Noy and Zhang (2023) report a 40% reduction in task completion time and an 18% improvement in evaluated quality for professional writing tasks using ChatGPT. Similarly, Brynjolfsson *et al.* (2025) observe a 15% increase in productivity within customer support roles, with the most pronounced gains occurring among lower-skilled workers, suggesting a potential equalisation of skill disparities through AI augmentation. However, these benefits are neither uniform nor invariant across skill levels. For highly skilled individuals, Brynjolfsson *et al.* (2025) document marginal performance gains accompanied by slight declines in quality, indicating a possible erosion of expert judgment due to reliance on standardized AI-generated outputs. In educational contexts, analogous trade-offs emerge: Bastani *et al.* (2025) find that unrestricted access to GPT models enhances practice performance but leads to a 17% decline in examination outcomes. Similarly, Li *et al.* (2025) report elevated performance and self-efficacy in programming courses employing AI assistance yet identify diminished knowledge transfer when cognitive offloading reaches excessive levels. These findings underscore those apparent contradictions in the literature. However, there is broad agreement on the necessity of human oversight, proper source attribution, and context-sensitive regulatory frameworks. While scholarly consensus affirms the utility of Gen-AI in editing, translation, ideation, and routine drafting, persistent concerns regarding hallucinations, algorithmic bias, privacy breaches, ambiguous authorship, and unresolved intellectual property issues still remain unresolved (Chetwynd, 2024; Khalifa & Albadawy, 2024; Sharma, 2025).

Critical Gaps in Literature and Study Justification

This critical literature review unravels a number of gaps. First, scholarly attention has been disproportionately directed toward higher education and academic integrity, with institutional policies predominantly focusing on plagiarism, student evaluation, and detection mechanisms (UNESCO, 2023). In contrast, professional and industrial domains such as law, consulting, journalism, engineering, corporate reporting, and public administration remain underexamined regarding the appropriate integration of GenAI in these fields. Second, a clear, operational distinction between AI-assisted support and full substitution remains absent. While current literature upholds the principle of human accountability, it rarely delineates how specific function, including grammar correction, translation, summarisation, drafting, analytical reasoning, and evidence production should be categorised according to their associated risks (Bastani *et al.*, 2025; Li *et al.*, 2025).

Third, prevailing policies tend to be principle-driven rather than procedural. Concepts such as transparency, responsibility, and human oversight are frequently asserted, yet there is limited specification regarding the nature and extent of required disclosures, the types of content necessitating independent verification, the responsible parties for conducting such validation, or the methods for documenting human involvement (Magesh *et al.*, 2025).

Fourth, empirical support for the efficacy of disclosure practices is sparse. Although disclosure is commonly advocated, insufficient research evaluates whether audiences comprehend statements about AI use, whether such disclosures influence trust, or whether they effectively

deter misrepresentation. Disclosure may constitute a necessary safeguard, but it is unlikely to mitigate the risks when the substantive content is flawed or misleading. Fifth, accountability is typically assigned broadly to "the human" actor without sufficient differentiation among stakeholders such as authors, supervisors, employers, editors, publishers, institutional authorities, and technology providers (Sharma, 2025). Fifth, concerns related to confidentiality and intellectual property remain largely theoretical. There is minimal cross-sectoral data on the types of sensitive information users input into GenAI systems, the effectiveness of organisational data governance measures, or the real-world performance of licensing and attribution frameworks. Sixth, policies reliant on detection technologies suffer from weak empirical foundations and pose risks of bias. AI-generated text can be readily modified to circumvent detection algorithms, while writing by non-native English speakers may be erroneously identified as machine-produced a concern particularly common in African and other multilingual contexts (Weber-Wulff *et al.* ,2023). Findings derived from specific iterations of tools such as ChatGPT, Bard, or proprietary legal applications may quickly become outdated following software updates. Consequently, governance frameworks should prioritise enduring principles and risk-based classifications over transient, model-specific performance metrics. Finally, voices from the Global South including African universities, regulatory bodies, publishers, and professional associations are arguably underrepresented in current Gen-AI discourse. Variations in digital literacy, regulatory infrastructure, linguistic diversity, and institutional capacity may significantly shape both the opportunities and challenges associated with GenAI-supported writing, necessitating more inclusive and context-sensitive policy development.

Challenges and Opportunities Identified in theLiterature

Across studies, the principal opportunities are faster drafting, improved language clarity, personalised feedback, translation, accessibility, knowledge retrieval and the diffusion of expert practices to less experienced writers. The major challenges are fabricated information and citations, plagiarism and false authorship, cognitive offloading, embedded bias, unequal access, confidential-data leakage, copyright uncertainty, diminished professional voice and unclear accountability (Bozkurt, 2024). The decisive variable is therefore not AI use alone, but the interaction among task risk, user expertise, system design, disclosure, verification and institutional controls. An analysis that is well-balanced must begin with an admission of real benefits of artificial intelligence. Gen-AI reduces the time needed to create routine messages, creates alternate outlines, helps to compress large volumes of information, makes technical information understandable at different levels, and increases clarity of language (Khalifa and Albadawy, 2024). Language support could make it possible to join global academic and business communication. Nevertheless, these gains in accessibility should not be confused with transferring of authorship. It is ethics to increase and not to replace the work of the human author, who decides what needs to be said or written, what evidence to consider and what to decide on it professionally, academically and industrially.

Cross Sector Case Studies

Academia: Performance gains, Learning Loss and Unfair Detection

This education case provides a reason why output quality does not necessarily mean learning. According to Bastani *et al.* (2025), 988 secondary school students were randomly assigned to a regular GPT interface, a secured GPT tutor, or a control group. Access to AI greatly increased the level of performance in practice sessions; however, without access to AI, students with the unlocked version performed 17% worse than the control group. The secured GPT tutor helped minimize this effect by offering teacher-developed hints instead of answers. This case clearly shows how educational designs can affect the process Gen-AI can be both cognitive scaffolding and thinking replacement. The second problem relates to academic enforcement. In their experiment with GPT detectors, Liang *et al.* (2023) discovered that detectors often considered writing by non-native English speakers as AI-generated. Thus, when applied at universities with linguistically diverse populations, such as universities in Africa, detectors-only decisions can penalize authentic writing and perpetuate linguistic inequality.

Professional Writing-Fabricated legal Authorities in Mata v Avianca

The United States case Mata v. Avianca, Inc. illustrates the consequences of unverified AI-assisted professional writing. Lawyers submitted a court filing containing non-existent cases and fabricated quotations produced through ChatGPT. The court found that the attorneys had abandoned their gatekeeping responsibilities and imposed sanctions (Mata v. Avianca, Inc., 2023). The failure was not merely technical hallucination; it involved professional over-reliance, inadequate source checking and misleading representations to a tribunal. The case supports policies requiring independent verification against authoritative databases, competence training, client-confidentiality safeguards and accountable human sign-off for all consequential documents.

Industry: Customer-support writing at a scale

In their study, Brynjolfsson *et al.* (2025) provide a real-world industry example based on data collected from 5,172 customer support agents. The introduction of a generative AI assistant on a staggered basis led to an overall improvement of 15% in resolution of issues per hour with higher improvements seen in novice and low-skilled workers. Moreover, the implementation improved the English skills as well as multiple aspects of client interaction. Nonetheless, the highest-skilled workers

demonstrated no significant improvement in terms of efficiency and even had a deterioration in terms of quality. Thus, this example shows the benefits and risks associated with the use of Gen-AI dissemination of good practice, but also loss of tacit knowledge as standard procedures replace expertise.

The Integrated Risk Based Framework Generative AI Use

Figure 2 presents the Integrated Risk-Based Framework for Responsible Generative AI Writing (IRF–GenAIW). It establishes a systematic approach to assessing and enforcing the responsible and accountable use of Gen-AI in academic, professional, and industrial writing contexts. It operates on the principle that the ethical implications of GenAI usage are not intrinsic but contingent upon contextual factors, including the intended purpose, the sensitivity of the content involved, the degree of Gen-AI involvement, potential impacts of inaccuracies, levels of stakeholder exposure, and the robustness of human oversight and validation. Consequently, the framework shifts regulatory paradigms from binary classifications such as permitted or prohibited toward a nuanced, risk-proportionate model of governance. It aligns the specific conditions and modalities of AI application with a comprehensive evaluation of multidimensional risks, tailored mitigation strategies and rigorous verification protocols, clearly defined institutional accountability, and mechanisms for iterative refinement of policies over time. Gen-AI use, including communication and writing, is achieved when the degree of disclosure, human oversight, verification, confidentiality safeguards, and accountability corresponds proportionally to the level of AI involvement, the sensitivity of the information processed, and the potential impact of inaccuracies. Consequently, the framework does not assess GenAI integration merely by determining whether AI was employed. Rather, it systematically examines the specific manner in which GenAI was utilized, the risks associated with that application, the protective measures implemented, the individuals responsible for validating the outputs, and the parties who assume liability for resulting outcomes. (UNESCO, 2023; NIST, 2023, Luo, 2024).

The IRF-GenAIW is a proposed synthesis. Its five-stage structure and safeguards are adapted from the following complementary sources:

National Institute of Standards and Technology (NIST, 2023 AI Risk Management Framework (NIST, 2023), European Union Artificial Intelligence Act (European Parliament & Council of the European Union, 2024), UNESCO human-centred GenAI guidance (Miao & Holmes, 2023). The framework extends the considerations beyond education to professional and industrial writing. OECD AI Principles (OECD Principles for trustworthy AI (OECD, 2024). Research-integrity scholarship (Resnik and Hosseini, 2025) Evidence on GenAI writing risks (Chelli *et al.*, 2024; Májovský *et al.*, 2023) and (Magesh *et al.*, 2025; Liang *et al.*, 2023; Perkins *et al.*, 2024; Weber-Wulff *et al.*, 2023; Luo, 2024).

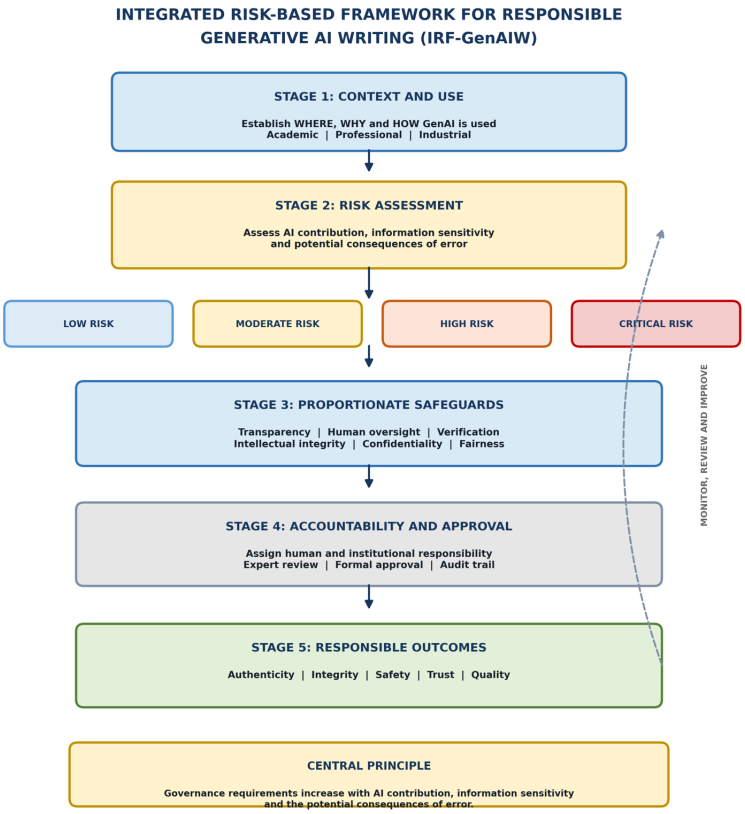

Figure 2: Proposed Integrated Risk Based Framework Generative AI Use

Framework Interpretation

Stage 1 - Context and use: This establishes where, why and how GenAI is being used across academic, professional and industrial writing. The initial component defines the context of generative artificial intelligence (GenAI) use, as ethical norms and implications vary significantly across different domains. The framework is intended to encompass a broad range of professionals and stakeholders including students, researchers, educators, editors, consultants, journalists, legal practitioners, engineers, public servants, corporate personnel, and others engaged in the creation of material with significant written output.

Stage 2 - Risk assessment: Classifies an application into one of four risk levels low, moderate, high, or critical based on the degree of Gen-AI involvement, the sensitivity of the data processed, and the potential impact of inaccuracies. The risk evaluation process initiates with an analysis of the manner and extent to which generative AI has contributed to the creation or modification of the document. Ethical risk tends to escalate progressively as the role of generative AI shifts from providing mechanical support to performing analytical functions or fully supplanting human input.

Stage 3 - Proportionate safeguards: This is essentially a multi-dimensional risk assessment. Each use should be evaluated against several parameters such as but not limited to; purpose and consequence, information sensitivity, intellectual contribution, evidential reliability, affected stakeholders and reversibility of harm. This stage matches the strength of transparency, verification, human oversight, confidentiality and fairness controls to the assessed level of risk.

Stage 4 - Accountability and approval (Core-Governance Principles): Users are required to provide appropriately detailed disclosures regarding the use of generative artificial intelligence (GenAI), specifying the particular system utilized, the rationale for its application, and the segments of the work it influenced. Such disclosures must be sufficiently precise to enable readers, editors, or supervisory authorities to assess the degree of human intellectual contribution. Disclosure is not a replacement for rigorous verification. Merely acknowledging the use of AI does not justify content that is factually erroneous, ethically problematic, or intellectually unsound. All assertions, supporting evidence, quotations, and references must undergo independent validation by the responsible individuals. Users must critically evaluate whether AI-generated outputs perpetuate stereotypes, marginalize underrepresented viewpoints, or systematically disadvantage specific cultural, linguistic, or social communities.

Stage 5 - Responsible outcomes: Seeks to preserve authenticity, integrity, safety, trust and quality while permitting beneficial and appropriately controlled GenAI use. Thus, the framework is designed to achieve five interconnected outcomes: Authenticity implies that written outputs must demonstrate genuine human intellectual engagement and original contribution whereas integrity means that assertions, citations, and supporting evidence must be accurate, verifiable, traceable, and subject to independent validation. Safety means that documents with significant implications undergo heightened review processes and require formal endorsement by qualified experts. Trust denotes that readers must be clearly informed about the extent of AI involvement and be able to ascertain the responsible individual or institutional authority. Finally, legitimacy entails that academic, professional, and industrial entities can leverage generative artificial intelligence while preserving adherence to established ethical principles and professional norms.

Illustrative Case Validation

Table 2 reports the framework illustrative case validation. Please note that this is not an Empirical proof of the framework but an illustrative validation of the Integrated Risk-Based Framework for Responsible Generative AI Writing (IRF–GenAIW). The rationale for illustrative case validation is that testing the framework helps demonstrate whether its categories produce coherent and proportionate decisions across different writing environments.

Tabel 2: Illustrative Case Validation of IRF–GenAIW Framework

Illustrative case	Likely risk	Required controls
Grammar correction in a university essay	Low	Ordinary author, no transfer of authorship
Gen-AI assisted drafting of a review section	moderate	Disclosure, citation checking, source verification and document human revision
Drafting legal, medical, financial or engineering advise	High	Expert verification, confidentiality assessment, audit trail and formal approve
Fabricated research results or undisclosed Gen-AI substitution	Critical	Prohibition, investigating and corrective procedures
Confidential client, patient or proprietary data entered into unapproved Gen-AI model	Critical	Prohibition unless a legally and institutionally approved secure environment exits

Source: Authors' Elaboration

Discussion

Generative artificial intelligence (Gen-AI) presents substantial potential for enhancing academic, professional, and industrial writing by facilitating processes such as idea generation, translation, summarization, reorganization, and linguistic refinement, while simultaneously reducing the time and cost associated with document production (Dwivedi *et al.*, 2023; Kasneci *et al.*, 2023). Nevertheless, despite the fluency of its outputs, such systems may generate fabricated information, erroneous citations, and unsubstantiated inferences risks that persist even in specialized domains requiring technical or domain-specific expertise (Chelli *et al.*, 2024; Magesh *et al.*, 2025; Májovský *et al.*, 2023). Consequently, human authors are obligated to validate significant assertions and maintain ultimate responsibility for the integrity of the final text. Current institutional policies remain inconsistent and overly concentrated on issues of plagiarism and academic dishonesty, frequently neglecting to differentiate between minimal-risk language support and high-risk delegation of tasks involving research, legal, technical, or public policy content (Luo, 2024). Moreover, automated AI-detection tools exhibit limited reliability, are easily circumvented through paraphrasing, and may introduce bias against non-native English speakers (Liang *et al.*, 2023; Perkins *et al.*, 2024; Weber-Wulff *et al.*, 2023). Critical gaps persist in understanding the implications of AI use in professional and industrial writing, including long-term impacts on skill erosion, data confidentiality, intellectual property rights, institutional liability, and the specific challenges faced by African and other Global South contexts.

To address these shortcomings, an Integrated Risk-Based Framework is proposed, comprising five sequential phases being; contextual analysis and intended use, risk evaluation, implementation of proportionate safeguards, accountability and authorization mechanisms, and the assurance of responsible outcomes. This framework categorizes applications according to risk level low, moderate, high, and critical and escalates governance requirements in accordance with the extent of AI involvement, sensitivity of information, stakeholder vulnerability, and potential for harm. Successful deployment necessitates more than mere disclosure; it demands rigorous source validation, meaningful human engagement, expert scrutiny, secure data management, training in AI literacy, and unambiguous assignment of accountability. Future empirical research should assess the framework's efficacy in improving accuracy, transparency, and trust across universities, regulatory bodies, publishers, industries, and public institutions, while ensuring that innovation is not unduly constrained.

The proposed framework makes a unique contribution by integrating fragmented principles of disclosure, verification, human authorship, confidentiality and accountability into a unified, risk-based governance model for Gen-AI-assisted writing. Unlike uniform policies that treat all Gen-AI use similarly, it evaluates applications according to their context, purpose, degree of human control and potential harm, thereby distinguishing acceptable assistance from substitution, deception and professional misrepresentation.

The framework is currently conceptually validated through its grounding in the reviewed literature and established ethical and governance principles. It has not yet been empirically tested or institutionally validated; therefore, its effectiveness, transferability and operational feasibility require assessment through expert review, pilot implementation and comparative application across different settings.

Its implications are context-specific:

- **Academia:** It can guide institutional policies on disclosure, authorship, assessment integrity, research verification and responsible student use.
- **Professional practice:** It can support ethical standards for AI-assisted documentation, advisory work and professional communication while preserving human responsibility.
- **Industry:** It can inform organisational risk classification, confidentiality safeguards, approval procedures, audit trails and accountability for AI-generated content.

Accordingly, the framework should be interpreted as a theoretically grounded governance model that offers testable propositions and practical guidance, rather than as an empirically validated solution.

Policy recommendations for Professional Sectors

Professional organizations and companies should formulate explicit rules regarding the use of AI in addition to existing duties of competence, confidentiality, candor, and care. Consequential writing, such as legal opinions, medical records, audit reports, engineering recommendations, and policy analysis needs to be reviewed and verified by an authorized person according to official sources. Organizations need to keep a list of approved tools, list prohibited data sources, identify situations requiring client/public disclosure, and maintain an audit trail of consequential documents. Education and professional development in fields should take into consideration failures of generative AI, including fabrication of authorities, outdated knowledge, automation bias, and increased confidence. Reviewers need to have enough time and access to the original sources, and approval needs to represent thorough review. Responsibility needs to stay with an identified professional or authorized individual when AI had any significant influence on the decision-making, consistent with COPE's broader position that AI cannot be responsible for human accountability requirements (COPE, 2023).

Policy Recommendations for Industry

Industry use of AI needs to be done through controlled trials with baselines of speed, accuracy, customer impact, rate of errors, and workforce impacts. Different results obtained by Brynjolfsson *et al.* (2025) show that one

productivity benchmark is insufficient because novices gain significantly while expert quality decreases. Thus, businesses should separate assessment of tasks by their type and employee experience, maintain escalation process to subject-matter experts, and watch out if AI-based text homogenizes organizational voice or negatively affects specific groups of customers. Data governance should consider prompt inputs, output storage, vendor access, model training, intellectual property, and incident reporting. Procurement contracts need to include security, privacy, provenance, and auditing standards. Employees need to be encouraged to report use of AI tools to avoid using unauthorized "shadow AI" due to unrealistic performance demands. Workforce management policy should invest in reskilling and allow learning opportunities for junior employees through drafting, review, and mentoring processes. Efficiency gains should be achieved through improved work design rather than used to increase workload or decrease quality controls.

CONCLUSION

Generative artificial intelligence (Gen-AI) use presents significant potential to enhance the efficiency, accessibility, and quality of academic, professional, and industrial writing. However, its integration introduces substantial risks, including the generation of fabricated evidence, citation inaccuracies, lack of transparency in authorship, breaches of confidentiality, violations of intellectual property, embedded biases, professional misrepresentation, and the erosion of human expertise. Current regulatory approaches are often disjointed, focusing predominantly on plagiarism detection while failing to comprehensively address the broader ethical and governance challenges associated with AI-assisted content creation across academia and industry. To overcome these shortcomings, the proposed Integrated Risk-Based Framework for Responsible Generative AI Writing ((IRF–GenAIW) establishes a structured approach that aligns safeguards with the degree and context of GenAI utilization. This includes proportionate measures for substantive human oversight, rigorous verification protocols, protection of sensitive information, and clearly defined accountability mechanisms. The framework operates on the principle that governance rigor must scale according to the extent of AI involvement, the sensitivity of the data processed, the level of stakeholder impact, and the severity of potential errors. Consequently, responsible implementation necessitates the preservation of authentic human authorship and professional discernment, enabling organizations to harness the advantages of generative AI while safeguarding intellectual integrity, operational safety, and public confidence.

Acknowledgment

The authors sincerely thank the anonymous reviewers for their insightful and professional comments and suggestions on the earlier version of this paper. Their comments and suggestions significantly improved the overall quality of this paper. The authors, not the publisher, are liable for all the potential errors and omissions in this paper.

Author Contributions

Dr. Richard Mulenga: Conceptualization, methodology, development of the review protocol and search strategy, literature searching, study screening and selection, data extraction, quality assessment, evidence synthesis, visualization, writing original draft, and project administration. Dr. Mildred Muhyila: Conceptualization, methodology, validation of the search and selection procedures, verification of data extraction and quality assessment, interpretation of findings, supervision, and writing review and editing. Both authors critically reviewed and approved the final manuscript and accept responsibility for the integrity of the work.

REFERENCES

Al-Kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024). Ethical challenges and solutions of generative AI: An interdisciplinary perspective. *Informatics, 11*(3), Article 58. https://doi.org/10.3390/informatics11030058

American Psychological Association. (2024). *The promise and perils of using AI for research and writing.* https://www.apa.org/topics/artificial-intelligence-machine-learning/ai-research-writing

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1).* National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), Article e2422633122. https://doi.org/10.1073/pnas.2422633122

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

Brynjolfsson, E., Li, D., & Raymond, L. R. (2025). Generative AI at work. *The Quarterly Journal of Economics, 140*(2), 889–942. https://doi.org/10.1093/qje/qjae044

Chelli, M., Descamps, J., Lavoué, V., Trojani, C., Azar, M., Deckert, M., Raynier, J.-L., Clowez, G., Boileau, P., & Ruetsch-Chelli, C. (2024). Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: Comparative analysis. *Journal of Medical Internet Research, 26*, Article e53164. https://doi.org/10.2196/53164

Chetwynd, E. (2024). Ethical use of artificial intelligence for scientific writing: Current trends. *Journal of Human Lactation, 40*(2), 211–215. https://doi.

org/10.1177/08903344241235160

Committee on Publication Ethics. (2023). *Authorship and AI tools: COPE position statement*. https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., . . . Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Eke, D. O. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity? *Journal of Responsible Technology, 13*, Article 100060. https://doi.org/10.1016/j.jrt.2023.100060

European Parliament, & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Faheemuddin, S. (2025). Exploring the role of artificial intelligence in predicting property value trends: A systematic review of machine learning applications in real estate pricing and risk assessment. *American Journal of Multidisciplinary Research and Innovation, 4*(5), 10–21. https://doi.org/10.54536/ajmri.v4i5.4815

Hidayatullah, M. H., Suryati, N., Cahyono, B. Y., & Mawaddah, N. (2025). A systematic literature review of artificial intelligence in academic writing: Challenges and opportunities. *Journal of Research on English and Language Learning, 6*(2). https://doi.org/10.33474/j-reall.v6i2.23821

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update, 5*, Article 100145. https://doi.org/10.1016/j.cmpbup.2024.100145

Li, S., Liu, J., & Dong, Q. (2025). Generative artificial intelligence-supported programming education: Effects on learning performance, self-efficacy and processes. *Australasian Journal of Educational Technology, 41*(3), 1–25. https://doi.org/10.14742/ajet.9932

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns, 4*(7), Article 100779. https://doi.org/10.1016/j.patter.2023.100779

Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the "originality" of students' work. *Assessment & Evaluation in Higher Education, 49*(5), 651–664. https://doi.org/10.1080/02602938.2024.2309963

Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. D., & Ho, D. E. (2025). Hallucination-free? Assessing the reliability of leading AI legal research tools. *Journal of Empirical Legal Studies, 22*(2), 216–242. https://doi.org/10.1111/jels.12413

Májovský, M., Černý, M., Kasal, M., Komarc, M., & Netuka, D. (2023). Artificial intelligence can generate fraudulent but authentic-looking scientific medical articles: Pandora's box has been opened. *Journal of Medical Internet Research, 25*, Article e46924. https://doi.org/10.2196/46924

Mata v. Avianca, Inc., 678 F. Supp. 3d 443 (S.D.N.Y. 2023). https://law.justia.com/cases/federal/district-courts/new-york/nysdce/1:2022cv01461/575368/54/

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1)*. U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

Organisation for Economic Co-operation and Development. (2024). *Recommendation of the Council on artificial intelligence (OECD/LEGAL/0449)*. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Perkins, M., Roe, J., Vu, B. H., Postma, D., Hickerson, D., McGaughran, J., & Khuat, H. Q. (2024). Simple techniques to bypass GenAI text detectors: Implications for inclusive education. *International Journal of Educational Technology in Higher Education, 21*, Article 53. https://doi.org/10.1186/s41239-024-00487-w

Resnik, D. B., & Hosseini, M. (2025). The ethics of using artificial intelligence in scientific research: New guidance needed for a new tool. *AI and Ethics, 5*, 1499–1510. https://doi.org/10.1007/s43681-024-00493-8

Sharma, M. (2025). The ethical considerations of

using GenAI and AI tools in academic writing in higher education: A systematic review. *Kalika Journal of Multidisciplinary Research, 3*(3). https://doi.org/10.3126/kjmr.v3i3.87215

Thawbaan, A., Dusserre, G., Yusuf, M. K., & Abdulkareem, Y. (2025). Sustainable cybersecurity in healthcare: An AI-integrated risk and resilience framework. *American Journal of Multidisciplinary Research and Innovation, 4*(4), 144–154. https://doi.org/10.54536/ajmri.v4i4.5086

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity, 19*, Article 26. https://doi.org/10.1007/s40979-023-00146-z

World Intellectual Property Organization. (2024). *Generative AI: Navigating intellectual property*. https://www.wipo.int/documents/d/frontier-technologies/docs-en-pdf-generative-ai-factsheet.pdf