Journal of Applied Social Sciences and Humanities

https://journal.krrishpub.com/index.php/jassh/index

RESEARCH ARTICLE

Leveraging Artificial Intelligence in Scholarly Publishing

Fatima Zahra Ouariach [a,*], Khouloud EL Meziani [b], Soufiane Ouariach [c]

[a] *Educational Sciences, Language didactics, Human and Social Sciences, Abdelmalek Essaadi University, Tetouan, 93000, Morocco.*
[b] *Department of Management, Abdelmalek Essaadi University, Tangier 90000, Morocco.*
[c] *Computer Science and University Pedagogical Engineering (S2IPU), Higher Normal School of Tetouan, 93000, Abdelmalek Essaadi University, Morocco.*

Abstract

The integration of Artificial Intelligence (AI) into scholarly publishing constitutes a structural transformation of historical significance, fundamentally reshaping how knowledge is produced, evaluated, and disseminated. This study presents a systematic analysis of AI adoption within the global research ecosystem, focusing on the critical period from 2021 to late 2025. Using a secondary data analysis framework, the paper examines the dual role of generative AI and large language models (LLMs) as both enablers of unprecedented efficiency and sources of emerging epistemic risk. Drawing on bibliometric evidence, industry reports, and peer-reviewed literature, the analysis reveals a rapid escalation in AI use among researchers, reaching 58% globally in 2025 compared to 37% in 2024. While the literature consistently demonstrates that AI substantially accelerates scholarly workflows—most notably in grant writing, literature synthesis, and preliminary review—it also exposes systemic vulnerabilities, including citation hallucination, opacity in reasoning, and erosion of academic integrity. These risks are compounded by the potential amplification of epistemic injustice, as AI systems trained on dominant linguistic and cultural corpora may marginalize non-Western and non-native English scholarship. The study is guided by two objectives: (i) to evaluate the operational efficacy of AI in streamlining research workflows and (ii) to assess the ethical and institutional implications of emergent "posthuman" authorship. Findings indicate that while AI-assisted tools can reduce grant preparation time by more than 90%, they simultaneously generate non-verifiable citations at rates that threaten the cumulative reliability of the scholarly record. Comparative analysis of detection tools and publisher policies further demonstrates that existing governance mechanisms are fragmented, biased, and insufficient for AI-scale knowledge production. The paper argues that academia is entering a posthuman phase of authorship in which human–machine collaboration destabilizes conventional notions of originality, accountability, and intellectual ownership. Without robust governance frameworks and a redefinition of scholarly integrity, the scientific record risks contamination by machine-generated simulacra of knowledge, undermining trust in research as a public good.

Keywords: Generative AI, Scholarly publishing, Research integrity, Epistemic injustice, Posthuman authorship, Academic Governance, Bibliometrics.

Article information:
Published by: **Krrish Scientific Publications Pvt. Ltd.**
DOI: https://doi.org/10.71426/jassh.v1.i1.pp1-8
Received: *30 Nov. 2025* | Revised: *26 Dec. 2025* | Accepted: *30 Dec. 2025* | Published: *31 Dec. 2025*

1. Introduction

Scholarly publishing is undergoing a transformation comparable to earlier paradigm shifts from manuscript to print and from print to digital dissemination. The rapid adoption of AI, particularly generative AI and LLMs, is reconfiguring how knowledge is produced, evaluated, and disseminated. What began as experimental assistance for writing and search has, by 2025, expanded into routine use across the entire research lifecycle, including literature review, grant writing, manuscript drafting, and peer review. This acceleration is strongly linked to unprecedented global investment in AI, which has fueled the deployment of tools capable of automating cognitive tasks that were previously central to academic labor [21, 22].

Empirical evidence indicates that adoption has out-

*Corresponding author
Email addresses: fati.ouariach.99@gmail.com, fatimazahra.ouariach@etu.uae.ac.ma (Fatima Zahra Ouariach), elmezianikhouloud@gmail.com, khouloud.elmeziani@etu.uae.ac.ma (Khouloud EL Meziani), ouariach.soufiane.97@gmail.com, Soufian.ouariach@etu.uae.ac.ma (Soufiane Ouariach).

Table 1: List of Acronyms

Acronym	Expansion
AI	Artificial Intelligence
GenAI	Generative Artificial Intelligence
LLM(s)	Large Language Model(s)
STM	Scientific, Technical and Medical (publishing)
TEQSA	Tertiary Education Quality and Standards Agency
AIES	AAAI/ACM Conference on AI, Ethics, and Society
ACS	American Chemical Society
DOI	Digital Object Identifier
L2	Second-language (Non-native English)
IT	Information Technology
API	Application Programming Interface
N/A	Not Available
USA	United States of America
UK	United Kingdom
HAI	Human-Centered Artificial Intelligence
KPI	Key Performance Indicator

paced governance. A global survey by Elsevier reports that 58% of researchers used AI tools in 2025, up from 37% in 2024, primarily to summarize literature and accelerate review processes [1]. Yet trust in these tools remains low: fewer than one-quarter of respondents consider them ethically developed or reliable. This mismatch between widespread use and limited confidence has produced a fragile research environment in which efficiency gains coexist with uncertainty about integrity, accountability, and authorship.

The risks are amplified by the opaque, "black-box" nature of LLMs. While these systems perform well on standardized benchmarks, their reasoning processes are not transparent, limiting reproducibility and raising concerns for high-stakes scientific applications. At the same time, the rise of AI-assisted paper mills, fabricated citations, and synthetic manuscripts has contributed to a surge in retractions, exposing structural vulnerabilities in peer review and editorial oversight [15]. These developments force the research community to confront unresolved questions: Who is responsible when machine-generated content enters the scholarly record, and how should credit, liability, and originality be defined in hybrid human–machine authorship?

Beyond integrity, AI adoption also raises equity concerns. Because most generative models are trained on English-language, Western-centric corpora, their outputs risk reinforcing existing epistemic hierarchies and marginalizing non-dominant knowledge systems. This dynamic threatens to deepen epistemic injustice by privileging certain writing styles, methodologies, and citation networks, while disadvantaging researchers from the Global South or non-native English backgrounds.

Against this backdrop, this study analyzes verified evidence from 2021–2025 to examine how AI is reshaping scholarly publishing at technical, ethical, and institutional levels. By synthesizing industry reports and peer-reviewed research, the paper evaluates the dual role of AI as both an efficiency-enhancing infrastructure and a destabilizing force for research integrity, authorship, and trust.

2. Literature Review

The scholarship on AI in scholarly publishing has progressed rapidly from exploratory commentary to problem-driven analyses of integrity, governance, and workflow transformation. The studies reviewed here synthesize sixteen perspectives that collectively map the emerging benefits of generative AI alongside its risks to trust, attribution, and equity.

A foundational concern is academic integrity in higher education. In a major report for TEQSA, Lodge argues that the inappropriate use of tools such as ChatGPT poses an immediate threat to assessment validity, with student usage estimates ranging from 10% to more than 60%. He emphasizes that institutions are increasingly unable to distinguish legitimate assistance from misconduct, and therefore require near-term mitigation while longer-term policy frameworks mature [2].

In contrast, evidence from research administration highlights substantial productivity gains. Rybiński reports that AI-assisted grant writing can compress preparation timelines from 30–50 days to 3–5 days, with an observed proposal success rate of around 50% compared with typical baselines of 10–20%. While these outcomes suggest greater efficiency and potentially wider access to funding opportunities, they also raise normative questions about whether competitive advantage shifts from scientific merit to prompt literacy and rhetorical optimization [3].

In medical evidence synthesis, Goyal et al. find that ChatGPT-4.0 can produce meta-analytic outputs showing high concordance with traditional approaches in cardiology contexts. However, they caution that limited transparency in model reasoning and data handling constrains clinical trust and reproducibility, underscoring the tension between apparent accuracy and methodological opacity [4].

The peer-review process has also become a testbed for LLM deployment. Liang et al. report measurable overlap between GPT-4-generated critiques and human reviewer comments across thousands of manuscripts. Although this indicates functional usefulness for preliminary feedback, the findings also suggest that automated reviews may privilege surface-level issues and risk homogenizing evaluation, potentially weakening the role of expert judgment in assessing novelty and conceptual rigor [5].

Beyond workflow performance, governance and ethics remain central. Malik et al. argue that institutional adoption of AI without continuous ethical oversight can amplify surveillance, erode trust, and embed inequities through algorithmic decision-making. Their analysis frames these risks as a form of epistemic injustice, where educational and scholarly practices become increasingly shaped by opaque metrics rather than human-centered pedagogy [6]. Complementing this, Craig and Kerr critique the assumption that authorship is inherently human, arguing that meaning and accountability are destabilized when text production becomes distributed across human prompts and machine generation, thereby challenging established norms in plagiarism, attribution, and intellectual responsibility [7].

At the institutional level, Clarivate's "Pulse of the Library" report documents a broad interest in AI integration among libraries, with many institutions evaluating deployment plans. The report also highlights a pronounced

skills gap, as upskilling and operational readiness remain key barriers to responsible implementation in knowledge management environments [8]. At the researcher level, Elsevier's global survey similarly shows rapid adoption but uneven confidence: high usage coexists with low trust, and attitudes vary across regions, implying that AI is reshaping scholarly practice through differing cultural and governance contexts [1].

Empirical work from developing educational contexts further underscores the equity dimension. Wiredu et al. report that generative AI can support learning outcomes, but also increases integrity risks when AI literacy and institutional safeguards are insufficient, potentially widening educational inequalities where oversight capacity is limited [9]. In parallel, Meakin finds that students increasingly bypass conventional library search systems in favor of conversational AI, shifting discovery from curated retrieval to probabilistic synthesis and raising concerns about provenance, verifiability, and the dilution of engagement with primary scholarly sources [10].

Within publishing integrity itself, Chauhan and Currie analyze how generative AI pressures research integrity norms in scholarly publishing ecosystems. Their discussion highlights the need for robust governance and clear accountability structures as AI becomes embedded across editorial and author workflows [11]. Related ethical critiques examine how generative systems can reproduce epistemic injustice through the biases and asymmetries encoded in training data. Kay et al. [12] provide a conceptual account of epistemic injustice in generative AI, while Hua et al. extend these concerns in the context of mental health, illustrating how model outputs can reinforce dominant epistemologies and marginalize alternative perspectives [13]. In scholarly publishing, such dynamics may manifest as systematic privileging of Anglophone and Western-centric citation networks.

Concerns about manipulation and fraud have intensified alongside adoption. Else documents how "tortured phrases" can signal fabricated or distorted scholarship, and Retraction Watch reports further emphasize the scale and visibility of integrity breaches in the AI era. Together, these accounts point to an arms race between industrialized paper production and detection or editorial safeguards [14, 15].

User acceptance and pedagogical legitimacy remain contested. Er et al. show that students often prefer instructor feedback over AI-generated feedback, even when human feedback is slower, suggesting that timeliness does not substitute for perceived credibility, nuance, and relational value in assessment [16]. A further integrity risk arises from citation hallucinations. Meyer et al. [17] discuss broader opportunities and challenges of LLMs in academia, while Gao et al. demonstrate how automated review-generation pipelines can inadvertently support plausible but unreliable scholarly text production. This line of work highlights how fabricated or non-verifiable references can contaminate the citation record and undermine cumulative knowledge building [18].

Finally, Stokel-Walker and Van Noorden describe the ongoing contest between increasingly capable text generators and imperfect detection tools. They emphasize that false positives, particularly for non-native English writers, create significant equity risks and may foster a presumption-of-guilt environment, suggesting that governance strategies must move beyond detection toward transparent disclosure norms and redesigned assessment and review practices [19].

3. Research Methodology

3.1. Research design

This study employs a secondary data analysis research design, a methodological choice necessitated by the velocity of change in the AI domain. Given the rapid evolution of Artificial Intelligence between 2021 and 2025, traditional primary data collection methods such as longitudinal surveys would likely yield obsolete results by the time of publication. Secondary data analysis allows for the aggregation of multiple high-quality, verified datasets from divergent sources including industry reports, bibliometric studies, and global surveys providing a meta-level view of the phenomenon that is both broad and deep. This method is particularly appropriate for "Research Education" as it enables the synthesis of pedagogical outcomes, administrative trends, and technological adoption rates across different institutional contexts without the limitations of a single-institution study.

3.2. Data collection

The Data was curated from high-impact industry reports and peer-reviewed studies published between 2021 and 2025. Key sources include the Stanford HAI AI Index 2025 [21], Elsevier Insights 2024 [1], and the Clarivate Pulse of the Library 2024 report [8]. Peer-reviewed articles were selected from top-tier journals in information science, higher education, and ethics. The selection criteria prioritized reports with global sample sizes (>1000 respondents) to ensure representativeness and peer-reviewed articles with reproducible methodologies. Specific attention was paid to datasets that disaggregated results by region (Global North vs. South) and language (Native vs. Non-Native English speakers) to address the equity dimensions of the study[20].

3.3. Data analysis

The analysis proceeds by categorizing data into "Efficiency Metrics" (adoption rates, time savings, success rates) and "Integrity Metrics" (retraction counts, detection accuracy, hallucination rates). The interpretation involves cross-referencing these datasets to identify contradictions such as the gap between high adoption rates and low trust levels and synthesizing them into a coherent narrative. For example, data on grant writing efficiency is cross-referenced with data on citation hallucinations to build a composite picture of the risks and rewards. The analysis also employs thematic coding to identify recurring motifs in the qualitative data, such as "epistemic injustice" and "posthuman authorship," allowing for the construction of theoretical frameworks to explain the observed trends.

4. Results and Analysis

The analysis of the collected data reveals a complex landscape where rapid adoption coexists with significant infrastructural and ethical vulnerabilities. The integration of AI is not a linear progress narrative but a disruptive event creating winners and losers based on geography, language, and institutional resources. The following tables and interpretations break down the key dimensions of this transformation.

Table 2: Comparative AI adoption and sentiment in research communities (2024-2025) [21]

Metric	Global Average	USA	China	United Kingdom
Active AI Usage (2025)	58%	N/A	N/A	N/A
Active AI Usage (2024)	37%	N/A	N/A	N/A
Belief AI Empowering	N/A	25%	64%	24%
Belief AI Saves Time	58%	54%	79%	57%
Belief AI Improves Quality	N/A	22%	60%	17%

The data presented in Table 2 illustrates a profound geopolitical divergence in the reception of AI technologies within the academy. While the global average for active AI usage has surged from 37% to 58% in a single year, representing a massive behavioral shift, the attitudes driving this adoption vary wildly. China emerges as the distinct leader in "techno-optimism," with nearly two-thirds (64%) of researchers viewing AI as empowering. This stands in stark contrast to the United States and the United Kingdom, where only a quarter of researchers share this sentiment. This disparity is likely driven by differing regulatory environments and cultural attitudes toward automation. In the West, the narrative is dominated by fear of job displacement, copyright infringement, and ethical breaches, reflected in the low belief that AI improves work quality (22% in the US, 17% in the UK). Conversely, the Chinese research ecosystem appears to view AI as a critical lever for productivity and advancement, potentially integrating it more deeply into national research strategies.

This trend suggests that future high-volume scientific output may increasingly originate from regions willing to integrate AI into the core research workflow, potentially creating a "productivity gap" between East and West. The high global agreement that AI "saves time" (58%) indicates that efficiency is the universal driver, but the perception of quality remains the primary friction point for Western adoption. Researchers in the West are using AI because they have to for efficiency, not because they believe in it, creating a cynical engagement with the technology that may undermine its effective governance.

Table 3 reveals the precarious state of "policing" AI in academia. While tools like Copyleaks demonstrate high accuracy in controlled environments, the broader ecosystem of detection is fraught with inconsistency. The most alarming finding is the persistent bias against non-native English speakers (L2 writers). As noted in the literature review, writing that is formulaic or uses limited vocabulary

Table 3: Efficacy and bias of major AI detection tools (2024)[22]

Detection Tool	Accuracy (Native English)	False Positive Rate (General)	Bias Against Non-Native Writers
Copyleaks	∼99–100%	Low (<1%)	Low
Turnitin	High (>95%)	<1% (claimed)	Moderate (flagging valid L2 writing)
GPTZero	Variable (90–95%)	Moderate	High
OpenAI Classifier	Low (<30%)	High (9%)	Significant (discontinued)

traits common in L2 writing is frequently misclassified as AI-generated because LLMs themselves are designed to produce "average" text. The Turnitin data acknowledges this difficulty by hiding scores below 20% to avoid false positives, effectively admitting that low-level detection is unreliable.

The failure of OpenAI's own classifier, which was discontinued due to low accuracy, serves as a bellwether for the industry. If the creators of the technology cannot reliably detect it, third-party tools face an uphill battle. The reliance on these tools by universities to adjudicate academic misconduct cases is ethically suspect. If a tool like GPTZero has a "High" bias against non-native writers, its use in global institutions constitutes a structural discrimination barrier, potentially penalizing students and researchers from the Global South for simply writing in a second language. This creates a "digital racial profiling" where algorithmic suspicion falls disproportionately on those already marginalized in the academic system. This data supports the argument that the "arms race" between generation and detection is unwinnable and that energy is better spent on reforming assessment and authorship norms.

Table 4: Publishers' regulatory frameworks on AI (2024–2025) [21]

Feature	Elsevier	Wiley	Springer Nature	ACS
AI as Author	Prohibited	Prohibited	Prohibited	Prohibited
Accountability	Author assumes full liability	Author fully accountable	Human accountability mandatory	Author responsible for accuracy
Disclosure	Required (AI Declaration)	Required in Methods/Ack.	Required in Methods	Required in Ack/Methods
GenAI Images	Prohibited (exceptions apply)	Restricted	Restricted	Disclosure required in captions

The regulatory landscape summarized in Table 4 shows a strong consensus among major publishers: AI cannot be an author. This seemingly simple rule is the "human firewall" attempting to protect the legal and ethical concept of authorship. All major publishers (Elsevier, Wiley, Springer, ACS) mandate that humans must take full responsibility for the content. This is a liability containment strategy; if an AI hallucinates a libelous statement or a medical error, the publisher needs a human legal entity to

hold accountable. The rejection of "AI as Author" is also a defense of the copyright system, which generally requires human creativity for protection.

However, the policies regarding the use of AI are more porous. While "AI as Author" is banned, "AI as Assistant" is permitted with disclosure. The nuance lies in the "GenAI Images" row. The prohibition or strict restriction of AI-generated images reflects the specific panic over fabricated data and deepfakes, which are harder to detect than text and have a higher potential for scientific fraud. The requirement for disclosure in "Methods" or "Acknowledgments" attempts to enforce transparency, but it relies entirely on the honor system. Given the pressure to publish and the competitive advantage of using AI (as seen in the grant writing data), it is highly probable that a significant volume of AI usage goes undeclared, rendering these policies partially performative. The policies create a "don't ask, don't tell" environment where smart AI use is rewarded, but clumsy AI use is punished.

Table 5: Retraction trends and causes in the AI era (2023–2024) [15]

Year	**Total Retractions**	**Key Drivers**	**Impacted Regions**
2023	∼10,000+	Fake peer review, "Tortured phrases"	Global
2024	∼14,000+	AI-generated images, Paper mills, Fake data	Emerging research economies
Trend	+40% Increase	Shift from manual manipulation to AI-scale fraud	Broadening to mainstream journals

Table 5 presents a grim trajectory for research integrity. The jump from 10,000 retractions in 2023 to over 14,000 in 2024 represents a 40% increase in a single year, a rate that far outpaces the growth of legitimate publishing. The "Key Drivers" column is telling; we have moved from "fake peer review" (a human coordination problem) to "AI-generated images" and "tortured phrases" (a machine scale problem). The mention of "Emerging research economies" as heavily impacted relates back to the pressures of "Publish or Perish" which are often more acute in systems using quantitative KPIs for promotion.

The presence of "tortured phrases" (e.g., "counterfeit consciousness" instead of "artificial intelligence") is a direct artifact of using spinning software to evade plagiarism detectors. This data confirms that AI is weaponizing academic fraud, allowing bad actors to generate fraudulent papers at a speed and scale that overwhelms the traditional peer review safeguards. The retraction mechanism, once a rare corrective, is becoming a routine sanitation process, suggesting that the "filter" of peer review is broken. The scale of fraud suggests that we are dealing with industrial "paper mills" that use AI to optimize their production lines, creating a flood of synthetic garbage that drowns out legitimate science.

Table 6 highlights the most disruptive finding of the report regarding efficiency. The reduction in grant preparation time by approximately 90% is transformative. In

Table 6: Impact of AI on grant writing efficiency and success [3]

Study Metric	**Traditional Human Process**	**AI-Assisted Process**	**Variance**
Preparation Time	30–50 Days	3–5 Days	∼90% Reduction
Success Rate	10–20% (Baseline)	50% (Observed)	+150% to +400%
Cost per Proposal	High (Human Hours)	Low (API/Sub Cost)	Significant Reduction

a research environment where faculty are overburdened with administrative tasks, the ability to draft a proposal in days rather than months is an irresistible value proposition. More critically, the observed success rate of 50% for AI-assisted proposals challenges the very notion of merit in funding. If an AI can structure a proposal more persuasively than a human, are funders rewarding the best science or the best rhetoric? This disparity creates an immediate equity issue: researchers who refuse to use AI or lack access to advanced models are competing with a severe handicap. It effectively mandates AI adoption for survival in the funding marketplace. This commodification of grant writing may lead to a system where "proposal quality" becomes decoupled from "research feasibility," as AI excels at the former but has no understanding of the latter. Table 7 ex-

Table 7: AI Hallucination rates in academic citation [17], [18]

Discipline	**Citation Accuracy**	**DOI Hallucination Rate**	**Source Reliability**
Natural Sciences	72.7%	29.1%	Moderate
Humanities	76.6%	61.7%	Low
Overall Trend	Variable	High in niche fields	Decreasing with model updates

poses the "Achilles' heel" of AI in research: the fabrication of authority. While the text generated by LLMs is often coherent, the citation accuracy is dangerously inconsistent. The "DOI Hallucination Rate" is particularly concerning; nearly 30% in sciences and over 60% in humanities involves the invention of DOIs. This means the AI is generating links to papers that do not exist. This is not merely an error; it is the creation of "ghost knowledge." In the Humanities, where citation is a primary form of evidence, a 60% error rate renders current models practically unusable for rigorous work without intense human oversight. The variance between disciplines likely reflects the training data density; widely cited scientific papers are better represented in the model's weights than niche humanities texts. This data underscores that while AI can write like an academic, it cannot yet reference like one, posing a threat to the genealogical integrity of scholarship.

The findings of this research report unequivocally demonstrate that the integration of Artificial Intelligence into scholarly publishing is a double-edged sword that is currently cutting deeply into the fabric of academic tradition.

4.1. Efficiency paradox

On the axis of efficiency, the data is staggering. The ability to reduce grant writing time by approximately 90%

(from months to days) and the 50% success rate of AI-assisted proposals suggests that AI is not merely a tool for convenience but a decisive competitive advantage. In a zero-sum funding environment, this necessitates the adoption of AI by all researchers merely to survive, creating a coercive adoption curve. The high adoption rates in China compared to the West indicate that this competitive pressure is being embraced unevenly, potentially leading to a geopolitical shift in research output volume where "AI-augmented" nations outpace "AI-hesitant" ones. However, this efficiency is paradoxical; while individual productivity rises, system-wide noise increases, potentially clogging peer review channels with plausible but mediocre or fabricated submissions.

4.2. Industrialization of fraud

On the axis of integrity, the findings are deeply concerning. The explosion in retractions to over 14,000 in 2024 is a direct downstream effect of the industrialization of fake science enabled by GenAI. The "tortured phrases" phenomenon reveals that bad actors are using AI to bypass the very safeguards (plagiarism detection) designed to protect the record. We are effectively in a "post-truth" phase of academic publishing where the provenance of text is unverifiable. The finding that AI tools frequently "hallucinate" citations inventing papers that do not exist poses a unique threat to the cumulative nature of science. If the "shoulders of giants" upon which we stand are digital mirages, the structural integrity of future research is compromised.

4.3. Failure of policing

The analysis of detection tools reveals a systemic failure. Tools like GPTZero and Turnitin exhibit significant bias against non-native English speakers, flagging their writing as AI-generated due to linguistic patterns rather than actual AI use. This creates a "presumption of guilt" for researchers from the Global South, exacerbating existing inequalities. The inability of these tools to keep pace with LLM advancement suggests that the "detection" strategy is a dead end. Reliance on the "honor system" for disclosure, as mandated by publishers, appears insufficient given the powerful incentives to use AI covertly.

4.4. Epistemic injustice

The data supports the hypothesis of algorithmic epistemic injustice. By relying on models trained primarily on English-language, Western-centric data, the research community risks amplifying dominant narratives while silencing diverse epistemologies. The "black box" nature of these models means that the biases encoded in their training data are opaque and difficult to challenge. This threatens to homogenize global knowledge production, reducing the rich diversity of human thought to a statistical average determined by a few corporate entities in the Global North.

5. Discussion

The philosophical depths of these results point toward a radical reconfiguration of the knowing subject. We are witnessing the emergence of the "posthuman author." Traditionally, the academic author was viewed as the sole originator of ideas, a sovereign individual whose intellectual property was sacrosanct. The widespread use of AI challenges this humanist ideal. If a researcher prompts an LLM to synthesize a literature review, and the LLM selects the sources, structures the argument, and drafts the prose, the locus of cognition has shifted from the individual to the human-machine assemblage. This is not merely "assistance"; it is "co-creation," yet our ethical and legal frameworks lack the vocabulary to accommodate this. We cling to binary categories human vs. machine while the reality of research practice has already become hybrid. This creates a dissonance where researchers must perform "human purity" for publishers while privately relying on machine labor.

Again, the data on bias and detection failures highlights a growing "epistemic injustice." The algorithmic governance of academia through detection tools, citation indexes, and automated screening is encoding the norms of the "center" (Global North, Native English) and punishing the "periphery." When a detection tool flags a valid Nigerian or Chinese manuscript as "AI" because of its sentence structure, it is committing an act of testimonial injustice, denying the author credibility based on an algorithmic prejudice. Conversely, the "black box" nature of LLMs, which are trained primarily on English-language data, reinforces the dominance of Anglophone epistemology. As these tools become the primary interface for knowledge discovery, they risk creating a feedback loop where Western knowledge is endlessly recycled and amplified, while indigenous and non-dominant knowledges are algorithmically erased.

The academy thus faces a paradox: AI democratizes access to tools of production (writing, coding), but it may simultaneously narrow the scope of what is considered valid knowledge. The "democratization" of grant writing success via ChatGPT might seem like a win for equity, allowing non-native speakers to compete with native speakers. However, if the criteria for success is merely the ability to mimic the rhetorical style of the dominant culture (which the AI does perfectly), then the system is not becoming more inclusive; it is simply becoming better at enforcing conformity. The AI helps you sound like "us," but it does not necessarily help you challenge "us."

The rise of "shadow IT" in academia where 58% of researchers use tools they don't trust indicates a breakdown in institutional culture. Researchers are trapped in a prisoner's dilemma: if they don't use AI, they fall behind in productivity; if they do use AI, they risk ethical breaches and "hallucinated" errors. The lack of clear, enforceable, and equitable guidelines exacerbates this tension. The "posthuman" turn is not a distant future; it is the operational reality of the present, and the academy is currently operating without a map.

6. Conclusion

This study concludes that the academic ecosystem has irreversibly crossed the event horizon of AI integration. The efficiency gains are too immense to roll back; the grant writing and literature synthesis capabilities of current models offer a solution to the unsustainable workload

of the modern researcher. However, this efficiency comes at the cost of transparency and trust. The current governance mechanisms publisher policies of "disclosure" and algorithmic detection tools are failing. Detection tools are scientifically unreliable and ethically compromised by bias. Publisher mandates are unenforceable. Consequently, the scholarly record is becoming a hybrid archive of human and machine outputs, with the distinction between the two becoming increasingly meaningless. The rise in retractions is not a temporary anomaly but a structural feature of this new era, signaling that the traditional peer review model is insufficient for vetting AI-scale output.

The implications for universities, funders, and publishers are profound. Universities must abandon the punitive approach to AI (bans and detectors) in favor of critical AI literacy. If detection is biased and inaccurate, using it for disciplinary action is a legal liability and an ethical failure. Assessment must shift from "product-based" (the essay) to "process-based" (the defense of the essay). Funders must redesign the grant application process; if a machine can write a winning proposal in three days, the proposal format itself is no longer a valid proxy for scientific merit. Evaluation must shift towards track record, replicability, and perhaps in-person defense of ideas to verify the "human" understanding of the project. Publishers must accept that the "version of record" can no longer be guaranteed by pre-publication review alone. Post-publication review and living documents may become necessary to scrub the record of hallucinatory errors that slip through the initial net. The "human accountability" clause in policies needs to be backed by better verification tools, such as watermarking or provenance tracking.

Future research must move beyond the binary of "is it AI?" to the nuance of "how does AI shape the thought?" Longitudinal studies are needed to track the long-term impact of AI assistance on the cognitive development of early-career researchers. Does offloading the literature review process to an AI result in a shallower understanding of the field? Additionally, urgent research is needed into "watermarking" and provenance technologies that can authenticate the human origin of critical data without relying on biased textual analysis. Finally, the geopolitical dimension requires attention: how will the divergent AI adoption rates in China and the West reshape the balance of scientific power in the coming decade? The answers to these questions will define the future of human knowledge.

Declarations and ethical statements

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding statement: The authors declare that no specific funding was received for this research.

Artificial Intelligence usage statement: During the preparation of this work, the authors used ChatGPT to assist with grammatical corrections. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Availability of data and materials: The data and/or materials that support the findings of this study are available from the corresponding author upon reasonable request.

CRediT authorship contribution statement

Fatima Zahra Ouariach: Conceptualization, Investigation, Writing – Original draft & Validation. **Khouloud EL Meziani**: Conceptualization, Methodology, Writing – Editing & Validation. **Soufiane Ouariach**: Conceptualization, Methodology, Writing – Editing & Validation.

Publisher's note

References

[1] [Online Available]: Insights 2024: Attitudes toward AI. *Elsevier*. Available from: https://www.elsevier.com/en-in/insights/attitudes-toward-ai.

[2] Lodge JM. The evolving risk to academic integrity posed by generative artificial intelligence: Options for immediate action. *Tertiary Education Quality and Standards Agency*. 2024 Aug;8. Available from:

[3] Rybiński K. AUTOMATION OF GRANT APPLICATION WRITING WITH THE USE OF CHATGPT. *Scientific Papers of Silesian University of Technology. Organization & Management/Zeszyty Naukowe Politechniki Slaskiej. Seria Organizacji i Zarzadzanie*. 2025 Jan 15(216). Available from: http://dx.doi.org/10.29119/1641-3466.2025.216.30

[4] Goyal A, Tariq MD, Ahsan A, Khan MH, Zaheer A, Jain H, Maheshwari S, Brateanu A. Accuracy of artificial intelligence in meta-analysis: A comparative study of ChatGPT 4.0 and traditional methods in data synthesis. *World Journal of Methodology*. 2025 Dec 20;15(4):102290.Available from: https://doi.org/10.5662/wjm.v15.i4.102290

[5] Liang W, Zhang Y, Cao H, Wang B, Ding DY, Yang X, Vodrahalli K, He S, Smith DS, Yin Y, McFarland DA. Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *NEJM AI*. 2024 Jul 25;1(8):AIoa2400196. Available from: https://ai.nejm.org/doi/abs/10.1056/AIoa2400196

[6] Malik S, Naz L, Ghafoor H, Sohail Z, Shaheen R. Artificial Intelligence in Education: Bridging Technology and Pedagogy for Student-Centered Outcomes. *The Critical Review of Social Sciences Studies*.2025 Sep 9;3(3):2394-405. Available from: https://doi.org/10.59075/ga8h2c83

[7] Craig C, Kerr I. The death of the AI author. *Edward Elgar Publishing eBooks*. 2025 Apr 8;52(1):250–85. Available from: https://doi.org/10.4337/9781800887305.00014

[8] [Online Available]: Pulse of the Library 2024. *Clarivate*.September 19, 2024. Available from: https://exlibrisgroup.com/blog/the-clarivate-pulse-of-the-library-2024-report-is-now-available/#:~:text=This%20comprehensive%20report%20is%20based,priority%20for%20the%20next%20year.

[9] Wiredu JK, Abuba NS, Zakaria H. Impact of generative AI in academic integrity and learning Outcomes: a case study in the Upper East region. *Asian Journal of Research in Computer Science*. 2024 Jul 30;17(8):70–88. Available from: https://doi.org/10.9734/ajrcos/2024/v17i7491

[10] Meakin L. Exploring the impact of generative artificial intelligence on higher education students' utilization of library resources: A critical examination. *Information Technology and Libraries*. 2024 Sep 23;43(3). Available from: https://doi.org/10.5860/ital.v43i3.17246

[11] Chauhan C, Currie G. The impact of generative artificial intelligence on research integrity in scholarly publishing. *American Journal of Pathology*. 2024 Oct 11;194(12):2234–8. Available from: https://doi.org/10.1016/j.ajpath.2024.10.001

[12] Kay J, Kasirzadeh A, Mohamed S. Epistemic injustice in generative AI. *InProceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 2024 Oct 16 (Vol. 7, pp. 684-697). Available from: https://doi.org/10.1609/aies.v7i1.31671

[13] Hua Y, Liu F, Yang K, Li Z, Na H, Sheu YH, Zhou P, Moran LV, Ananiadou S, Clifton DA, Beam A. Large language models in mental health care: a scoping review. *Current Treatment Options in Psychiatry*. 2025 Jul 25;12(1):27. Available from: https://doi.org/10.1007/s40501-025-00363-y

[14] Else H. Tortured phrases' give away fabricated.*Nature*. 2021 Aug 19;596:328-9. Available from: https://doi.org/10.1038/d41586-021-02134-0

[15] [Online Available]: AI unreliable in identifying retracted research papers, says study. *Retraction Watch*. Available from: https://retractionwatch.com/2025/11/19/ai-unreliable-identifying-retracted-research-papers-study/

[16] Er E, Akçapınar G, Bayazıt A, Noroozi O, Banihashem SK. Assessing student perceptions and use of instructor versus AI-generated feedback. *British Journal of Educational Technology*. 2025 May;56(3):1074-91. Available from: https://doi.org/10.1111/bjet.13558

[17] Meyer JG, Urbanowicz RJ, Martin PC, O'Connor K, Li R, Peng PC, Bright TJ, Tatonetti N, Won KJ, Gonzalez-Hernandez G, Moore JH. ChatGPT and large language models in academia: opportunities and challenges. *BioData mining*. 2023 Jul 13;16(1):20. Available from: https://doi.org/10.1186/s13040-023-00339-9

[18] Gao Z, Brantley K, Joachims T. Reviewer2: Optimizing review generation through prompt generation. *arXiv preprint arXiv:2402.10886*. 2024 Feb 16. Available from: https://doi.org/10.48550/arXiv.2402.10886

[19] Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. *Nature*. 2023 Feb 6;614(7947):214–6. Available from: https://doi.org/10.1038/d41586-023-00340-6

[20] Kalra MP, Mathur A, Patvardhan C. Detection of AI-generated Text: An Experimental Study. *In2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)* 2024 Jul 27 (pp. 552-557). IEEE. Available from: https://ieeexplore.ieee.org/abstract/document/10731116

[21] [Online Available]: The 2025 AI Index Report. Available from: https://hai.stanford.edu/ai-index/2025-ai-index-report

[22] [Online Available]: AI Detection and assessment – an update for 2025. Available from: https://nationalcentreforai.jiscinvolve.org/wp/2025/06/24/ai-detection-assessment-2025/