Guarding truth in an AI world: ethics for the next generation of scholarship

Farooq Mubarak
Department of Management and Entrepreneurship, Information Systems Science Unit, Turku School of Economics, University of Turku, Turku, Finland

Journal of Information, Communication and Ethics in Society

Received 6 September 2025
Revised 24 October 2025
4 January 2026
Accepted 30 January 2026

Abstract

Purpose – In light of the limited published literature addressing the complex ethical, epistemic and systemic implications of artificial intelligence (AI) integration in research, this paper aims to build foundational conceptual frameworks that clarify these challenges. It aims to provide a comprehensive examination of the multifaceted issues involved and to propose principled strategies for responsible AI adoption that preserve scholarly integrity and foster sustainable research practices. It is hoped that this effort can start academic discussions on how AI can be used more responsibly for guarding academic integrity.

Design/methodology/approach – This study adopts a conceptual approach to explore the ethical, epistemic and systemic impacts of AI in scholarly research. Due to the emerging nature of this field, the analysis draws on a selective review of key academic and policy sources. Synthesizing diverse perspectives from research ethics, science and technology studies and critical AI scholarship, the paper identifies core challenges and proposes strategies for responsible and equitable AI adoption.

Findings – While AI tools dramatically enhance literature synthesis, data analysis and academic writing, they simultaneously introduce risks such as misattribution, data fabrication and erosion of trust. AI disrupts traditional notions of authorship and originality, raising urgent questions about accountability. Without careful human oversight, it risks compromising the reliability of research. Institutions must update policies to promote ethical use and ensure fair access to AI tools.

Originality/value – This paper offers a timely and comprehensive synthesis of the multifaceted ethical, epistemic and systemic challenges posed by AI integration in scholarly research. The study bridges technological innovation with foundational academic values and charts a clear path for responsible AI adoption, emphasizing human oversight, transparency and equitable access. Unlike prior studies that focus narrowly on technical capabilities or isolated ethical concerns, this work situates AI within the broader cultural and institutional ecosystems, providing actionable insights for researchers, policymakers and academic leaders committed to safeguarding the integrity and future of scholarship.

Keywords Artificial intelligence, Ethics, Research ethics, Information ethics, Scholarly integrity, Authorship, Artificial agents, Epistemic responsibility, Academic integrity, Knowledge production, Transparency, Academic policy, Authorship ethics, AI and authorship ethics

Paper type Conceptual paper

1. Introduction

1.1 Contextual background

The rapid emergence of generative artificial intelligence (AI) has begun to transform the landscape of academic research in profound and complex ways. Advanced language models, such as GPT-4, possess the capacity to generate highly coherent, contextually relevant and

Journal of Information, Communication and Ethics in Society
Emerald Publishing Limited
1477-996X
DOI 10.1108/JICES-09-2025-0241

stylistically polished text, allowing researchers to expedite tasks such as literature reviews, preliminary manuscript drafting and synthesis of complex data sets (Floridi and Chiriatti, 2020; van Dis *et al.*, 2023). At first glance, these capabilities promise remarkable gains in efficiency, potentially reducing barriers associated with language proficiency, workload and time constraints. By automating routine tasks, AI tools can theoretically free scholars to focus on higher-order intellectual work, including critical analysis, theoretical development and methodological innovation.

However, the incorporation of AI into research introduces a series of ethical, epistemic and practical dilemmas. Unlike human authors, AI lacks consciousness, reflective understanding and moral judgment. Its outputs, although linguistically fluent, are generated through pattern recognition across vast corpora rather than deliberate reasoning or conceptual innovation (Bender *et al.*, 2021). Consequently, questions arise regarding authorship, accountability and the integrity of scholarly communication. When human researchers rely extensively on AI without disclosure, the origin of intellectual contribution becomes obscured, raising concerns about transparency, credibility and the ethical representation of knowledge (Thorp, 2023).

1.2 Ethical and epistemic challenges

The use of AI in academic writing challenges long-standing ethical norms regarding authorship and intellectual responsibility. Authorship traditionally implies active cognitive engagement, the generation of original insights and accountability for methodological rigor (ICMJE, 2023). By contrast, AI operates without intentionality, generating outputs without awareness of content validity or ethical implications. This discrepancy poses an ethical tension: when AI substantially contributes to a manuscript, to what extent can the human researcher claim full authorship? The absence of clear disclosure may inadvertently constitute misrepresentation, eroding the trust upon which scholarly communication depends. Recent analyses therefore emphasize the need for explicit disclosure policies and clearer authorship frameworks to safeguard academic integrity and editorial accountability when AI tools are used (Cheng *et al.*, 2025; Resnik and Hosseini, 2025).

Furthermore, AI-generated content raises epistemic concerns related to the nature of originality and the reliability of knowledge. Text produced by AI is inherently derivative, synthesizing existing sources rather than creating genuinely novel ideas (Bender *et al.*, 2021). While stylistically coherent, such outputs may inadvertently introduce inaccuracies, fabricated references or misinterpreted data. Without meticulous human oversight, these errors can propagate through the literature, compromising the cumulative integrity of scholarly knowledge (Ji *et al.*, 2023). Therefore, ethical AI integration necessitates not only transparency in authorship but also rigorous verification and critical evaluation of outputs. Emerging research on generative model hallucination highlights these epistemic risks and calls for systematic verification protocols and transparent reporting of AI involvement in scholarly production (Kay *et al.*, 2024; Tang *et al.*, 2024).

While the ethical and epistemic implications of AI vary with the type of system in question, these issues are further shaped and amplified by the institutional incentives and cultural norms that govern scholarly work. Productivity imperatives, metrics-based evaluation and the rapid uptake of new technologies all interact with AI adoption, altering not merely the frequency of AI use but also the manner and norms of its deployment in research (Arar *et al.*, 2025; Qadhi *et al.*, 2024; Dabis and Csáki, 2024). Accordingly, a typological framework of AI systems is advanced in Section 1.3 to clarify how differences in technological design give rise to distinct ethical challenges, each shaped by the institutional and disciplinary contexts within which research is conducted.

1.3 Artificial intelligence typology and ethical distinctions

Building on the ethical reflections in Section 1.2, it becomes necessary to recognize that "artificial intelligence" in academic contexts is not a singular entity but a constellation of distinct technological forms. Each class of AI system embodies different capacities, risks and epistemic assumptions, making a typological approach essential for meaningful ethical analysis (Bender *et al.*, 2021). This distinction helps to avoid overgeneralization and to specify how varied AI tools reshape the norms of scholarly authorship, reasoning and accountability.

Generative AI systems, such as ChatGPT, Gemini or Claude, are primarily designed to produce synthetic text or code that resembles human output (Floridi and Chiriatti, 2020). While such tools can enhance productivity and linguistic fluency, they raise profound questions about authorship integrity, intellectual ownership and the authenticity of scholarly voice (Stokel-Walker and Van Noorden, 2023). Their ability to simulate coherent argumentation without genuine understanding or verification introduces subtle risks of misinformation, conceptual distortion and citation fabrication.

Recent evidence indicates that unchecked reliance on generative AI may undermine the cognitive dimensions of scholarly work by promoting surface-level synthesis over deeper analytical engagement. Rather than fostering reflective reasoning or methodological innovation, excessive dependence on these tools can encourage efficiency-driven, low-effort outputs that compromise intellectual rigor (Gonsalves, 2024; Lee *et al.*, 2025; Oğuz and Alsancak-Akın, 2024). This cognitive displacement effect – where researchers outsource evaluative or interpretive judgment to AI systems – raises profound questions about the cultivation of scholarly expertise and the integrity of academic authorship. Safeguarding the epistemic quality of research, therefore, requires not only regulatory oversight but also a reaffirmation of the human role in interpretation, critique and creativity.

By contrast, AI tools for data analysis, including machine learning and deep learning models, do not generate text; but shape empirical reasoning by identifying statistical relationships and predictive patterns (Obermeyer *et al.*, 2019). The ethical concerns here revolve around transparency, reproducibility and bias in training data. The opacity of these algorithms, often referred to as the "black box" problem, complicates peer verification and raises questions about methodological accountability (Lipton, 2018). As newer analytical systems automate decision-making in data-intensive disciplines, scholars emphasize the need for explainable AI to preserve interpretive rigor and epistemic trust (Paul and Weinberger, 2024).

A third and increasingly influential category consists of AI research assistants – systems such as *Scite*, *Elicit* and *Consensus* – that retrieve, organize or summarize scholarly literature. While these tools streamline information management, they may also narrow the epistemic landscape by amplifying already dominant voices and marginalizing alternative perspectives. Research on scholarly recommender systems and algorithmic curation shows that popularity and visibility biases are endemic to automated discovery tools, which tend to privilege well-established sources and thereby limit exposure to peripheral or emerging scholarship (Färber *et al.*, 2023; Klimashevskaia *et al.*, 2024). Because many ranking algorithms rely on metrics such as citation counts and prior visibility, such systems risk reinforcing entrenched hierarchies in knowledge production and constraining intellectual diversity. As scholars increasingly depend on these platforms to determine relevance and credibility, the epistemic architecture of research itself becomes partially algorithmic, raising important questions about autonomy, inclusivity and the preservation of critical heterodoxy in academic inquiry (Renkema and Tursunbayeva, 2024).

In sum, recognizing the heterogeneity of AI systems enables a more precise and ethically grounded engagement with technology in academia. Each category (generative, analytical and assistive) poses distinct epistemic vulnerabilities and governance needs. A differentiated ethical framework is therefore essential not only for responsible use but also for preserving the diversity, creativity and integrity that underpin scholarly inquiry.

These typological distinctions reveal that the ethical challenges posed by AI are not intrinsic to technology alone but emerge through the social and institutional contexts in which these tools are embedded. The following section examines how systemic pressures in academia, such as publication expectations, resource inequalities and inconsistent policy responses; influence researchers' motivations and capacities to use AI responsibly. Recent scholarship highlights that research integrity concerns in the age of AI stem not only from the tools themselves but from the institutional and cultural systems that govern their use. When organizational incentives emphasize productivity over reflection, or when policies fail to articulate ethical standards, AI integration can inadvertently weaken norms of transparency and accountability (Chen *et al.*, 2024; Bjelobaba *et al.*, 2025).

1.4 Systemic pressures and research culture

Modern academic environments exert significant pressure on researchers, often prioritizing publication volume, citation metrics and institutional recognition over methodological rigor or ethical conduct (Yeh *et al.*, 2024). Within this context, the availability of AI tools may create incentives to prioritize output speed over reflective scholarship, particularly for early-career researchers or those operating under resource constraints. Disparities in access to AI technologies further exacerbate inequities. Scholars with robust institutional support can use AI responsibly, while those in under-resourced settings may struggle to meet publication expectations without resorting to ethically questionable shortcuts (Hvistendahl, 2023).

Academic culture also plays a critical role in shaping AI adoption practices. Without clear disciplinary norms and institutional guidelines, researchers may adopt AI inconsistently, varying in disclosure practices, authorship attribution and oversight rigor. Addressing these cultural and systemic factors is essential to fostering a research environment that promotes responsible, transparent and ethical AI use. Educational programs, professional development initiatives and institutional policies can help bridge knowledge gaps and standardize expectations for AI integration.

1.5 Motivation for the study

Currently, the scholarship on the topic is very limited; nevertheless, it is important to build the conceptual foundations for future scholarly work. Despite the challenges outlined in the preceding sections, AI offers substantial opportunities to enhance scholarly work. It can accelerate labor-intensive processes, support multilingual scholars and facilitate the analysis of large, complex data sets. For example, AI can highlight emerging trends in literature, suggest relevant references or propose analytical frameworks that human researchers may overlook. When harnessed responsibly, these capabilities can augment human reasoning rather than replace it, creating space for deeper critical engagement and innovation.

Recognizing the dual nature of AI – as a tool for productivity and a potential source of misconduct – underscores the need for a comprehensive investigation. This manuscript aims to examine the ethical, philosophical and practical dimensions of AI in research, exploring its implications for authorship, scholarly integrity, systemic equity and institutional governance. By doing so, it provides a framework for understanding how AI can be integrated responsibly to enhance research quality without undermining trust or epistemic rigor.

1.6 Structure of the manuscript

The manuscript is organized into eight sections. Section 2 examines the ethical and philosophical implications of AI, with a focus on authorship, originality, accountability and epistemic trust. Section 3 analyzes forms of AI-enabled misconduct, detailing common practices, motivations and consequences. Section 4 addresses risks to the scholarly ecosystem, including systemic, epistemic and societal impacts. Section 5 presents strategies for mitigation and responsible AI use, encompassing policy, technology, education and cultural interventions. Section 6 evaluates policy and institutional implications, emphasizing governance, oversight and equitable access. Section 7 explores implications for research and practice, considering methodological, communicative and professional dimensions. Section 8 concludes, synthesizing insights and providing recommendations for the ethical integration of AI into scholarly research.

Since, the published material on this topic is scarce, the manuscript has been structured with relevant sections according to the aims of building a conceptual understanding. Through this structured approach, the manuscript balances a critical examination of AI's risks with constructive guidance for its responsible use, aiming to inform researchers, institutions and publishers on best practices that uphold scholarly integrity and epistemic trust.

2. Ethical and philosophical implications of artificial intelligence in research

2.1 Challenges to authorship and intellectual responsibility

The integration of generative AI into research workflows has prompted a reevaluation of traditional conceptions of authorship. Historically, authorship has been associated with deliberate intellectual contribution, ethical accountability and the capacity to make informed judgments regarding research content (ICMJE, 2023). AI, however, lacks consciousness, intentionality and moral agency. It produces outputs through statistical pattern recognition rather than reflective reasoning, creating a fundamental tension between perceived contribution and actual intellectual engagement.

This discrepancy raises important ethical questions. If AI is responsible for generating significant portions of a manuscript, yet human authors claim full credit without disclosure, the integrity of the scholarly record is compromised (Thorp, 2023). Such practices obscure accountability and may mislead reviewers and readers about the source of ideas, potentially eroding trust in the research process. Even when human oversight is involved, the absence of clear reporting can blur the line between legitimate assistance and unethical misrepresentation. Establishing guidelines for disclosure, authorship attribution and responsibility is therefore critical to maintaining transparency and preserving epistemic integrity.

Furthermore, authorship ethics intersect with institutional and cultural norms. Researchers operating under competitive pressures, particularly in high-stakes publication environments, may face difficult choices regarding the use of AI. The absence of universally accepted standards creates variability in how AI contributions are reported and understood, emphasizing the need for coherent policy frameworks that align individual practices with broader scholarly norms. By addressing these challenges, institutions can ensure that AI enhances rather than undermines ethical scholarship.

2.2 Applying ethical frameworks to artificial intelligence in research

The ethical evaluation of AI's role in research can be situated within established philosophical frameworks. From a consequentialist perspective, moral assessment depends on the outcomes of AI use**;** its ability to enhance epistemic quality, equity and collective benefit while minimizing harms such as misinformation, bias or erosion of trust in scholarship. A deontological approach, in contrast, emphasizes the intrinsic duties of researchers: honesty, transparency and respect for intellectual ownership, irrespective of the efficiency or outcomes

that AI tools might deliver (Madanchian and Taherdoost, 2025). Virtue ethics extends this reasoning by focusing on moral character and the cultivation of scholarly virtues such as integrity, humility and accountability in engagement with AI systems (Hagendorff, 2022; Hayes *et al.*, 2025).

Together, these frameworks highlight that the ethical integration of AI in research is not merely a procedural or regulatory matter but a reflection of deeper moral commitments. Consequentialism draws attention to the societal and epistemic impacts of AI; deontology reinforces the inviolability of ethical duties; and virtue ethics underscores the personal cultivation of moral character. Considering these perspectives collectively provides a coherent foundation for evaluating responsible AI practices in research, linking moral reasoning with the operational principles of scholarly integrity.

2.3 Originality and the nature of knowledge

A central philosophical concern involves the concept of originality. Traditional scholarship values novel insights, creative reasoning and critical synthesis, reflecting the human capacity for reflective thought (Bender *et al.*, 2021). AI-generated outputs, by contrast, are derivative, synthesizing existing knowledge patterns rather than producing genuine innovation. While these outputs may appear stylistically polished, they do not inherently contribute new conceptual perspectives or theoretical frameworks.

This distinction has practical and ethical consequences. If originality is defined primarily in terms of surface-level coherence, AI-generated content may be misinterpreted as original intellectual work. However, genuine scholarly originality requires intentionality, critical evaluation and creative synthesis – capacities that remain uniquely human. Researchers must therefore engage with AI outputs critically, ensuring that generated content is supplemented with reflective reasoning, methodological rigor and novel interpretation. Ethical research practice involves not merely using AI to streamline writing but integrating it as a tool within a reflective and accountable scholarly process.

Philosophically, AI challenges conventional epistemology by raising questions about the nature of knowledge itself. Knowledge is traditionally conceived as justified true belief, grounded in evidence, reasoning and conscious deliberation. AI outputs, while contextually coherent, lack awareness and cannot independently justify claims or evaluate evidence. Incorporating AI without human verification risks presenting information as knowledge when it lacks epistemic grounding. Scholars must therefore retain interpretive authority, ensuring that AI serves as an instrument rather than a substitute for human judgment.

2.4 Accountability and ethical oversight

Accountability is a cornerstone of ethical research, encompassing responsibility for methodological soundness, data accuracy and the ethical treatment of research subjects. AI complicates these responsibilities. Outputs generated by AI may contain errors, fabricated references or misleading conclusions if not critically evaluated (Ji *et al.*, 2023). Researchers remain ultimately accountable for these outputs, highlighting the ethical imperative for thorough review, verification and documentation.

Ethical oversight must extend beyond individual manuscripts to consider broader systemic factors. Institutional policies, peer review standards and disciplinary norms play a key role in shaping how AI is used and evaluated. Without clear guidelines, researchers may inadvertently engage in practices that compromise integrity, such as overreliance on AI for data interpretation or selective disclosure of AI-generated content. Ethical governance frameworks should address both individual and structural responsibilities, ensuring that AI adoption aligns with professional standards and maintains public trust in scholarly research.

2.5 *Epistemic trust and scholarly reliability*

The credibility of academic knowledge depends on the assumption that research outputs are produced by agents capable of reasoned judgment and ethical reflection. AI-generated content, although coherent and contextually appropriate, lacks evaluative understanding and moral responsibility (van Dis *et al.*, 2023). Overreliance on AI without human oversight can undermine trust in research findings, potentially diminishing confidence both within academic communities and among the broader public.

Maintaining epistemic trust requires deliberate human engagement. Researchers must critically assess AI outputs, validate references and data and ensure that conclusions are logically and empirically justified. Peer reviewers, editors and institutions also play a crucial role by establishing standards for AI use, detecting potential misuse and promoting transparency. Together, these measures help preserve the reliability and credibility of the scholarly record in an era of increasing AI integration.

2.6 *Equity, access and global considerations*

Ethical considerations extend to equity and access. Not all researchers have equal access to advanced AI tools, computational resources or training opportunities. Disparities in technological infrastructure, financial support and institutional guidance create unequal opportunities for ethical AI integration (Hvistendahl, 2023). Scholars in resource-limited environments may face difficult choices, balancing the need to publish against the risk of ethical compromise.

Global considerations also highlight the responsibility of institutions and publishers to establish equitable practices. Providing access to AI training, transparent guidelines and support resources can help mitigate disparities and foster a more inclusive research ecosystem. By promoting equitable access, institutions can ensure that the benefits of AI (*for instance*, efficiency, accessibility and innovation) are realized across diverse scholarly contexts, while minimizing ethical risks.

2.7 *Synthesis and forward-looking perspective*

The ethical and philosophical implications of AI in research are complex and interdependent. Challenges surrounding authorship, originality, accountability, epistemic trust and equity reveal that the integration of AI is not merely a technical or procedural issue but a profound ethical transformation within the epistemic culture of science. Recognizing this, scholars must approach AI use not only as a matter of compliance but as a site of moral reasoning and philosophical reflection.

Applying ethical frameworks such as deontology, virtue ethics and consequentialism enables a structured understanding of these responsibilities (Floridi, 2021; van Wynsberghe, 2021). A deontological lens underscores adherence to scholarly duty and truthfulness in representation; virtue ethics emphasizes the cultivation of intellectual integrity, humility and practical wisdom; and consequentialism draws attention to the societal and epistemic outcomes of AI-mediated research practices. Together, these frameworks provide a moral architecture for evaluating not just what researchers *can* do with AI, but what they *ought* to do in preserving academic credibility and public trust.

Looking ahead, responsible AI integration demands a shift from instrumental to reflective adoption; where efficiency is balanced by ethical deliberation, and automation is guided by conscious human oversight. Institutions and publishers must embed ethical literacy into research training, develop transparent accountability structures and ensure equitable access to AI tools and guidance. For individual scholars, maintaining authorship integrity, validating AI outputs and exercising epistemic humility are essential components of ethical practice.

Ultimately, the path forward lies in cultivating a research culture where technological innovation and moral responsibility advance together. By grounding AI adoption in ethical reasoning and philosophical reflection, academia can ensure that AI becomes not a substitute for human judgment, but a catalyst for more transparent, inclusive and trustworthy knowledge creation. In the subsequent sections, this manuscript will examine concrete forms of AI-enabled misconduct, systemic risks, mitigation strategies and policy frameworks that operationalize these ethical considerations in practice.

3. Forms of artificial intelligence-enabled misconduct

3.1 Undisclosed artificial intelligence authorship

One of the most significant forms of AI-enabled misconduct is the failure to disclose AI assistance in manuscript preparation. Undisclosed use occurs when researchers rely on AI to generate substantial portions of text, analyses or references without acknowledging its contribution. This practice misrepresents the origin of intellectual work, potentially inflating claims of human authorship and obscuring responsibility for content (Thorp, 2023). In extreme cases, undisclosed AI authorship can amount to deceptive practice, undermining transparency and the integrity of peer review processes.

Undisclosed AI use is particularly problematic because it blurs ethical boundaries. While human oversight may occur, the lack of disclosure prevents reviewers and editors from accurately assessing the role of AI in generating ideas, structuring arguments or interpreting data. This compromises accountability, as errors, misinterpretations or fabricated information may be inadvertently attributed to human authors. Transparent reporting of AI contributions is therefore essential for maintaining scholarly trust and ensuring that evaluative judgments accurately reflect human intellectual labor.

3.2 Fabrication of data and references

AI systems are capable of producing text that appears factual and well-structured but may contain fabricated data or nonexistent references. Generative AI often generates citations that are plausible in format yet do not correspond to actual publications, creating the risk of propagating misinformation (Ji *et al.*, 2023). Similarly, AI-generated statistical summaries or data sets may not be grounded in empirical research, leading to conclusions that are logically consistent but factually incorrect.

The ethical implications of such fabrication are substantial. Misrepresentation of data or references can distort the scientific record, mislead other researchers and compromise the cumulative reliability of knowledge. When fabricated outputs are published and subsequently cited, errors may propagate through citation networks, amplifying the impact of misconduct. Human verification of all AI-generated data, references and analyses is therefore indispensable to prevent the unintentional dissemination of false information.

3.3 Spectrum of opaque human–artificial intelligence collaboration

In practice, many interactions between humans and AI in research are hybrid in nature, involving a blend of human and machine input. A researcher might generate an initial draft with AI and then refine it manually; create references with AI and subsequently replace some with verified sources; or use AI to paraphrase existing work, obscuring the boundaries of originality. These forms of collaboration occupy a spectrum of opacity, where human intent and machine capacity intersect in ways that are ethically and epistemically complex. *These situations constitute what this paper defines as ethical* “Gray Zones”*: ethically ambiguous situations in which the boundaries between acceptable and unacceptable AI-assisted*

practices are unclear, and where responsibility, transparency and accountability may be difficult to determine.

Hybrid human–AI practices pose particular challenges because they can generate outputs that appear superficially legitimate. For instance, AI-assisted paraphrasing may rearticulate scholarly work in ways that evade traditional plagiarism detection, raising concerns about the misappropriation of intellectual contributions (Halupa, 2023). Likewise, manuscripts that combine authentic and fabricated citations may exploit the credibility of genuine references to lend undue authority to fabricated ones. *Another type of Gray Zone arises when a researcher uses AI to propose alternative data analyses but selectively reports only those results that support a preferred conclusion. In this case, AI assists decision-making, but the selective reporting creates an ethically ambiguous situation where responsibility and transparency are unclear.* This intermixing of legitimate and opaque practices complicates editorial scrutiny and challenges conventional safeguards, emphasizing the importance of examining *ethical gray zones* rather than reducing such behavior to simple notions of "cheating."

3.4 Predatory publishing and artificial intelligence misconduct

The rise of predatory journals adds another layer of complexity. Such journals, already criticized for lax or absent peer review, may become willing or unwitting venues for AI-generated manuscripts. Because predatory outlets prioritize revenue over scholarly standards, they often lack key qualities like meaningful peer review, editorial oversight and transparent editorial procedures (Ferris and Winker, 2017). This creates an environment where fraudulent manuscripts can proliferate, further polluting the academic record and confusing readers who struggle to distinguish legitimate from illegitimate scholarship.

Moreover, predatory publishers may themselves adopt AI tools to automate editorial processes, compounding the problem. Automated peer review, combined with AI-generated submissions, risks creating a feedback loop in which machine-produced content is "validated" by machine-driven evaluation, with minimal human oversight.

3.5 Student misconduct and the pipeline problem

Although much of the debate has focused on professional researchers, students are also engaging in AI-enabled misconduct, often as a precursor to behaviors that may carry over into academic careers. Studies suggest that a significant proportion of university students have already experimented with AI tools for writing assignments, with some using them to generate essays wholesale (Cotton *et al.*, 2023). While educational institutions are developing detection tools and honor codes, enforcement remains inconsistent, and many students perceive AI use as a "gray area" (*within the broader* "Gray Zone" *introduced in* Section 3.3) *rather than a clear violation. For example, a student might use AI to generate potential research hypotheses for a term paper and then selectively test or modify only the ones they deem plausible. Because the student is actively engaging with the material, the work is neither fully original nor entirely AI-generated, creating a situation where it is unclear whether academic misconduct has occurred. Another student might employ AI to summarize peer-reviewed articles for a literature review, then paraphrase portions of the summaries without full attribution. While the student is contributing some analysis, the partial reliance on AI introduces ethical ambiguity regarding authorship and proper acknowledgment.*

This "pipeline problem" matters for research integrity because today's students are tomorrow's researchers. If habits of AI-enabled misconduct are normalized at the educational stage, they risk becoming entrenched in professional scholarship. Addressing AI misconduct,

therefore, requires interventions not only in publishing but also in pedagogy and academic training, *including explicit instruction and scenario-based exercises to help students recognize and navigate ethically ambiguous situations involving AI. Such measures can also help students develop a deeper understanding of responsible research practices and the long-term consequences of ethical lapses. Early reflection and awareness among students contribute to shaping a research culture grounded in integrity and accountability.*

3.6 Plagiarism and redundant publication

Another form of AI-enabled misconduct involves plagiarism and redundant publication. Although AI-generated text is often unique in phrasing, it draws heavily on existing sources and patterns within its training data (Bender *et al.*, 2021). Researchers who fail to properly attribute ideas or who repurpose AI-generated content across multiple publications risk violating norms of originality and intellectual honesty. Redundant publication, in which the same AI-generated content is submitted to multiple venues without disclosure, further undermines the integrity of the research ecosystem.

This type of misconduct is particularly insidious because AI outputs may evade conventional plagiarism detection tools due to subtle rewording or structural alterations. Researchers and institutions must therefore remain vigilant, emphasizing the importance of critical oversight, ethical attribution and adherence to publication standards. Educating scholars on the boundaries between AI assistance and unethical reuse is essential for preventing inadvertent or deliberate misconduct.

3.7 Misrepresentation of methodology

AI can also facilitate misrepresentation of research methodology, either intentionally or unintentionally. Researchers may rely on AI-generated descriptions of experimental procedures, statistical analyses or theoretical frameworks without fully understanding or validating the content. This practice can lead to manuscripts that appear methodologically sound but are based on inaccurate or fabricated processes.

The consequences of misrepresented methodology extend beyond individual manuscripts. Such practices can undermine the reproducibility of research, erode confidence in peer review and distort cumulative knowledge within a field. Human researchers must exercise critical evaluation and retain full comprehension of methodological choices, ensuring that AI outputs are used as tools rather than substitutes for intellectual engagement. Ethical oversight mechanisms should address both the accuracy of methodological reporting and the researcher's responsibility for all AI-assisted content.

3.8 Ethical reflection and intentionality

A common thread across these forms of misconduct is the absence of intentional ethical reflection. Unlike human authors, AI does not possess awareness of ethical norms, and its outputs cannot be judged as responsible or accountable (Floridi and Chiriatti, 2020). The ethical burden falls entirely on researchers to evaluate, verify and disclose AI contributions, ensuring that outputs adhere to established standards of integrity. Neglecting this responsibility, whether due to oversight, convenience or competitive pressure, constitutes misconduct with potentially serious implications for the scholarly ecosystem.

Promoting intentional ethical reflection requires institutional guidance, training and policy support. Researchers must cultivate an awareness of the boundaries between AI assistance and ethical authorship, integrating reflective practices into all stages of research. By fostering a culture of accountability, the academic community can mitigate the risks associated with AI-enabled misconduct while harnessing the benefits of technological innovation.

4. Risks to the scholarly ecosystem

4.1 Peer review under pressure: Challenges to scholarly gatekeeping

The adoption of AI in research presents tangible risks to the integrity of the peer review system. Peer review functions as the primary mechanism through which scholarly quality, rigor and credibility are evaluated. When manuscripts incorporate AI-generated text or analyses without disclosure, reviewers may unknowingly assess content that lacks authentic human reasoning or oversight (Thorp, 2023). This undermines the evaluative process, as judgments are based on text that may not reflect true intellectual engagement or methodological competence.

Moreover, AI can produce text that mimics conventional academic discourse with high fluency, potentially masking errors or misinterpretations. Reviewers may focus on stylistic clarity rather than scrutinizing methodological rigor or data validity, leading to the inadvertent acceptance of flawed research (van Dis *et al.*, 2023). The subtlety of this issue highlights a systemic vulnerability: the peer review process was not designed to detect outputs generated by nonconscious agents. Without updated review protocols, the scholarly ecosystem risks the proliferation of publications that appear credible but may lack substantive originality or accuracy.

4.2 Proliferation of low-quality research

AI-enabled research, when misused, may contribute to the proliferation of low-quality publications. Researchers can generate large volumes of manuscripts rapidly, emphasizing quantity over quality in competitive academic environments (Yeh *et al.*, 2024). This flood of publications strains editorial and peer review capacity, diluting attention to methodological rigor, novelty and ethical standards. Journals may inadvertently accept AI-generated content that has not undergone thorough verification, further compromising scholarly standards.

The increased volume of publications also complicates literature synthesis and meta-analytical research. When low-quality studies dominate citation networks, subsequent research may rely on flawed evidence, perpetuating inaccuracies and undermining cumulative knowledge. Scholars may spend additional time discerning reliable sources, which diminishes overall research efficiency and slows progress. The risk of contaminating the scientific record highlights the systemic impact of unregulated AI integration on the broader research ecosystem.

4.3 Amplification of inequities

The integration of AI tools into research amplifies existing inequities within the academic ecosystem. Access to advanced AI systems, computational resources and training opportunities is unevenly distributed across institutions, disciplines and geographic regions (Hvistendahl, 2023). Researchers at well-resourced institutions can leverage AI to enhance productivity, refine manuscript quality and navigate publication processes more efficiently. Conversely, scholars in under-resourced settings may lack access to these tools, placing them at a competitive disadvantage.

These disparities extend beyond individual productivity. Institutions with widespread AI access may produce disproportionate research output, influencing funding decisions, reputational rankings and citation metrics. This dynamic risks entrenching structural inequities, where researchers with technological advantages dominate scholarly discourse, while others struggle to maintain visibility and credibility. Addressing these inequities requires deliberate policy interventions, equitable access initiatives and capacity-building programs to ensure that AI enhances rather than exacerbates disparities.

Beyond access-related disparities, it is equally critical to examine inequities embedded within the AI systems themselves. Large language models and other AI tools are typically trained on data sets that disproportionately reflect Western, English-language and commercially

dominant sources (Bender *et al.*, 2021; Sheng *et al.*, 2021; Floridi and Chiriatti, 2020). As a result, these systems can systematically amplify existing biases in scholarly discourse, marginalize non-Western knowledge systems and perpetuate patterns of epistemic exclusion. Such dynamics constitute a form of epistemic injustice, whereby certain perspectives, languages and epistemologies are systematically underrepresented or misrepresented in AI-mediated research outputs (Dotson, 2014; Helm *et al.*, 2024; Van Slyke *et al.*, 2025).

Specifically, this dynamic constitutes a form of digital neocolonialism (for instance, consider Salami, 2024), wherein technological infrastructures and epistemic authority are concentrated within a small number of Western institutions and corporations. The reproduction of such asymmetries not only perpetuates inequitable access to technological resources but also shapes what counts as legitimate knowledge within global research ecosystems. Researchers and institutions, therefore, bear a responsibility to critically assess these embedded biases and implement strategies to mitigate their effects. Recommended approaches include diversifying training data sets to incorporate underrepresented voices, actively evaluating model outputs for cultural and linguistic bias and promoting transparency around the limitations of AI-assisted knowledge generation (Sheng *et al.*, 2021; Floridi and Chiriatti, 2020). By integrating these practices, AI adoption can support more equitable, inclusive and epistemically just research environments.

4.4 Threats to public trust

Public trust in academic research is predicated on the assumption that findings are the product of rigorous human reasoning, ethical conduct and empirical validation. AI-generated research, particularly when undisclosed or misused, threatens this trust. If nonexperts, policymakers, or the general public perceive that scholarly outputs may be produced or substantially influenced by AI without accountability, confidence in the reliability and legitimacy of research may diminish (van Dis *et al.*, 2023).

Such erosion of trust has tangible consequences. Public skepticism can reduce engagement with research findings, hinder the implementation of evidence-based policy and increase susceptibility to misinformation. The reputational risk extends to journals and institutions, which may face criticism for publishing work perceived as lacking authentic scholarly oversight. Protecting public trust necessitates transparent disclosure practices, robust editorial oversight and clear communication regarding the role of AI in research production.

4.5 Impacts on research collaboration

The integration of AI into research teams is not merely a technological shift – it is a potential reshaping of intellectual authority and power within collaborative projects. AI tools can amplify the influence of team members with advanced proficiency, enabling them to disproportionately guide project direction, interpret data and draft manuscripts. These dynamics introduce new ethical and epistemic challenges, as technological skill can overshadow scholarly insight, subtly restructuring the hierarchy of intellectual contributions.

AI adoption is reshaping collaborative practices within the scholarly ecosystem in ways that are both enabling and disruptive. On one hand, AI can enhance coordination, streamline workflow and support complex analyses, particularly in multidisciplinary teams. Automated literature summaries, draft analyses and AI-generated experimental proposals can accelerate progress and reduce routine cognitive burdens. On the other hand, overreliance on AI may inadvertently diminish critical discussion, reflective debate and collective decision-making, undermining the richness of intellectual exchange that characterizes robust collaboration.

Crucially, disparities in AI access, familiarity and proficiency can create power imbalances within research teams. Team members with advanced AI literacy or privileged access may disproportionately shape project design, interpretative frameworks and manuscript composition, while those less proficient or excluded from these tools risk marginalization. This dynamic can produce subtle yet profound shifts in power: AI-literate collaborators may steer the intellectual direction of a project, prioritize certain lines of inquiry and influence authorship and credit allocation. Such shifts raise important ethical and epistemic concerns, including equitable recognition of contributions, fairness in decision-making and the potential entrenchment of hierarchical structures based on technological competence rather than scholarly insight.

These considerations highlight that the implications of AI extend beyond individual usage to the collective culture of research. Institutions and teams must anticipate and mitigate the effects of unequal AI proficiency, promoting transparency in tool usage, inclusive decision-making and equitable recognition of contributions. By attending to these dynamics, scholarly communities can harness AI's benefits without compromising collaborative integrity, intellectual diversity or ethical standards (Zárate-Torres *et al.*, 2025; Asghar *et al.*, 2025).

4.6 Influence on citation and knowledge networks

The integration of AI into research may alter patterns of citation, reference networks and knowledge dissemination. AI-generated text can suggest or generate references based on patterns rather than critical evaluation, potentially promoting sources that are more accessible, stylistically compatible or algorithmically favored rather than epistemically justified (Bender *et al.*, 2021). Over time, such practices may bias citation networks, influence perceived authority of sources and shape the trajectory of disciplinary knowledge.

These systemic effects underscore the importance of careful human oversight and critical evaluation of AI-generated references and recommendations. Scholars, editors and reviewers must remain vigilant to ensure that knowledge propagation remains grounded in rigorous evaluation rather than algorithmic convenience. Failing to do so risks distorting the cumulative understanding of research fields and perpetuating epistemic inaccuracies.

5. Strategies for mitigation and responsible artificial intelligence use

Building on the typological framework presented earlier, it is possible to develop mitigation strategies that are tailored to the distinct ethical and epistemic challenges posed by each AI system. Generative AI, analytical AI models and AI research assistants each present unique risks, ranging from authorship integrity and reproducibility to epistemic inclusivity and citation equity. By aligning mitigation efforts with these differentiated risks, interventions can be made more precise and effective, ensuring that AI adoption supports responsible, transparent and equitable research practices. Section 5.1 offers mitigation strategies related to various AI types.

While the typology of AI systems provides a structured lens for understanding the distinct ethical and epistemic risks associated with different tools, it is equally important to consider mitigation strategies that apply across AI applications more broadly. The following sections from Sections 5.2–5.7 will discuss general approaches to responsible AI use, highlighting practices that foster transparency, accountability and integrity in research regardless of the specific system used. Section 5.8 introduces human-in-the-loop (HITL) and human-on-the-loop (HOTL) frameworks. By situating typology-specific insights within this wider context, the discussion ensures that both targeted and overarching strategies for ethical AI adoption are addressed.

5.1 Mitigation strategies informed by artificial intelligence typology

Generative AI: generative systems, such as large language models, primarily raise concerns related to authorship integrity, originality and the potential for unintentional plagiarism. Mitigation strategies should focus on ensuring transparency and accountability in AI-assisted writing. Researchers are encouraged to explicitly disclose the use of generative AI in manuscripts, maintain active editorial oversight and apply rigorous peer-review standards that can detect and address automated content generation (Qadhi *et al.*, 2024; Arar *et al.*, 2025). Such measures help safeguard the authenticity of scholarly outputs while promoting responsible adoption.

Analytical AI models: AI tools used for data analysis, prediction or modeling pose distinct risks, including bias in training data, limited interpretability and challenges in reproducibility. Strategies to mitigate these risks include the use of explainable AI frameworks, systematic auditing of data sets, adherence to open data and open methods practices and clear documentation of algorithmic assumptions (Dabis and Csáki, 2024; Qadhi *et al.*, 2024). By prioritizing transparency and interpretability, these measures reduce the likelihood of biased conclusions and enhance the credibility of AI-mediated research findings.

AI research assistants: Systems designed to retrieve, filter or summarize literature – such as Scite, Elicit and Consensus – can unintentionally reinforce epistemic hierarchies or limit exposure to diverse scholarly perspectives. Mitigation approaches for this category involve promoting source diversity, providing training on algorithmic limitations, encouraging critical evaluation of recommendations and monitoring citation patterns for potential biases (Arar *et al.*, 2025). Such interventions help maintain epistemic inclusivity and prevent the narrowing of research agendas due to algorithmic influence.

Integrative considerations: Across all AI types, mitigation strategies are most effective when tailored to the specific functions and risks of the system in question while remaining sensitive to disciplinary norms and institutional contexts. Embedding these practices within professional development programs, policy guidelines and research workflows ensures that AI adoption enhances productivity without compromising ethical standards or the epistemic integrity of scholarly work (Arar *et al.*, 2025; Dabis and Csáki, 2024; Qadhi *et al.*, 2024).

5.2 Transparent disclosure of artificial intelligence contributions

A foundational strategy for mitigating the risks associated with AI in research is the transparent disclosure of AI contributions. Researchers should explicitly report the extent and nature of AI involvement in manuscript preparation, data analysis or literature synthesis (Thorp, 2023). Such transparency allows peer reviewers, editors and readers to evaluate the work with a clear understanding of the human and machine contributions, safeguarding accountability and scholarly integrity.

Implementing disclosure practices requires clear guidelines from journals, institutions and professional bodies. Disclosure statements should specify which sections of a manuscript involved AI assistance, the type of AI tool used and the degree of human oversight applied. By establishing consistent reporting standards, the academic community can normalize ethical AI integration and reduce ambiguity surrounding authorship, ensuring that AI is positioned as a supportive tool rather than a surrogate for human reasoning.

5.3 Critical evaluation and verification

Even with disclosure, the responsible use of AI demands rigorous critical evaluation and verification of all outputs. Researchers must assess AI-generated content for accuracy, coherence and methodological soundness, particularly when it pertains to empirical data,

statistical analysis or reference management (Ji *et al.*, 2023). Blind reliance on AI-generated material can perpetuate errors, compromise research validity and undermine epistemic trust.

Verification practices should include cross-checking references, validating statistical analyses and confirming that AI-generated summaries or interpretations align with original source material. Critical evaluation is particularly crucial when AI outputs propose methodological frameworks or theoretical interpretations, as these may appear plausible while lacking genuine empirical or conceptual grounding. By maintaining active oversight, researchers can harness AI's efficiencies while ensuring that intellectual and ethical responsibilities remain intact.

5.4 Education and ethical training

Education and training constitute a key strategy for promoting responsible AI use. Many researchers may lack familiarity with the capabilities, limitations and ethical considerations associated with generative AI (Floridi and Chiriatti, 2020). Structured training programs, workshops and institutional guidance can equip scholars with the knowledge needed to integrate AI ethically, critically and effectively.

Ethical training should emphasize principles of authorship, transparency, accountability and verification. Researchers should learn to differentiate between AI assistance and substantive intellectual contribution, understand the risks of data or reference fabrication and recognize the importance of human oversight. By fostering an ethical mindset alongside technical competence, institutions can cultivate a research culture that balances innovation with integrity.

5.5 Technological safeguards

Technological interventions can also support responsible AI integration. For instance, software tools designed to detect AI-generated content, validate references or identify inconsistencies in data can complement human oversight (van Dis *et al.*, 2023). Integrating these tools into editorial workflows, research management platforms and institutional review processes enhances the capacity to identify potential errors or misuse before publication.

However, technological safeguards are not substitutes for ethical reflection. AI detection tools have limitations, including false positives and false negatives, and may be circumvented by sophisticated manipulation. Therefore, technology should function as a supportive measure, enhancing human judgment rather than replacing it. Combining technological tools with education, transparent disclosure and rigorous verification ensures a multi-layered approach to ethical AI use.

5.6 The artificial intelligence detection paradox

The increasing reliance on AI detection tools to monitor scholarly writing introduces a paradox in the management of AI-assisted research. While these tools are designed to identify unacknowledged AI-generated content, empirical studies demonstrate that their effectiveness is limited. Detection algorithms often produce high false-positive rates and can be circumvented through minor paraphrasing or stylistic adjustments (Erol *et al.*, 2025; Weber-Wulff *et al.*, 2023).

For instance, paraphrasing AI-generated text has been shown to significantly lower the detection capabilities of existing tools. Studies have revealed that paraphrasing can reduce detection accuracy from over 70% to as low as 4.6%, without appreciably modifying the input semantics (Krishna *et al.*, 2023). *Such inconsistencies* risk unjustly penalizing researchers, particularly those writing in a second language or using AI-assisted phrasing tools for accessibility.

Given these limitations, the scholarly community's primary defense should not rest solely on forensic detection. Instead, fostering a culture of transparency, accountability, community norms and explicit disclosure is paramount. Editors and institutions are encouraged to implement policies that require authors to clearly state the extent and nature of AI involvement in manuscript preparation (Akbar, 2025). By prioritizing provenance and cultural norms over purely technological solutions, research communities can more effectively mitigate risks associated with AI use while maintaining trust and fairness in scholarly evaluation.

In sum, a sustainable approach to integrity requires reframing this paradox: the goal should not be perfect detection, but transparent provenance. Policies that mandate disclosure of AI involvement, promote human verification and cultivate ethical literacy offer a more constructive and equitable means of addressing AI-assisted writing. Rather than policing textual output through unreliable algorithms, institutions and publishers should foster a culture of responsibility, openness and reflective judgment that aligns technological innovation with academic integrity.

5.7 Policy frameworks and institutional guidelines

Developing clear policy frameworks and institutional guidelines is essential for consistent and ethical AI adoption. Universities, research institutes and journals should establish standards for AI disclosure, authorship attribution, data verification and methodological reporting. Policies should specify acceptable AI applications, reporting requirements and consequences for unethical use (Yeh *et al.*, 2024).

Institutional guidelines can also foster equity in AI access, ensuring that all researchers, regardless of geographic location or resource availability, have the knowledge and tools necessary for responsible integration. Policies should encourage collaboration, mentorship and capacity-building, reducing disparities and promoting a culture of integrity. By embedding these principles into institutional governance, organizations can guide researchers toward practices that balance innovation, efficiency and ethical responsibility.

5.8 Cultural and normative interventions

Beyond policies and technology, cultural and normative interventions are critical. Academic communities must cultivate norms that value reflective practice, critical evaluation and ethical integrity alongside productivity metrics. Celebrating responsible AI integration, recognizing transparency and emphasizing quality over quantity can shift incentives away from purely output-driven goals (Hvistendahl, 2023).

Peer networks, mentorship programs and professional societies can reinforce ethical norms, providing forums for discussion, sharing best practices and modeling responsible AI use. When ethical reflection and accountability become ingrained in research culture, the likelihood of AI-enabled misconduct diminishes and AI becomes a tool for enhancing, rather than compromising, the scholarly ecosystem.

5.9 Operationalizing human oversight: Human-in-the-loop and human-on-the-loop frameworks

The integration of AI into research necessitates structured human oversight to safeguard ethical, epistemic and methodological standards. Current references to human oversight are often reactive; limited to detecting errors or correcting outputs, yet ethical AI integration requires a proactive and systematic approach (Amershi *et al.*, 2014; Doshi-Velez and Kim, 2017). Two complementary frameworks, HITL and HOTL, provide clear operational models for researchers.

5.9.1 Human-in-the-loop. The HITL model positions the human as the primary driver of research activities, ensuring that AI remains a tool rather than an autonomous agent. HITL consists of three stages:

(1) *Human direction*: The researcher defines the research question, selects the methodology and determines the analytical framework. Ethical and epistemic standards, disciplinary norms and methodological rigor are embedded at this stage, providing a foundation for responsible AI use.

(2) *AI execution*: AI performs the assigned tasks – such as drafting sections, summarizing literature, coding data sets or performing initial analyses – strictly according to the human-defined parameters. AI functions as an augmentative tool, executing instructions but not making autonomous interpretive judgments.

(3) *Human curation and validation*: Researchers critically evaluate AI outputs, contextualize findings, correct errors and take full responsibility for final decisions. This stage ensures interpretive accuracy, mitigates bias and preserves the cognitive and ethical integrity of research. The process also supports transparency, reproducibility and accountability in scholarly outputs.

HITL is particularly suitable for tasks where nuanced human judgment is essential, such as theory development, interpretive analysis or ethical decision-making, allowing AI to enhance productivity without compromising oversight.

5.9.2 Human-on-the-loop. For high-volume, computationally intensive, or partially autonomous AI tasks, the HOTL framework emphasizes supervisory oversight rather than direct intervention (Kandikatla and Radeljic, 2025). HOTL is structured as follows:

(1) *AI execution with supervision*: AI carries out tasks independently within defined parameters, such as large-scale data processing, predictive modeling or automated summarization.

(2) *Human monitoring*: Researchers maintain continuous or periodic oversight, monitoring outputs for alignment with research objectives, ethical standards and methodological integrity. Monitoring can include automated alerts, audits or sampling of outputs to ensure correctness.

(3) *Targeted intervention*: Humans intervene selectively when anomalies, biases or ethical issues are detected, providing corrections or contextual guidance. While AI operates autonomously, ultimate responsibility for interpretation, decision-making and dissemination remains with the human researcher.

HOTL is well-suited for projects where AI handles repetitive or data-intensive tasks, reducing cognitive load while maintaining human accountability and ethical oversight.

5.9.3 Practical implications and integration of human-in-the-loop and human-on-the-loop frameworks. Operationalizing HITL and HOTL frameworks in research settings requires more than individual effort; it demands systemic institutional support, formal training programs and clear governance policies. Researchers must be equipped with a thorough understanding of AI capabilities and limitations, including the interpretive constraints, potential biases in training data and contextual risks that may arise when AI outputs are applied uncritically. Training should emphasize critical engagement, ethical reasoning and the preservation of scholarly judgment, ensuring that researchers can effectively supervise AI outputs without ceding responsibility for interpretation or analysis. By cultivating these competencies, institutions foster a workforce capable of integrating AI tools in ways that advance research productivity while safeguarding intellectual rigor.

Equally important is the development of institutional policies and guidelines that delineate the boundaries of AI use, outline disclosure obligations and clarify accountability structures. Such frameworks should specify which tasks may be delegated to AI and which require direct human oversight, establishing clear standards for both HITL and HOTL approaches. By explicitly separating AI-executed functions from human-driven decision-making, these policies help maintain epistemic integrity and ethical accountability across the research lifecycle. When implemented effectively, HITL and HOTL frameworks do not merely manage risk – they actively enhance the quality, transparency and reliability of scholarly outputs, embedding responsible AI use as a core component of research culture. Together, these frameworks move beyond reactive error-checking toward embedded ethical engagement throughout the research process.

6. Policy and institutional implications

6.1 Governance and oversight

Effective governance is essential for ensuring that AI integration into research maintains ethical and scholarly standards. Institutions, journals and funding agencies play a pivotal role in establishing clear frameworks that define acceptable use, disclosure requirements and accountability measures (Yeh *et al.*, 2024). Without robust oversight, AI adoption can exacerbate misconduct risks, distort research outputs and undermine public trust.

Governance mechanisms may include mandatory reporting of AI use in manuscripts, standardized authorship attribution protocols and integration of AI verification into peer review workflows. Funding agencies can require compliance with ethical AI practices as part of grant conditions, reinforcing institutional and individual accountability. Such oversight must be adaptive, reflecting ongoing technological developments, to ensure that policies remain relevant and effective as AI capabilities evolve.

6.2 Equitable access and capacity building

AI adoption can unintentionally reinforce systemic inequities unless institutions implement measures to ensure equitable access. Researchers in under-resourced regions or institutions may lack access to advanced AI tools, computational infrastructure or training opportunities (Hvistendahl, 2023). Policies should address these disparities by providing resources, training programs and institutional support that enable all researchers to use AI responsibly and effectively.

Capacity-building initiatives might include workshops, mentorship programs and collaborative platforms that foster knowledge sharing about ethical AI use. By democratizing access to AI technologies and promoting ethical literacy, institutions can mitigate inequities, prevent competitive disadvantage and ensure that the benefits of AI integration are distributed fairly across the scholarly ecosystem.

6.3 Editorial standards and journal policies

Academic journals occupy a critical position in shaping norms for AI use. Editorial policies should articulate clear expectations for disclosure, verification of AI-generated content and ethical compliance. Journals can adopt guidelines specifying how AI contributions should be reported, the degree of human oversight required and the types of AI-supported analyses deemed acceptable (Thorp, 2023).

Editors may also implement additional review steps to assess AI-generated content, including cross-referencing citations, verifying data integrity and evaluating methodological soundness. Standardizing these practices across journals reduces ambiguity, fosters consistency and reinforces accountability. By maintaining rigorous editorial standards, journals safeguard the credibility of the published record and establish norms for responsible AI integration.

6.4 *Legal and regulatory considerations*

AI adoption in research raises potential legal and regulatory implications. Intellectual property rights, authorship attribution and data privacy are all areas of concern. Since AI cannot hold copyright, the human user is generally regarded as the author, but failure to disclose AI involvement could lead to disputes over intellectual property or misattribution of work (Bender *et al.*, 2021).

Regulatory bodies may also impose guidelines related to the ethical use of AI, particularly in research involving human subjects, sensitive data or clinical applications. Institutions must remain informed about evolving regulations and integrate compliance measures into research governance frameworks. Legal clarity helps prevent disputes, reinforces ethical norms and ensures that AI is used within established regulatory boundaries.

6.5 *Incentives and research culture*

Generally speaking, policy effectiveness is closely tied to research culture and incentives. Institutions must balance the pressures of productivity metrics with the promotion of responsible, ethical AI use (Yeh *et al.*, 2024). Incentives that reward transparency, methodological rigor and ethical oversight (rather than solely publication volume) can shift behavior toward responsible integration of AI technologies.

Mentorship programs, recognition awards and professional development opportunities can reinforce these values. By embedding ethical expectations into institutional evaluation systems, researchers are encouraged to prioritize quality, integrity and accountability. A culture aligned with these principles ensures that AI enhances scholarly work without compromising ethical or epistemic standards.

6.6 *Global policy coordination*

AI adoption and its associated risks transcend national and institutional boundaries. Global coordination of policies can help harmonize ethical standards, foster collaboration and prevent exploitation or misconduct. International academic organizations, consortia and professional societies can provide guidelines that transcend local regulatory discrepancies, promoting consistent practices in AI integration worldwide (Floridi and Chiriatti, 2020).

Such coordination may include shared training materials, standardized disclosure requirements and collaborative monitoring mechanisms. By aligning policies internationally, the research community can reduce inequities, facilitate cross-border collaboration and ensure that AI supports ethical, high-quality scholarship globally.

7. Implications for research and practice

7.1 *Methodological implications*

The integration of AI into research has profound methodological implications, influencing study design, data analysis and knowledge synthesis. AI can efficiently generate literature reviews, summarize complex data sets and suggest analytical frameworks, potentially enhancing methodological rigor when applied judiciously (Floridi and Chiriatti, 2020). Researchers can leverage these capabilities to identify patterns, generate hypotheses and explore relationships within data that may not be immediately apparent through traditional approaches.

However, methodological risks arise when AI outputs are adopted uncritically. AI-generated analyses or experimental designs may lack contextual nuance, misinterpret variables, or introduce subtle biases, undermining the validity and reliability of research findings (Ji *et al.*, 2023). Ensuring methodological integrity requires that human researchers critically evaluate AI contributions, verify assumptions and maintain oversight over all

stages of study execution. AI should serve as a supplementary tool to human reasoning rather than a replacement for deliberate methodological judgment.

7.2 Impacts on research communication

AI's influence extends to research communication, particularly the preparation and dissemination of manuscripts. AI-generated text can enhance clarity, improve linguistic precision and facilitate multilingual communication, reducing barriers for non-native English speakers and enhancing global scholarly participation (van Dis *et al.*, 2023). Such improvements can increase accessibility and broaden the reach of research findings, contributing to more inclusive knowledge dissemination.

While AI offers clear advantages in enhancing linguistic clarity and accessibility, excessive reliance on such tools simultaneously introduces complex risks for the authenticity and diversity of scholarly expression. One key concern is the potential homogenization of academic prose: repeated use of AI-generated language may erode individual scholarly voice, resulting in a uniform, "bland" writing style across disciplines. Agarwal *et al.* (2025) discuss how AI writing suggestions can lead to a homogenization of writing styles, particularly favoring Western norms. The result is a more uniform and standardized prose style or formulaic style; linguistically polished but often detached from the intellectual and cultural identity of the author.

Additionally, AI-generated text can inadvertently mask conceptual weaknesses or methodological gaps, creating the impression of rigor and coherence where underlying deficiencies remain. Rivero and Yin (2025) highlight concerns about the potential erosion of individual writing styles and the homogenization of academic expression due to AI tool usage. Researchers must therefore balance efficiency gains with the preservation of intellectual authenticity, critical engagement and stylistic diversity, ensuring that AI enhances rather than diminishes the richness of scholarly communication.

7.3 Professional development and skill acquisition

AI integration reshapes the skills required for contemporary research practice. Proficiency with AI tools, understanding of ethical guidelines, and the ability to critically evaluate algorithmically generated outputs are increasingly essential competencies (Yeh *et al.*, 2024). Training programs, workshops and mentorship initiatives should focus on developing these skills, enabling researchers to leverage AI effectively while maintaining ethical and methodological standards.

This shift in professional development has broader implications for career progression, collaboration and interdisciplinary engagement. Researchers adept at combining AI capabilities with human critical reasoning may gain competitive advantages in productivity and publication efficiency, while others risk falling behind if support and training are insufficient. Institutions must ensure equitable access to resources and guidance, fostering inclusive skill development across research communities.

7.4 Implications for interdisciplinary research

AI's capabilities for rapid synthesis and pattern recognition are particularly relevant for interdisciplinary research. Scholars can use AI to integrate findings from multiple fields, identify connections and generate cross-disciplinary insights that would be difficult to achieve manually (Floridi and Chiriatti, 2020). Such applications can accelerate the generation of novel frameworks and collaborative research agendas, promoting innovation across domains.

Nonetheless, interdisciplinary integration via AI carries risks of superficial understanding. AI may produce summaries that overlook domain-specific nuances, leading to misinterpretations or oversimplified conclusions. Human oversight, domain expertise and critical engagement remain essential to ensure that interdisciplinary research maintains depth, accuracy and conceptual rigor.

7.5 Implications for peer review and editorial practice

AI also affects peer review and editorial processes. Reviewers may encounter manuscripts with AI-generated content that is highly polished but lacks underlying rigor or originality (Thorp, 2023). This necessitates the development of new review competencies, including the ability to detect potential AI-generated outputs, assess methodological fidelity and evaluate ethical disclosure practices.

Editorial practices must adapt accordingly. Journals may implement AI screening protocols, require explicit reporting of AI contributions and provide training for reviewers to maintain consistency and fairness in manuscript evaluation. By aligning peer review and editorial standards with the realities of AI integration, the scholarly community can safeguard research quality while fostering innovation in review processes.

7.6 Broader implications for research ethics

Finally, AI integration has wide-ranging implications for research ethics. Ethical responsibility remains primarily with human researchers, encompassing accuracy, accountability, transparency and equity. AI's role as a tool rather than an agent underscores the importance of human judgment at every stage of the research process (Bender *et al.*, 2021).

Ethical frameworks must address disclosure, authorship, verification and the equitable distribution of AI resources. Institutions, journals and funding agencies play a critical role in enforcing ethical standards, providing guidance and fostering a culture of responsible AI use. Ensuring that AI supports rather than undermines ethical research practices is essential for maintaining trust, credibility and the integrity of the scholarly ecosystem.

8. Conclusion

8.1 Reconceptualizing the role of artificial intelligence in research

The integration of AI into scholarly practice represents a profound shift in the research landscape, necessitating a reconceptualization of traditional notions of authorship, originality and methodological engagement (Floridi and Chiriatti, 2020). AI provides researchers with unprecedented capabilities to synthesize literature, analyze complex data sets and generate coherent scholarly text, offering efficiencies that can enhance productivity and inclusivity. However, these benefits are accompanied by significant ethical, epistemic and systemic risks, including misattribution of authorship, fabrication of data and erosion of trust within the scholarly ecosystem (Thorp, 2023).

Addressing these challenges requires a balanced approach that emphasizes human oversight, ethical responsibility and critical evaluation. AI should be positioned as a tool that augments human reasoning rather than a replacement for cognitive engagement. By reconceptualizing AI as a supportive instrument within the research process, scholars can leverage its potential while preserving intellectual authenticity, methodological rigor and ethical accountability.

8.2 Ethical and epistemic imperatives

Ethical imperatives in AI-assisted research remain paramount. Researchers must ensure transparency in disclosing AI contributions, verify the accuracy and validity of outputs and

maintain accountability for all content incorporated into manuscripts (Bender *et al.*, 2021). Epistemically, AI-generated outputs are derivative, synthesizing existing knowledge without independent reasoning or moral judgment. As such, the responsibility to critically interpret, contextualize and validate information rests entirely with human authors (Ji *et al.*, 2023).

Institutions, journals and funding agencies play a critical role in enforcing these imperatives by establishing clear guidelines, monitoring compliance and fostering a culture of ethical reflection. When ethical and epistemic standards are rigorously upheld, AI can enhance research quality without compromising scholarly integrity or public trust.

8.3 Systemic and cultural considerations

The broader research ecosystem is profoundly affected by AI adoption. Equitable access to AI tools, capacity-building initiatives and training programs are essential to prevent exacerbation of existing disparities (Hvistendahl, 2023). Academic cultures that prioritize transparency, quality and critical engagement over mere output volume provide the conditions necessary for responsible AI integration. Peer review, editorial oversight and collaborative practices must adapt to ensure that AI-enhanced manuscripts meet rigorous standards of accuracy, originality and accountability.

Cultural interventions, including mentorship, professional development and ethical modeling, reinforce individual responsibility while promoting systemic integrity. By embedding ethical and reflective practices into the norms of scholarly communities, the academic ecosystem can harness AI's benefits while mitigating risks associated with misconduct, inequity and epistemic erosion.

8.4 Implications for policy and governance

Policy and governance frameworks are central to shaping responsible AI use. Clear institutional guidelines, journal policies and international coordination efforts help standardize disclosure, ethical use and verification practices (Yeh *et al.*, 2024; Floridi and Chiriatti, 2020). Legal and regulatory considerations, including intellectual property and data privacy, further define the boundaries of acceptable AI application.

Effective governance requires adaptability, anticipating technological developments while maintaining foundational principles of accountability, transparency and scholarly rigor. Policies should also foster equitable access, promote capacity-building and align incentives with ethical practice. By integrating policy, governance and cultural norms, the research community can create a robust framework that ensures AI enhances rather than undermines the scholarly ecosystem.

8.5 Directions for future research and practice

Looking forward, continued investigation is necessary to understand the evolving dynamics of AI in research. Empirical studies examining the prevalence, impact and ethical considerations of AI-generated content will provide valuable insights for shaping policies, editorial standards and institutional practices (van Dis *et al.*, 2023). Methodological innovations that incorporate AI responsibly, while retaining human oversight and critical engagement, may redefine best practices in multiple disciplines.

Training programs and educational initiatives should evolve in parallel, equipping researchers with the skills to integrate AI ethically, critically and effectively. Cross-institutional and interdisciplinary collaborations can also foster shared understanding, resource sharing and equitable implementation of AI tools. By proactively addressing both opportunities and risks, the research community can navigate the challenges of AI integration while sustaining the integrity, credibility and advancement of scholarly knowledge.

8.6 Concluding reflection

AI's role in research is neither inherently beneficial nor intrinsically harmful; its impact depends on how scholars, institutions and publishers exercise responsibility, oversight and ethical judgment. When integrated with transparency, rigorous verification and equitable access, AI has the potential to augment human intellectual capacity, enhance research quality and broaden the inclusivity of scholarly participation. Conversely, neglecting ethical and epistemic responsibilities risks undermining trust, integrity and the cumulative value of knowledge.

The path forward requires deliberate reflection, coordinated policy and a cultural commitment to responsible practice. By aligning technological innovation with ethical principles and scholarly rigor, the academic community can ensure that AI serves as a transformative force for advancing knowledge while safeguarding the core values of scientific research.

Ultimately, the future of AI in research must rest upon enduring ethical foundations. Deontological commitments to truth and responsibility, virtue-ethical emphasis on intellectual humility and integrity and consequentialist concern for societal and epistemic well-being together offer a balanced moral framework for guiding AI's role in scholarship. Embedding these perspectives into research governance and everyday academic practice will ensure that technological progress remains anchored in human judgment, ethical reflection and collective accountability.

References

Agarwal, D., Naaman, M. and Vashistha, A. (2025), "AI suggestions homogenize writing toward Western styles and diminish cultural nuances", *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1-21.

Akbar, M.S. (2025), "Beyond detection: designing AI-resilient assessments with automated feedback tool to foster critical thinking", *arXiv preprint arXiv:2503.23622*.

Amershi, S., Cakmak, M., Knox, W.B. and Kulesza, T. (2014), "Power to the people: the role of humans in interactive machine learning", *AI Magazine*, Vol. 35 No. 4, pp. 105-120, doi: 10.1609/aimag.v35i4.2513.

Arar, K.H., Özen, H., Polat, G. and Turan, S, others (2025), "Artificial intelligence, generative artificial intelligence and research integrity: a hybrid systemic review", *Smart Learning Environments*, Vol. 12 No. 1, p. 44, doi: 10.1186/s40561-025-00403-3.

Asghar, M.Z., Iqbal, J., Özbilen, F.M., Abedin, J., Järvenoja, H. and Widanapathirana, U. (2025), "The nexus of artificial intelligence literacy collaborative knowledge practices and inclusive leadership development among higher education students in Bangladesh China Finland and Turkey", *Discover Computing*, Vol. 28 No. 1, pp. 1-30.

Bender, E.M., Gebru, T., McMillan-Major, A. and Mitchell, S. (2021), "On the dangers of stochastic parrots: Can language models be too big?", Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pp. 610-623, doi: 10.1145/3442188.3445922.

Bjelobaba, S., Waddington, L., Perkins, M. and Zarate, J. (2025), "Maintaining research integrity in the age of GenAI: an analysis of ethical challenges and recommendations to researchers", *International Journal for Educational Integrity*, Vol. 21 No. 1, p. 18, doi: 10.1007/s40979-025-00191-w.

Chen, Z., Chen, C., Yang, G., He, X., Chi, X., Zeng, Z. and Chen, X. (2024), "Research integrity in the era of artificial intelligence: challenges and responses", *Medicine*, Vol. 103 No. 27, p. e38811, doi: 10.1097/MD.0000000000038811.

Cheng, A., Calhoun, A. and Reedy, G. (2025), "Artificial intelligence-assisted academic writing: recommendations for ethical use", *Advances in Simulation*, Vol. 10 No. 1, p. 22.

Cotton, D., Cotton, P. and Shipway, J. (2023), "Chatting and cheating: ensuring academic integrity in the era of ChatGPT", *Innovations in Education and Teaching International*, Vol. 61 No. 2, pp. 1-13, doi: 10.1080/14703297.2023.2190148.

Dabis, A. and Csáki, C. (2024), "AI and ethics: investigating the first policy responses of higher education institutions to the challenge of generative AI", *Humanities and Social Sciences Communications*, Vol. 11 No. 1, pp. 1-13.

Doshi-Velez, F. and Kim, B. (2017), "Towards a rigorous science of interpretable machine learning", *arXiv Preprint arXiv:1702.08608*.

Dotson, K. (2014), "Conceptualizing epistemic oppression", *Social Epistemology*, Vol. 28 No. 2, pp. 115-138, doi: 10.1080/02691728.2013.782585.

Erol, G., Ergen, A., Gülşen Erol, B., Kaya Ergen, Ş., Bora, T.S., Çölgeçen, A.D., … Güngör, A. (2025), "Can we trust academic AI detective? Accuracy and limitations of AI-output detectors", *Acta Neurochirurgica*, Vol. 167 No. 1, pp. 1-12.

Färber, M., Coutinho, M. and Yuan, S. (2023), "Biases in scholarly recommender systems: impact, prevalence, and mitigation", *Scientometrics*, Vol. 128 No. 5, pp. 2703-2736, doi: 10.1007/s11192-023-04636-2.

Ferris, L. E. and Winker, M. A. (2017), "Ethical issues in publishing in predatory journals", *Biochemia Medica*, Vol. 27 No. 2, pp. 279-284.

Floridi, L. (2021), "Establishing the rules for building trustworthy AI", *Ethics, Governance, and Policies in Artificial Intelligence*, Springer International Publishing, Cham, pp. 41-45.

Floridi, L. and Chiriatti, M. (2020), "GPT-3: its nature, scope, limits, and consequences", *Minds and Machines*, Vol. 30 No. 4, pp. 1-3, doi: 10.1007/s11023-020-09548-1.

Gonsalves, C. (2024), "Generative AI's impact on critical thinking: revisiting bloom's taxonomy", *Journal of Marketing Education*, Vol. 48 No. 1, doi: 10.1177/02734753241305980.

Hagendorff, T. (2022), "A virtue-based framework to support putting AI ethics into practice", *Philosophy and Technology*, Vol. 35 No. 3, pp. 1-25.

Halupa, C. (2023), "Algiarism: artificial intelligence-assisted plagiarism", *Proceedings of the 15th International Conference on Education and New Learning Technologies (EDULEARN23)*, *IATED*, pp. 1018-1024, doi: 10.21125/edulearn.2023.0363

Hayes, P., Fitzpatrick, N. and Ferrández, J.M. (2025), "From applied ethics and ethical principles to virtue and narrative in AI practices", *AI and Ethics*, Vol. 5 No. 2, pp. 1217-1239, doi: 10.1007/s43681-024-00472-z.

Helm, P., Bella, G., Koch, G. and Giunchiglia, F. (2024), "Diversity and language technology: how language modeling bias causes epistemic injustice", *Ethics and Information Technology*, Vol. 26 No. 1, p. 8, doi: 10.1007/s10676-023-09742-6.

Hvistendahl, M. (2023), "AI and the global research divide", *Science*, Vol. 379 No. 6624, pp. 14-15, doi: 10.1126/science.abm2567.

ICMJE (2023), "Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals", International Committee of Medical Journal Editors, available at: www.icmje.org/recommendations

Ji, H., Choi, J., Liu, Y. and Zhao, Y. (2023), "Fabrication of data and references in AI-generated research: a systematic review", *Journal of Research Integrity*, Vol. 10 No. 2, pp. 45-58, doi: 10.1016/j.jri.2023.03.001.

Kandikatla, L. and Radeljic, B. (2025), "AI and human oversight: a risk-based framework for alignment", *arXiv Preprint arXiv:2510.09090*.

Kay, J., Kasirzadeh, A. and Mohamed, S. (2024), "Epistemic injustice in generative AI", *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 7, pp. 684-697.

Klimashevskaia, A., Jannach, D., Elahi, M. and Trattner, C. (2024), "A survey on popularity bias in recommender systems", *User Modeling and User-Adapted Interaction*, Vol. 34 No. 5, pp. 1777-1834.

Krishna, K., Song, Y., Karpinska, M., Wieting, J. and Iyyer, M. (2023), "Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense", *Advances in Neural Information Processing Systems*, Vol. 36, pp. 27469-27500.

Lee, H.P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R. and Wilson, N. (2025) "The impact of generative AI on critical thinking: self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers", *Proceedings of the 2025 CHI conference on human factors in computing systems*, pp. 1-22.

Lipton, Z.C. (2018), "The mythos of model interpretability", *Communications of the ACM*, Vol. 61 No. 10, pp. 36-43.

Madanchian, M. and Taherdoost, H. (2025), "Ethical theories, governance models, and strategic frameworks for responsible AI adoption and organizational success", *Frontiers in Artificial Intelligence*, Vol. 8, p. 1619029.

Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (2019), "Dissecting racial bias in an algorithm used to manage the health of populations", *Science*, Vol. 366 No. 6464, pp. 447-453.

Oğuz, H. and Alsancak-Akın, S. (2024), "Using generative AI in teaching and learning: ethical concerns and reduced critical thinking in higher education", *Journal of the Scholarship of Teaching and Learning*, Vol. 24 No. 4, pp. 162-173.

Paul, D. and Weinberger, M. (2024), "Transparency, accountability, and explainability in machine learning: a survey of current practices", *Philosophy and Technology*, Vol. 37 No. 4, p. 14, doi: 10.1007/s13347-024-00837-6.

Qadhi, S.M., Alduais, A., Chaaban, Y. and Khraisheh, M. (2024), "Generative AI, research ethics, and higher education research: insights from a scientometric analysis", *Information*, Vol. 15 No. 6, p. 325, doi: 10.3390/info15060325.

Renkema, M. and Tursunbayeva, A. (2024), "The future of work of academics in the age of artificial intelligence: state-of-the-art and a research roadmap", *Futures*, Vol. 163, p. 103453.

Resnik, D.B. and Hosseini, M. (2025), "The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool", *AI and Ethics*, Vol. 5 No. 2, pp. 1499-1521.

Rivero, E. and Yin, P. (2025), "Navigating the contradictions of AI: critiquing AI's standardized English and developing creative bilingual possibilities", *In Frontiers in Education*, Vol. 10, p. 1616458.

Salami, A.O. (2024), "Artificial intelligence, digital colonialism, and the implications for Africa's future development", *Data and Policy*, Vol. 6, p. e67.

Sheng, E., Chang, K.-W., Natarajan, P. and Peng, N. (2021), "Societal biases in language generation: progress and challenges", *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 330-345, available at: https://aclanthology.org/2021.acl-long.330/

Stokel-Walker, C. and Van Noorden, R. (2023), "What ChatGPT and generative AI mean for science", *Nature*, Vol. 614 No. 7947, pp. 214-216.

Tang, A., Li, K.K., Kwok, K.O., Cao, L., Luong, S. and Tam, W. (2024), "The importance of transparency: declaring the use of generative artificial intelligence (AI) in academic writing", *Journal of Nursing Scholarship*, Vol. 56 No. 2, pp. 314-318.

Thorp, H. (2023), "Generative approach to research integrity", *Science*, Vol. 381 No. 6658, pp. 16-17, doi: 10.1126/science.adk1852.

van Dis, E.A., Bollen, J., Zuidema, W., van Rooij, R. and Bender, E.M. (2023), "Ethical use of artificial intelligence for scientific writing: current perspectives", *Journal of Research Ethics*, Vol. 19 No. 3, pp. 123-135, doi: 10.1177/17470161231180449.

Van Slyke, C., Sarabadani, J. and Mosafer, H. (2025), "Epistemic injustice in generative AI", in Aubert, B., Chin, W., Parsons, J., Boudreau, M.C., Corbett, J. and Gupta, A. (Eds), *31st Americas Conference on Information Systems: Intelligent Technologies for a Better Future, AMCIS 2025, Association for Information Systems*, pp. 201-205.

van Wynsberghe, A. (2021), "Sustainable AI: AI for sustainability and the sustainability of AI", *AI and Ethics*, Vol. 1 No. 3, pp. 213-218.

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., … Waddington, L. (2023), "Testing of detection tools for AI-generated text", *International Journal for Educational Integrity*, Vol. 19 No. 1, pp. 1-39.

Yeh, M.H., Wang, J., Du, X., Park, S., Tao, L., Im, S. and Li, Y. (2024), "Challenges and future directions of Data-Centric AI alignment", *arXiv Preprint arXiv:2410.01957*.

Zárate-Torres, R., Rey-Sarmiento, C.F., Acosta-Prado, J.C., Gómez-Cruz, N.A., Rodríguez Castro, D.Y. and Camargo, J. (2025), "Influence of leadership on human–artificial intelligence collaboration", *Behavioral Sciences*, Vol. 15 No. 7, p. 873.

Corresponding author
Farooq Mubarak can be contacted at: wohi_farooq@yahoo.com