ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26SEP073

# Responsible Generative AI in Doctoral Education: Balancing Innovation, Academic Integrity, and Researcher Autonomy

Ranjeeta Basra
Independent Researcher, India
Phd in Public Administration, Indira Gandhi National Open University (IGNOU), New Delhi, India

Corresponding Author

**Abstract**
Generative artificial intelligence (AI) is increasingly used in higher and doctoral education for literature exploration, academic writing, coding, data analysis, and access to knowledge. This paper examines how generative AI can support doctoral education while preserving academic integrity, researcher autonomy, and independent scholarly judgement. An integrative literature review was conducted across recent scholarly studies, systematic reviews, editorial and publisher guidance, and institutional policy literature, focusing on AI literacy, cognitive offloading, authorship and disclosure, hallucinated citations, and doctoral supervision. The review indicates that generative AI can improve efficiency, accessibility, feedback, and research support, but unregulated use can encourage over-reliance, weaken critical engagement, obscure authorship and originality, and introduce unverifiable or fabricated information. Based on these findings, the paper proposes the TRUST framework, Transparency, Responsibility, Understanding, Supervision, and Tiered Use,as a conceptual governance model for responsible AI integration in doctoral education. The framework places human accountability and researcher agency at the centre of AI-supported scholarship and proposes differentiated boundaries according to research function and intellectual risk. The paper concludes that doctoral programmes should move beyond blanket prohibition or AI detection alone and instead develop transparent, literacy-based, and supervisory approaches that treat generative AI as an assistive technology subject to human oversight.

**Keywords:** Generative AI; AI in Education; Doctoral Education; Academic Integrity; AI Literacy; Researcher Autonomy; Responsible AI

## 1. Introduction

Doctoral education has long been organised around the principle that a thesis represents an original, independently defensible contribution to knowledge, produced through sustained critical engagement between the candidate and the field. The emergence of generative AI tools , large language models capable of producing fluent text, summarising literature, drafting code, and assisting with statistical analysis , has begun to reshape almost every stage of this process, from the framing of a research problem to the final defence of a dissertation. Surveys of student AI use report adoption rates that have risen sharply within a short period, with recent higher-education data indicating that the great majority of students now use at least one generative AI tool in their academic work (Azevedo et al., 2025; Humble, 2025).

For doctoral candidates specifically, AI tools offer practical assistance with literature searching, synthesis of large bodies of scholarship, language editing, coding for data analysis, and the structuring of complex arguments. These capabilities are attractive to candidates who face time pressure, language barriers, or the sheer volume of material a doctoral thesis must engage with. At the same

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26SEP073

time, the same capabilities raise questions that strike at the heart of what a doctorate is meant to certify: independent scholarly capability. If literature synthesis, argument construction, or data interpretation is substantially produced by an AI system, the intellectual ownership and authenticity of the resulting work becomes uncertain (Bittle & El-Gayar, 2025; Eke, 2023).

This tension is not simply a matter of detecting misconduct. It concerns the deeper question of how doctoral training cultivates the analytical reasoning, creativity, and scholarly judgement that AI tools can, if used uncritically, allow candidates to bypass rather than develop (Risko & Gilbert, 2016). Universities, journal publishers, and professional bodies have responded with an expanding set of policies on disclosure, permissible use, and authorship, but these policies remain uneven across institutions and disciplines, and doctoral education in particular still lacks a settled, widely adopted framework (Humble, 2025; Azevedo et al., 2025).

This paper addresses that gap. It synthesises recent scholarly, institutional, and publisher-level literature on generative AI in higher education and research to examine, specifically for the doctoral context: (a) the opportunities AI offers across the stages of thesis research and writing; (b) the risks it poses to originality, critical thinking, and academic integrity; (c) the evolving norms around authorship disclosure and transparency; and (d) the institutional and pedagogical mechanisms , AI literacy, supervisory oversight, and disclosure protocols , through which responsible use can be supported. Building on this synthesis, the paper proposes an integrated framework intended to help doctoral programmes move past a binary choice between prohibition and unregulated adoption, toward a position in which AI use strengthens rather than substitutes for the candidate's own scholarly contribution.

The educational question is therefore broader than whether AI should be permitted. Doctoral education is simultaneously a learning environment and a process for developing independent researchers. Responsible AI integration must consequently protect the developmental function of doctoral study: candidates should learn to formulate problems, evaluate evidence, make methodological choices, construct arguments, and defend conclusions while using AI selectively as an assistive technology. This paper treats that principle as the central educational criterion for responsible generative AI use.

## 2. Literature Review

### 2.1 Generative AI in Higher and Doctoral Education

Generative AI tools became widely accessible to the academic community from late 2022 onward, and their use has expanded far beyond casual writing assistance into core research activities such as literature review, coding, and manuscript preparation (Golan et al., 2023; Graf & Bernardi, 2023). Reviews of the literature note that this expansion has occurred faster than institutional policy has been able to keep pace with, producing what several authors describe as a period of reactive rather than proactive governance (García-López & Trujillo-Liñán, 2025). For doctoral researchers in particular, AI tools are increasingly embedded in reference management, qualitative coding software, and manuscript-preparation workflows, meaning that a candidate may be using AI extensively without necessarily treating each instance as a discrete ethical decision.

### 2.2 Educational and Research Opportunities

A substantial body of literature documents genuine benefits of AI adoption in research settings. AI tools can accelerate literature discovery and synthesis, assist with language editing for non-native English speakers, support exploratory data analysis, and lower barriers to participation for researchers with limited institutional resources (Golan et al., 2023). In graduate research contexts specifically, studies report that structured, reflective use of generative AI is associated with improved research self-efficacy and, under the right conditions, can support rather than replace higher-order thinking such as evaluation and synthesis (Gonsalves, 2024). Workshop-level discussions of AI in academic search similarly emphasise its potential to support summarisation, recommendation, and synthesis functions that extend traditional literature-review practice, provided these functions remain transparent and auditable (Bittle & El-Gayar, 2025).

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26SEP073

### 2.3 Academic Integrity, Originality, and Educational Risk

Set against these benefits is a growing body of evidence that unregulated AI use threatens the originality and authenticity that doctoral and other advanced research is meant to demonstrate. Systematic reviews of generative AI and academic integrity report that while AI can enhance engagement and efficiency, it simultaneously introduces significant risks of academic dishonesty, including undisclosed authorship of substantial portions of text (Bittle & El-Gayar, 2025). Assessment-focused studies similarly find that generative AI can be used to circumvent the very processes designed to demonstrate a student's or candidate's own capability, complicating both detection and the marking or examination process itself (Kofinas et al., 2025). Eke (2023) frames this as a structural threat to academic integrity rather than an isolated behavioural problem, arguing that the ease and fluency of AI-generated text make conventional norms of authorship and originality difficult to enforce without new safeguards.

Detection-based responses to this risk have shown notable limitations. Reviews of AI-text detection tools report variability in accuracy and a susceptibility to circumvention through adversarial prompting, alongside a documented tendency toward false positives for multilingual or non-native English writers (as synthesised in Frontiers in Education, 2026). This body of evidence suggests that a purely technological, detection-based approach to safeguarding integrity is unlikely to be sufficient on its own, and may introduce new forms of unfairness even as it attempts to address one.

### 2.4 Cognitive Offloading and Researcher Autonomy

A closely related concern is the effect of AI reliance on the cognitive and analytical skills that doctoral training is intended to cultivate. The concept of cognitive offloading , the use of external tools to reduce the mental effort required for a task , provides a useful theoretical lens here (Risko & Gilbert, 2016). Recent empirical work applying this concept to generative AI reports that habitual, unstructured reliance on AI tools is associated with reduced engagement in deep reasoning and a decline in independent analytical effort, particularly on unfamiliar or ill-structured tasks (Gonsalves, 2024). At the same time, this literature converges on a more nuanced conclusion: the cognitive consequences of AI use are shaped less by whether a tool is used and more by how it is used. Research distinguishing between passive AI consumption, assisted learning, and reflective AI collaboration finds that these three modes of interaction produce markedly different outcomes, ranging from reduced reflective engagement under passive use to enhanced analytical reasoning under genuinely reflective, dialogic use (Gonsalves, 2024).

A related concern for doctoral education is that heavy or passive reliance on generative AI may create an appearance of independent reasoning while reducing the candidate's direct engagement with the cognitive work of evaluating, questioning, and revising ideas. Recent educational research suggests that the effects of AI on critical thinking depend strongly on the mode of use, reinforcing the importance of reflective rather than passive collaboration (Gonsalves, 2024).

### 2.5 Authorship, Disclosure, and Scholarly Accountability

As AI tools have become embedded in manuscript preparation, publishing bodies have converged on a broadly consistent position regarding authorship and disclosure. The Committee on Publication Ethics (COPE), the International Committee of Medical Journal Editors (ICMJE), the World Association of Medical Editors (WAME), and other editorial associations agree that AI tools cannot be listed as authors because they cannot take responsibility for a work's accuracy, originality, or integrity, nor can they assert the absence of conflicts of interest (COPE, 2023; ICMJE, 2026). These bodies instead require that authors disclose whether and how AI tools were used, typically distinguishing between AI used for writing assistance , reported in acknowledgements , and AI used for data collection, analysis, or coding, which is reported in the methods section (ICMJE, 2026). The ICMJE's most recent recommendations reinforce that human authors retain full responsibility for manuscript content regardless of AI involvement, and that undisclosed AI use may itself constitute a form of scientific misconduct (ICMJE, 2026).

Individual publishers have translated these principles into practical policy. Major publishers permit generative AI in manuscript preparation on condition that authors provide oversight, verify all outputs, and disclose their use, while treating human authors as fully accountable for the final content (as summarised in comparative policy analyses; arXiv, 2026). Structured disclosure frameworks, such

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26SEP073

as the Artificial Intelligence Disclosure (AID) framework, have been proposed to standardise how AI contributions are reported across research, writing, and educational contexts, drawing on contributor-role taxonomies already used to describe human co-authorship (Weaver, 2024). For doctoral research, which typically culminates in both a thesis and subsequent journal publications, this convergence toward disclosure-based governance , rather than prohibition , offers a template that doctoral programmes can adapt for thesis submission itself.

### 2.6 Institutional AI Governance and AI Literacy

At the institutional level, universities have moved quickly, if unevenly, to formulate AI-use policies. Comparative document analyses of university guidelines find that most institutions now explicitly extend their policies to doctoral candidates alongside undergraduate and taught postgraduate students, and that a majority frame AI's opportunities as outweighing its risks, provided use is transparent (Humble, 2025). Analyses of large public university systems highlight recurring themes across institutional policy: requirements for AI literacy among both faculty and students, explicit academic-integrity provisions, and guidance on course or research design that reduces incentives for undisclosed AI substitution (Azevedo et al., 2025).

For doctoral supervision specifically, emerging guidance emphasises the supervisor's role in setting expectations, requiring written disclosure of AI use, and integrating verification checkpoints into the supervisory process rather than relying solely on end-of-thesis detection (Azevedo et al., 2025; Humble, 2025). Broader governance frameworks for higher education similarly recommend embedding AI literacy , covering bias awareness, verification routines, and documentation habits , directly into research-methods training and supervisory meetings, as a more durable safeguard than after-the-fact policing (Azevedo et al., 2025; Humble, 2025). Taken together, this literature suggests that responsible AI governance in doctoral research is converging on a small number of recurring elements: transparency through disclosure, human accountability for final content, structured AI literacy training, and supervisory or institutional oversight built into the research process rather than bolted on at submission.

## 3. Objectives

Building on the literature reviewed above, this paper pursues the following objectives:

To examine the principal opportunities that generative AI tools offer across the stages of doctoral research, including literature exploration, academic writing, data analysis, and knowledge accessibility.

To critically analyse the risks that unregulated AI adoption poses to originality, critical thinking, researcher autonomy, and academic integrity in doctoral research.

To review current authorship, disclosure, and publishing-ethics norms relevant to AI use in scholarly work, and to assess their applicability to doctoral thesis research.

To synthesise institutional policy and AI-literacy literature relevant to doctoral supervision and governance.

To propose an integrated, practically applicable framework through which doctoral candidates, supervisors, and institutions can balance innovation with academic integrity.

## 4. Research Methodology

This paper adopts an integrative literature review design, appropriate for synthesising a rapidly evolving and cross-disciplinary body of work spanning higher-education policy, research ethics, cognitive science, and scholarly publishing (Bittle & El-Gayar, 2025). Rather than testing a specific empirical hypothesis, an integrative review is used here to consolidate disparate strands of evidence , institutional policy analyses, systematic reviews, publisher guidelines, and cognitive-science research on AI-assisted work , into a coherent conceptual account and a practically oriented framework.

Sources were identified through targeted searches of academic databases and publisher repositories for literature published primarily between 2023 and 2026, corresponding to the period of widespread generative AI adoption in higher education. Search terms combined generative AI with academic integrity, doctoral or graduate research, cognitive offloading, authorship disclosure, and institutional AI policy. Priority was given to peer-reviewed systematic reviews, publisher and editorial-association position statements (COPE, ICMJE, WAME), and institutional policy analyses, supplemented by

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26SEP073

empirical studies on cognitive effects of AI use. Sources describing purely undergraduate assessment contexts were included only where their findings on integrity, disclosure, or cognitive effects were plausibly transferable to doctoral research. As with any integrative review, this approach is interpretive rather than exhaustive, and the resulting framework should be read as a conceptual synthesis intended for further empirical testing rather than a validated instrument.

## 5. Findings from the Integrative Review

The literature reviewed in Section 2 establishes that AI adoption in doctoral research carries genuine risks alongside its benefits. This section examines those risks in closer detail, organised around five specific and interrelated problems that recur across the evidence: the erosion of critical thinking, ambiguity over original contribution, unresolved questions of authorship and ownership, concrete research-quality risks arising from AI hallucination, and the changing demands placed on doctoral supervision.

### 5.1 Decline in Critical Thinking and Intellectual Struggle

Doctoral research is fundamentally a process of sustained intellectual struggle: formulating a question, testing it against evidence, revising it in light of contradiction, and defending the resulting argument under scrutiny. This struggle is not incidental to the doctorate; it is the mechanism through which analytical reasoning and scholarly judgement are actually built. Generative AI tools can supply a plausible answer, argument, or synthesis almost instantly, which risks allowing a candidate to arrive at a conclusion without passing through the reasoning that would normally produce it. The cognitive-offloading literature reviewed earlier shows that this risk is not hypothetical: habitual, unreflective AI use is associated with reduced engagement in deep reasoning, particularly on unfamiliar or ill-structured problems of exactly the kind a doctorate is meant to pose (Risko & Gilbert, 2016; ScienceDirect, 2025). The related concept of epistemic confinement is especially concerning in this context, since a candidate may experience a strong subjective sense of having reasoned independently while, in fact, operating entirely within the analytical boundaries an AI system has already supplied (Gonsalves, 2024). Because doctoral examiners typically assess a candidate's capacity for independent argument during the viva or defence, this gap between the appearance and the substance of independent reasoning is not easily detected from the written thesis alone.

### 5.2 The Question of Original Contribution

Doctoral degrees are awarded, in nearly every disciplinary and national context, on the basis of a 'new and significant contribution to knowledge.' This requirement presupposes that the contribution can be meaningfully attributed to the candidate. Generative AI complicates this attribution in ways that current thesis-examination practice is not yet well equipped to resolve. When AI tools assist not only with language polishing but with structuring arguments, generating interpretive framings, or suggesting theoretical connections, it becomes genuinely difficult to identify where AI assistance ends and the candidate's own intellectual contribution begins. Systematic reviews of generative AI and academic integrity note this same difficulty at the level of individual assessments, finding that GenAI can be used in ways that make it hard to determine, after the fact, how much of a piece of work reflects independent effort (Bittle & El-Gayar, 2025; Kofinas et al., 2025). For doctoral research the stakes are higher than for a single assignment, since the entire credential rests on a judgement about originality that supervisors and examiners are asked to certify.

### 5.3 Authorship, Ownership, and Attribution Ambiguity

Closely related to the question of original contribution is the unresolved question of how AI-assisted contributions should be attributed within a thesis. Publisher and editorial bodies have reached a relatively settled position for journal articles: AI tools cannot be listed as authors or contributors, and disclosure is required wherever they are used (COPE, 2023; ICMJE, 2026). Doctoral thesis regulations, by contrast, rarely specify an equivalent disclosure requirement, leaving candidates and supervisors to apply informal or inconsistent standards. This creates a practical ambiguity: a candidate may treat AI-assisted drafting as equivalent to using a grammar checker or a supervisor's feedback, while an examiner may treat the same assistance as a substantive, undisclosed contribution to the argument. Without a shared standard for what must be disclosed and how, doctoral programmes risk

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26SEP073

applying authorship expectations to theses that are less rigorous than those already adopted by the journals in which the same candidates will eventually seek to publish.

### 5.4 Research Quality Risks: Hallucinated and Fabricated Citations

A further, more concrete problem concerns the factual reliability of AI-assisted literature work. Generative AI models are known to produce hallucinated citations: references that are plausibly formatted, attributed to real researchers, and topically appropriate, but that do not correspond to any actual publication. Early empirical audits of this problem found that a large proportion of AI-generated references in literature-review tasks were entirely fabricated, with fabrication rates reported as high as 55% for earlier models and still around 18% for more advanced models (Walters & Wilder, 2023). More recent large-scale evidence indicates that the problem is worsening in the published literature itself: an audit of 2.5 million biomedical papers found that the rate of papers containing at least one fabricated citation rose roughly twelve-fold in three years, from approximately one in 2,828 papers in 2023 to one in 277 papers in early 2026, with the increase closely tracking the growing use of generative AI in manuscript preparation (Topaz et al., 2026). Crucially, these fabricated citations are typically not obviously defective: they are correctly formatted, attributed to real scholars, and carry plausible publication details, which makes them difficult for supervisors and examiners to detect through ordinary reading (Topaz et al., 2026). Some legal scholars have argued that hallucinated citations may meet the threshold for research misconduct when they function as evidentiary support for a claim and the researcher shows indifference to the risk of fabrication (Accountability in Research, 2026). For doctoral candidates, whose reference lists commonly run into the hundreds of entries, this risk is amplified: unverified AI-assisted literature searching can silently introduce citation errors that undermine the credibility of an entire thesis, even when the candidate had no intention to deceive.

### 5.5 The Changing Role of the Doctoral Supervisor

Taken together, these four challenges point to a shift already visible in emerging institutional guidance: the doctoral supervisor's role is expanding from assessing a finished written product to actively verifying the process by which that product was produced. Guidance on supervising with AI recommends that supervisors require written disclosure of AI use, build verification checkpoints into the candidacy timeline, and treat research supervision as an ongoing quality system rather than a single end-of-thesis review (Azevedo et al., 2025; Humble, 2025). Consistent with this, broader governance analyses recommend that research supervision incorporate scheduled verification, provenance tracking of AI-assisted content, and structured human-AI collaboration checkpoints throughout candidacy (Azevedo et al., 2025; Humble, 2025). In practice, this means supervisors and examiners increasingly need to assess a candidate's underlying reasoning and decision-making directly , for example, through oral questioning that probes how and why particular interpretive or methodological choices were made , rather than relying on the written thesis and its reference list as sufficient evidence of independent scholarly capability.

## 6. Discussion

### 6.1 Opportunities and Educational Value

The reviewed literature confirms that AI tools offer doctoral candidates meaningful support across the research lifecycle: faster and broader literature synthesis, assistance with language and structural clarity in writing, support for exploratory data analysis and coding, and improved accessibility for candidates working in a second language or with limited institutional resources (Golan et al., 2023; Graf & Bernardi, 2023). These benefits are genuine and are unlikely to diminish as tools continue to mature. The central question for doctoral education is therefore not whether these tools should be used, but under what conditions their use strengthens, rather than substitutes for, the candidate's own scholarly contribution.

### 6.2 Key Challenges and Risks

Three challenges recur across the literature reviewed. First, unregulated AI use threatens originality and complicates the assessment of a candidate's independent contribution, a concern documented across systematic reviews of GenAI and academic integrity (Bittle & El-Gayar, 2025; Kofinas et al.,

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26SEP073

2025). Second, habitual or passive AI reliance is associated with cognitive offloading and, in some documented cases, an illusion of independent reasoning that masks a narrowing of analytical boundaries (Risko & Gilbert, 2016; ResearchGate, 2026) ,a risk with particular significance for doctoral research, whose central purpose is to extend rather than reproduce existing boundaries of knowledge. Third, authorship and disclosure norms remain inconsistently applied at the doctoral level even though publishing bodies have converged on relatively clear principles for journal manuscripts (ICMJE, 2026; Weaver, 2024), leaving a gap between publication-stage governance and thesis-stage governance for the same candidate and often the same content.

**Table 1. TRUST Framework for Responsible Generative AI in Doctoral Education**

The TRUST framework is proposed as a conceptual model for responsible integration of generative AI into doctoral education. It translates the review findings into five governance principles that preserve human accountability and researcher agency.

### 6.3 The TRUST Framework

| Element | Purpose | Risk Addressed | Implementation |
|---|---|---|---|
| Transparency | Disclosure | Hidden or ambiguous AI contribution | AI-use statement and provenance record |
| Responsibility | Accountability | Unverified output and authorship ambiguity | Human verification and final-author responsibility |
| Understanding | AI literacy | Bias, hallucination and uncritical dependence | Structured training and verification skills |
| Supervision | Governance | Weak oversight during research development | Supervisory checkpoints and defence of research decisions |
| Tiered Use | Risk-based boundaries | Overuse in high-risk intellectual tasks | Differentiate permissible AI use by research stage and function |

To address this gap, this paper proposes an integrated framework- TRUST, comprising five interdependent elements that doctoral programmes can adapt to their own disciplinary and institutional context.

Transparency: Candidates disclose where and how AI tools were used at each stage of the research (literature review, analysis, writing), following the acknowledgement/methods distinction already established in publisher guidelines (ICMJE, 2026).

Responsibility: Consistent with COPE and ICMJE positions on authorship, the candidate retains full accountability for the accuracy, originality, and integrity of the thesis regardless of AI involvement; AI cannot be listed as a contributor or co-author (COPE, 2023; ICMJE, 2026).

Understanding: Structured AI-literacy training is embedded in doctoral coursework and supervisory meetings, covering not only tool operation but critical evaluation of AI outputs, bias awareness, and verification practices (Azevedo et al., 2025; Humble, 2025).

Supervision: Supervisors build verification checkpoints into the candidacy timeline, for example, requiring candidates to explain and defend AI-assisted analysis or drafting at milestone reviews, rather than relying solely on end-of-thesis detection software (Azevedo et al., 2025; Humble, 2025).

Tiered use: Institutions distinguish permissible levels of AI involvement by research stage and function, for instance, treating AI-assisted literature scoping or language editing differently from AI-generated data interpretation or original argumentation, which should remain substantially the candidate's own work.

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26SEP073

This framework is deliberately structured around disclosure and shared accountability rather than prohibition or detection alone, reflecting the consensus emerging from publisher and institutional policy literature that transparency-based governance is both more enforceable and more consistent with how AI tools are actually being integrated into research practice (García-López & Trujillo-Liñán, 2025; Humble, 2025). It is intended as a starting point for adaptation, not a fixed set of rules, since appropriate tiers of use will vary meaningfully across disciplines, for example, between computational fields where AI-assisted coding is well established and humanities fields where original argumentation is the primary scholarly contribution.

## 7. Implications and Recommendations

Universities and doctoral schools should issue explicit, discipline-sensitive AI-use policies for doctoral candidates, rather than relying on generic undergraduate-level guidance, and should update these policies as tools and norms evolve (Humble, 2025; Azevedo et al., 2025).

Doctoral programmes should require a standardised AI-disclosure statement at thesis submission, modelled on existing journal-level disclosure practice, specifying which tools were used and for which research functions (ICMJE, 2026; Weaver, 2024).

Supervisors should incorporate AI-literacy discussions and verification checkpoints into regular supervisory meetings rather than treating AI governance as a one-time compliance exercise (Azevedo et al., 2025; Humble, 2025).

Doctoral training programmes should teach candidates to use AI for reflective collaboration, questioning, checking, and extending AI outputs, rather than passive consumption, given the documented cognitive differences between these modes of use (Gonsalves, 2024).

Institutions should invest in AI-literacy training for both candidates and supervisors, recognising that responsible use depends as much on human judgement and verification habits as on institutional rules (Azevedo et al., 2025; Humble, 2025).

Detection technologies should be used cautiously and in combination with, not as a substitute for, disclosure-based governance, given documented accuracy limitations and risks of disadvantaging non-native English writers (Slimi, 2026).

## 8. Limitations and Future Research

As an integrative literature review, this paper synthesises existing evidence and policy rather than generating new primary data, and its proposed framework has not yet been empirically tested with doctoral candidates or supervisors. Much of the underlying literature on cognitive offloading and academic integrity draws on undergraduate and taught-postgraduate populations, and the extent to which these findings transfer directly to doctoral-level research, which typically involves more prolonged, supervised, and specialised inquiry, warrants direct empirical investigation. Future research should test the TRUST framework's elements through mixed-methods studies involving doctoral candidates across disciplines, evaluate the actual effectiveness of disclosure-based governance relative to detection-based approaches, and examine how disciplinary norms in fields such as the humanities, social sciences, and STEM shape appropriate thresholds for AI involvement.

## 9. Conclusion

Generative AI is now embedded in the everyday practice of doctoral research, offering genuine gains in efficiency, accessibility, and research support. The evidence reviewed here shows that these gains coexist with well-documented risks to originality, critical thinking, and academic integrity, and that neither uncritical adoption nor blanket prohibition offers a workable path forward. The more durable response emerging across publisher, institutional, and scholarly literature is one grounded in transparency, shared human accountability, structured literacy-building, active supervision, and differentiated levels of permissible use. The TRUST framework proposed in this paper offers one way of organising these elements for doctoral education specifically, where the stakes for originality and independent scholarly capability are especially high. The task facing doctoral education is therefore not to resist AI, but to build the institutional and pedagogical infrastructure through which it strengthens, rather than substitutes for, the candidate's own intellectual contribution.

## Conflict of Interest Statement

The author declares no conflict of interest.

## AI Use Disclosure

Generative AI tools were used during manuscript development for language refinement and editorial support. The author retained responsibility for the study concept, literature interpretation, framework development, factual verification, and final manuscript content.

## References

1.Resnik, D. B., & Hosseini, M. (2026). Hallucinated citations produced by generative artificial intelligence may constitute research misconduct when citations function as data in scholarly papers. 2.Accountability in Research. https://doi.org/10.1080/08989621.2026.2645390

Azevedo, L., Robles, P., Best, E., & Mallinson, D. J. (2025). Institutional policies on artificial intelligence in higher education: Frameworks and best practices for faculty. New Directions for Adult and Continuing Education, 2025(188), 70–78. https://doi.org/10.1002/ace.70013

3.Bittle, K., & El-Gayar, O. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. Information, 16(4), 296. https://doi.org/10.3390/info16040296

4.Committee on Publication Ethics. (2023). COPE position statement: Authorship and AI tools. https://doi.org/10.24318/cCVRZBms

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

6.Evangelista, E. D. L. (2025). Ensuring academic integrity in the age of ChatGPT: Rethinking exam design, assessment strategies, and ethical AI policies in higher education. Contemporary Educational Technology, 17(1), ep559. https://doi.org/10.30935/cedtech/15775

7.Gallent Torres, C., Zapata González, A., & Ortego Hernando, J. L. (2023). El impacto de la inteligencia artificial generativa en educación superior: una mirada desde la ética y la integridad académica. RELIEVE, 29(2). https://doi.org/10.30827/relieve.v29i2.29134

8.García-López, I. M., & Trujillo-Liñán, L. (2025). Ethical and regulatory challenges of generative AI in education: A systematic review. Frontiers in Education, 10, 1565938. https://doi.org/10.3389/feduc.2025.1565938

9.Golan, R., Reddy, R., Muthigi, A., & Ramasamy, R. (2023). Artificial intelligence in academic writing: A paradigm-shifting technological advance. Nature Reviews Urology, 20, 327–328. https://doi.org/10.1038/s41585-023-00746-x

10.Gonsalves, C. (2024). Generative AI's impact on critical thinking: Revisiting Bloom's taxonomy. Journal of Marketing Education, 48(1), 4–19. https://doi.org/10.1177/02734753241305980

Graf, A., & Bernardi, R. E. (2023). ChatGPT in research: Balancing ethics, transparency and advancement. Neuroscience, 515, 71–73. https://doi.org/10.1016/j.neuroscience.2023.02.008

11.Humble, N. (2025). Higher education AI policies,A document analysis of university guidelines. European Journal of Education, 60(3), e70214. https://doi.org/10.1111/ejed.70214

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

12.Kofinas, A. K., Tsay, C.-H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. British Journal of Educational Technology, 56, 2522–2549. https://doi.org/10.1111/bjet.13585

13.Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Slimi, Z. (2026). A systematic critical review of generative AI's impact on authorship, pedagogy, and integrity (2023–2025). Frontiers in Education, 11, 1769680. https://doi.org/10.3389/feduc.2026.1769680

14.Topaz, M., Roguin, N., Gupta, P., Zhang, Z., & Peltonen, L.-M. (2026). Fabricated citations: An audit across 2.5 million biomedical papers. The Lancet, 407(10541), 1779–1781. https://doi.org/10.1016/S0140-6736(26)00603-3

15.Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. Scientific Reports, 13, 14045. https://doi.org/10.1038/s41598-023-41032-5

Weaver, K. D. (2024). The Artificial Intelligence Disclosure (AID) framework: An introduction. College & Research Libraries News, 85(10), 407–409. https://doi.org/10.5860/crln.85.10.407