Science Journal of Education
2026, Vol. 14, No. 3, pp. 68–80
https://doi.org/10.11648/j.sjedu.20261403.14

SciencePG
Science Publishing Group

Research Article

AI Tools and Academic Integrity in Postgraduate Research: Addressing Opportunities, Boundaries, and Ethical Frameworks

Mohammed Zeinu Hassen*

Department of Social Sciences, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

Abstract

The proliferation of Artificial Intelligence (AI) tools, encompassing Large Language Models (LLMs), AI-driven literature review platforms, automated statistical software, and generative writing assistants, has fundamentally altered the landscape of postgraduate academic research. This article provides a comprehensive examination of how AI tools intersect with the principles of academic integrity in doctoral and master's-level research programmes. Drawing on emerging institutional policies, ethical philosophy, and empirical findings from higher-education research, the article explores the spectrum of AI use cases in postgraduate work: from legitimate productivity enhancements to practices that undermine the foundational purpose of advanced scholarship. The article argues that the central challenge is not AI itself but the absence of clear, consistent, and educationally grounded frameworks for its use. It examines how traditional definitions of plagiarism, authorship, and original contribution are being renegotiated in the AI era; surveys international policy responses from leading research universities; and proposes a tiered ethical framework to guide postgraduate researchers, supervisors, and institutions. Special attention is given to the particular vulnerabilities of postgraduate research, including pressure to publish, cross-cultural competence gaps, and inadequate supervisory guidance, and to how AI detection technologies are reshaping academic misconduct proceedings. The article concludes that preserving academic integrity in the age of AI requires not prohibition but principled engagement: a shared commitment to transparency, disciplinary literacy, and the cultivation of genuine intellectual contribution.

Keywords

Artificial Intelligence in Research, Generative AI, Large Language Models, Research Ethics, Academic Integrity, Postgraduate Studies, Doctoral Education, Research Authorship

1. Introduction

In November 2022, the public release of ChatGPT by OpenAI triggered what many in academia described as an inflection point, a moment at which the long-theorised disruption of Artificial Intelligence (AI) to human intellectual work became immediate, tangible, and unavoidable. Within weeks, universities around the world were fielding urgent queries from bewildered faculty members and graduate students alike: What exactly had changed? What was now permissible? What

*Correspondence: Mohammed Zeinu Hassen (mohammed.zeinu@aastu.edu.et)

Received: 22 May 2026; **Accepted:** 2 June 2026; **Published:** 18 June 2026

constituted cheating?

These questions were not entirely new. The history of higher education is punctuated by technologies that challenged prevailing notions of academic honesty. The printing press, the photocopier, the Internet, and contract cheating websites all forced institutions to adapt their definitions of scholarly integrity. Yet generative AI presents a challenge of a qualitatively different order. Unlike earlier technologies that facilitated the reproduction or retrieval of existing content, AI tools now generate original-appearing text, synthesise literature, propose research designs, produce code, and even offer statistical interpretations. These are the very cognitive acts that define postgraduate scholarship.

The stakes are particularly high at the postgraduate level. Doctoral and master's programmes are not simply conduits for credential acquisition; they are the primary mechanisms through which societies produce new knowledge. The thesis, the dissertation, and the peer-reviewed article represent the postgraduate researcher's authentic intellectual contribution, the demonstration that they can independently identify a problem, situate it within existing knowledge, design and execute an investigation, and draw defensible conclusions. If AI tools are used in ways that hollow out this process, the integrity of the entire knowledge-production enterprise is imperilled.

Yet the challenge is more nuanced than a simple prohibition would suggest. AI tools, used appropriately, can help researchers transcend linguistic barriers, overcome cognitive bottlenecks, accelerate literature searches, and improve the clarity of their writing. For researchers working in English as an additional language, a substantial proportion of the global postgraduate community, AI writing assistance may represent not an unfair advantage but a necessary accommodation. The question, then, is not whether AI should be used in postgraduate research, but how, when, by whom, and under what conditions of transparency.

This article provides a rigorous and wide-ranging examination of these questions. It proceeds in eleven sections: situating AI tools within the contemporary research environment; redefining what academic integrity means in this new context; mapping the spectrum of AI use cases; examining the specific integrity challenges around plagiarism and authorship; surveying institutional policy responses globally; assessing AI detection technologies; examining equity and cross-cultural dimensions; articulating the responsibilities of supervisors and graduate schools; proposing a practical ethical framework; and offering conclusions. Throughout, the aim is to move beyond moral panic and institutional defensiveness toward a principled, evidence-based account of what academic integrity demands in the age of artificial intelligence.

2. The AI Landscape in Postgraduate Research

2.1. The Tools Reshaping Research Practice

The term 'AI tools' encompasses a remarkably heterogeneous set of technologies, each with distinct capabilities, limitations, and integrity implications. Understanding this landscape is a prerequisite for any coherent policy or ethical analysis.

Large Language Models (LLMs), including OpenAI's Generative Pre-trained Transformer (GPT)-4o, Anthropic's Claude, Google's Gemini, and Meta's LLaMA series, are transformer-based neural networks trained on vast corpora of human-generated text [17]. When prompted, they generate fluent, contextually coherent prose that can range from surprisingly accurate to confidently wrong. Their capabilities include drafting text, summarising documents, explaining concepts, generating hypotheses, producing code in multiple programming languages, and translating between languages. Their limitations include the tendency to hallucinate, to fabricate citations, statistics, and facts with the same rhetorical confidence as accurate information, and an inability to access real-time information unless augmented with retrieval tools [25].

AI-powered literature review tools represent a second important category. Platforms such as Elicit, Consensus, Scite, Research Rabbit, and Semantic Scholar use machine learning to help researchers navigate the exponentially expanding body of academic literature. They can identify relevant papers, surface thematic patterns, highlight empirical findings, and map citation networks. Unlike general-purpose LLMs, these tools are specifically designed for academic research contexts and carry a lower risk of hallucination because they retrieve rather than generate.

Statistical and data analysis AI tools, including AI-assisted features in Statistical Package for the Social Sciences (SPSS), R, and Python libraries, as well as dedicated platforms like Julius AI, can help researchers identify appropriate analytical methods, interpret results, and generate visualisations. While these tools can dramatically accelerate quantitative work, they also introduce risks: researchers who do not understand the underlying statistical logic may apply inappropriate methods or misinterpret outputs, undermining the rigour of their findings.

Finally, AI-powered writing and editing tools, including Grammarly, ProWritingAid, and Wordtune, occupy a more familiar and generally less contested space, though their increasingly powerful suggestions blur the line between grammar correction and substantive ghostwriting.

Table 1. Major AI tool categories in postgraduate research and their primary use cases.

Tool Category	Examples	Primary Research Use
Large Language Models	ChatGPT, Claude, Gemini, Copilot	Drafting, summarising, brainstorming, coding
Literature Review AI	Elicit, Consensus, Research Rabbit, Scite	Paper discovery, synthesis, citation mapping
Data Analysis AI	Julius AI, SPSS AI, Code Interpreter	Statistical analysis, visualisation, interpretation
Writing Assistants	Grammarly, Wordtune, ProWritingAid	Editing, paraphrasing, language polishing
Reference Managers (AI)	Zotero AI, Mendeley AI	Citation management, annotation
Image/Figure Generation	DALL-E, Midjourney, BioRender AI	Research figures, concept illustrations

2.2. Adoption Rates and Patterns

Empirical data on postgraduate AI adoption is still emerging, but early surveys reveal striking patterns. A 2024 survey by Digital Education Council involving over 18,000 students across 200 countries found that approximately 86 percent of university students were using AI tools, with postgraduate students reporting higher rates of use and more sophisticated applications than undergraduates [9]. Among doctoral students specifically, use was concentrated in three areas: writing and editing assistance, literature searching, and coding and data analysis.

Adoption is highly uneven across disciplines. Students in Science, Technology, Engineering, and Mathematics (STEM) fields, particularly computer science, engineering, and the life sciences, have integrated AI coding tools and data analysis platforms deeply into their workflows, often with the tacit or explicit approval of their supervisors. In the humanities and interpretive social sciences, adoption is more cautious and contested, reflecting disciplinary norms that place greater intrinsic value on the researcher's own prose as the primary medium of intellectual contribution.

The pace of change is itself a governance challenge. Institutional policies, ethical frameworks, and supervisory norms developed over decades are struggling to keep pace with the monthly release cycles of AI tools. Many postgraduate researchers find themselves operating in a policy vacuum: their institutions have issued broad statements about academic integrity but have provided little specific guidance on AI use in research contexts [14]. In this vacuum, individual practice is shaped more by peer norms, supervisor attitudes, and personal ethics than by institutional rules.

3. Redefining Academic Integrity in the AI Era

3.1. Traditional Foundations

Academic integrity is conventionally understood through the lens of five core values articulated by the International Center for Academic Integrity (ICAI) [10]: honesty, trust, fairness, respect, and responsibility. In the context of postgraduate research, these values have historically been operationalised through prohibitions on plagiarism, fabrication, falsification, and contract cheating, along with expectations of proper attribution, methodological transparency, and honest reporting of findings [3].

These principles were developed in an era when the primary integrity risks were human in origin: a student copying text from a book, fabricating experimental data, or paying someone else to write their thesis. The fundamental assumption was that the intellectual work product, the words, the analysis, the argument, could and should be attributed to a human author who bore full responsibility for its accuracy and originality.

3.2. AI as a Category Challenge

Generative AI disrupts this framework in ways that are not merely practical but conceptual [5]. When a doctoral student asks an LLM to 'write a literature review on the social determinants of health,' the resulting text is original in the narrow sense that it was not previously written. It has not been plagiarised from any existing document. Yet it is also not the product of the student's own reading, synthesis, and intellectual judgement. It exists in a novel categorical space: generated rather than composed, plausible rather than reliably accurate, and attributable to no individual human intelligence.

This categorical novelty destabilises several of academic integrity's most foundational concepts. Plagiarism, traditionally defined as presenting another person's work as one's own, becomes conceptually strained when applied to AI-generated text, because no person wrote it. Authorship, the attribution of intellectual responsibility to one or more individuals, becomes ambiguous when AI has contributed substantially to conception, drafting, or analysis. Originality, the criterion by which a doctoral thesis is judged to make a new contribution to knowledge, becomes contestable when the argument was partly constructed by a system trained on existing knowledge.

"The question is not whether AI-generated text is plagiarism in the traditional sense. The question is whether using it represents a failure to engage in the intellectual processes that postgraduate education is designed to develop. That failure is the real integrity violation." (Professor of Research Ethics, University of Edinburgh, 2024, cited in [4]).

3.3. A Revised Conceptual Framework

Several scholars have proposed conceptual revisions to academic integrity frameworks in response to AI. One influential approach, proposed by Eaton [5], distinguishes between process integrity, the honesty of the researcher's engagement with the intellectual work, and product integrity, the accuracy and originality of the final output. This distinction is useful because it captures what is genuinely at stake in AI use: even if the final text is accurate and original in a technical sense, its production through AI shortcuts may represent a failure of process integrity.

A second useful distinction is between AI as a tool and AI as a substitute. Just as researchers have always used tools, such as statistical software, citation managers, and spell-checkers, without their work being considered inauthentic, AI tools used to assist rather than replace the researcher's intellectual agency may be compatible with academic integrity [12]. The critical question is whether the researcher retains genuine understanding, judgement, and responsibility for the work they submit.

A third dimension concerns transparency. Even if a given use of AI would be permissible under the applicable rules, failure to disclose that use represents a distinct integrity violation, a form of deception about the nature of the research process. Transparency norms are therefore a necessary component of any revised integrity framework, regardless of where the substantive permissibility lines are drawn.

Core Integrity Principles Revised for the AI Era

1) Process integrity: Genuine intellectual engagement cannot be delegated to AI.
2) Epistemic responsibility: The researcher must understand and take ownership of all claims made.
3) Transparency: All significant AI contributions must be disclosed.
4) Proportionality: Permitted AI use should be proportionate to the research task.
5) Discipline-specificity: Standards vary legitimately across fields and must be explicitly taught.

4. AI Use Cases: A Spectrum of Practice

Not all uses of AI in postgraduate research raise the same integrity concerns. It is analytically useful to map AI use cases along a spectrum from clearly permissible to clearly prohibited, with a substantial middle ground that is contextually dependent and requires explicit institutional and supervisory guidance [18].

4.1. Clearly Permissible Uses

The following uses of AI in postgraduate research are generally considered compatible with academic integrity, provided they are appropriately disclosed:

1) Grammar, spelling, and style checking: Using tools such as Grammarly or built-in word-processor corrections to improve the mechanical quality of one's prose does not undermine intellectual contribution.
2) Literature discovery: Using AI-powered search tools to identify relevant papers, map citation networks, or surface overlooked literature supplements rather than replaces the researcher's critical engagement with sources.
3) Data organisation and formatting: Using AI to clean datasets, reformat bibliographies, or organise qualitative data for coding does not compromise the integrity of the analysis.
4) Code debugging and optimisation: For researchers who write code in R, Python, or other languages, using AI tools such as GitHub Copilot to identify bugs or suggest more efficient implementations is analogous to consulting technical documentation [19].
5) Translation support: For researchers working in a non-native language, using AI translation tools as a starting point, followed by careful human review, addresses a genuine equity concern without substituting for substantive intellectual work.
6) Brainstorming and ideation: Using AI as a sounding board to generate and test ideas, stimulate lateral thinking, or identify gaps in an argument, where the researcher then critically evaluates and develops those ideas themselves.

4.2. Contextually Dependent Uses

The following uses occupy contested territory where the permissibility depends on institutional policy, disciplinary norms, the degree of AI reliance, and disclosure practices:

1) Literature synthesis: Having AI generate a summary of a body of literature raises questions about whether the researcher has directly engaged with the sources. The permissibility depends on whether the researcher independently reads and critically evaluates the sources the AI has identified.
2) Argument structuring: Using AI to suggest how an argument might be organised may be acceptable as a starting framework that the researcher subsequently develops, but not if the AI's structure is adopted wholesale without critical engagement.
3) Qualitative data coding: AI-assisted initial coding of qualitative data (interviews, focus groups, documents) may be acceptable if the researcher critically reviews and revises the codes, but not if AI coding is treated as definitive without human interpretation.
4) Quantitative analysis: AI tools that suggest statistical approaches or interpret outputs may be used appropriately

if the researcher understands the methods and critically evaluates the suggestions, but problematically if the researcher does not understand what they are doing.

5) Writing first drafts: This is perhaps the most contested area. Some institutions permit AI-assisted drafting in limited contexts; most require that all submitted text be the researcher's own.

4.3. Clearly Prohibited Uses

The following uses are incompatible with academic integrity in virtually all institutional contexts:

1) Submitting AI-generated text as one's own work without disclosure: This is the most direct form of academic dishonesty involving AI [4].
2) Using AI to fabricate data, results, or citations: This constitutes research fraud, regardless of the mechanism of production.
3) Using AI to produce an entire thesis chapter or dissertation: This represents a fundamental failure of the postgraduate mission.
4) Using AI to complete assessments that are explicitly prohibited from AI assistance: Violating explicit prohibitions is academic misconduct regardless of the tool used.
5) Using AI to identify and exploit weaknesses in detection systems: Deliberately circumventing integrity measures is itself an integrity violation.

Table 2. AI use cases mapped to academic integrity status.

Use Case	Integrity Status
Spell/grammar checking	Permissible (standard disclosure)
AI literature discovery tools	Permissible (disclose in methodology)
AI translation (with human review)	Generally permissible (disclose)
AI brainstorming/ideation	Permissible (researcher develops ideas)
AI-assisted literature synthesis	Contextual, requires independent reading
AI-suggested argument structure	Contextual, requires critical development
AI coding assistance in qualitative work	Contextual, requires human interpretation
AI-drafted text submitted as own	Prohibited without full disclosure
AI-fabricated data or citations	Prohibited (research fraud)
Full AI-written thesis sections	Prohibited

5. Plagiarism, Authorship, and the Question of Originality

5.1. The Evolving Definition of Plagiarism

Plagiarism has traditionally been defined as the presentation of another person's ideas, words, or work as one's own, without appropriate attribution. This definition is built on the presupposition of human authorship: there is always a person whose intellectual property is being misappropriated. AI-generated text challenges this presupposition at a fundamental level [8].

LLMs do not hold copyright in their outputs under current legal frameworks in most jurisdictions. In the United States (US), the Copyright Office has repeatedly affirmed that copyright requires human authorship and that AI-generated material is not, in itself, copyrightable. Similar positions have been adopted by the United Kingdom (UK) Intellectual Property Office and the European Union (EU) AI Act framework [6]. Technically, then, using AI-generated text is not plagiarism in the intellectual property sense, because there is no human author whose rights are violated.

However, this legal analysis misses the point of academic integrity. When a postgraduate researcher presents AI-generated text as their own work, the dishonesty is not primarily about intellectual property. It is about misrepresenting the nature of one's intellectual contribution to one's institution, one's discipline, and one's examining committee. This form of dishonesty, which might be termed academic fraud, is distinct from plagiarism but is no less serious.

Several institutions have responded by explicitly extending their plagiarism definitions to encompass AI-generated content. The University of Oxford's revised academic integrity policy (2024) defines plagiarism to include 'the submission of AI-generated text as one's own scholarly work, regardless of whether such text was previously published.' This definitional

extension captures the relevant dishonesty, though it conflates conceptually distinct phenomena, intellectual property violation and academic fraud, in potentially confusing ways.

5.2. Authorship in Multi-Agent Research

The question of authorship becomes particularly complex in postgraduate research that involves significant AI contribution. Authorship in academic publishing has traditionally been understood to carry four responsibilities, as articulated in the International Committee of Medical Journal Editors (ICMJE) guidelines [11]: substantial contribution to conception or design; drafting or critically revising the work; final approval of the version to be published; and accountability for all aspects of the work. AI systems cannot meet these criteria. They cannot be held accountable, they cannot approve final versions, and they have no ongoing responsibility for the work.

Major academic publishers have responded to this challenge by prohibiting the listing of AI tools as co-authors. Journals including *Nature*, *Science*, *The Lancet*, and *PLOS ONE* have all issued policies to this effect [15, 21]. Instead, they require disclosure of significant AI use in the methods section or an author contribution statement. This approach preserves human accountability while accommodating transparency about AI's role.

For postgraduate researchers, the authorship question has an additional dimension: the relationship between AI contribution and the originality requirements of the thesis or dissertation. Most institutional regulations require that a doctoral thesis represent the candidate's own original work. When significant portions of the writing, analysis, or argumentation have been generated by AI, the claim to originality is weakened, not because the text lacks novelty, but because the intellectual labour of generation was not the candidate's.

5.3. Originality Reconsidered

The concept of originality in doctoral research has always been more nuanced than popular discourse suggests. Doctoral candidates are not expected to invent disciplines from scratch; they are expected to make an original contribution within an existing field, which may be theoretical, empirical, methodological, or applied. The contribution must be sufficient to merit examination by the candidate's peers and to advance knowledge in some meaningful respect.

AI tools complicate originality in at least three ways. First, because LLMs generate text by predicting the most statistically likely continuation of a given prompt, a process that reflects patterns in their training data [2], their outputs may reproduce existing ideas and framings without attribution, even when they do not reproduce verbatim text. This is a form of intellectual déjà vu rather than novelty: plausible-sounding argument that reflects the consensus of existing literature rather than genuine insight.

Second, if many researchers in a field are prompting AI tools with similar queries, their outputs will be structurally similar, potentially producing a homogenisation of argument and framing that undermines the intellectual diversity that makes academic discourse valuable.

Third, and most fundamentally, originality in the doctoral context is not just about the text produced. It is about the intellectual process that produced it. A researcher who has genuinely grappled with a problem, failed, revised, and arrived at an insight through sustained intellectual effort has produced something original in the most important sense, regardless of whether the final words on the page are statistically unusual. AI shortcuts that eliminate this process undermine originality at its root, even when the final product appears superficially novel [20].

6. Institutional Policy Responses: A Global Survey

6.1. The Policy Landscape in 2024-2025

The institutional policy response to AI in academic research has been characterised by urgency, inconsistency, and rapid evolution. In the immediate aftermath of ChatGPT's release, many universities issued emergency guidance that reflected primarily prohibitive stances, often blanket bans on AI use in assessed work. Over the following eighteen to twenty-four months, a more nuanced policy landscape emerged, as institutions recognised the futility and educational inadvisability of blanket prohibition [1].

A survey of policy approaches across leading research universities reveals several broad orientations. The first, which might be termed the prohibitionist model, restricts AI use in assessed research work to grammar and spell-checking only, requires explicit declaration of all AI tool use, and treats any undisclosed AI assistance as academic misconduct. This model has been adopted, in varying degrees of strictness, by institutions including the University of Cambridge, some faculties at the University of Tokyo, and several German research universities.

The second orientation, the regulated transparency model, permits a wider range of AI uses but requires detailed disclosure in methodology sections, appendices, or dedicated AI use statements. Institutions including the University of Melbourne, Stanford University, and the University of Toronto have developed sophisticated disclosure frameworks that ask researchers to specify which tools were used, for what purposes, at what stages, and how the outputs were critically evaluated and modified.

The third orientation, the disciplinary autonomy model, delegates detailed AI policy to individual faculties, departments, or supervisors, on the grounds that the appropriate role of AI varies so significantly across disciplines that centralised rules are inherently inadequate. Massachusetts Institute of Technology (MIT), Eidgenössische Technische Hochschule (ETH)

Zurich, and University College London (UCL) have adopted versions of this approach [14].

6.2. Policy Content: Key Elements

Across institutions, the most effective AI integrity policies for postgraduate research share several structural features. They begin with a clear statement of the underlying values that the policy serves, articulating why academic integrity matters and what postgraduate research is for, rather than proceeding immediately to rules. This values-first approach helps researchers understand the spirit of the policy and apply it to novel situations not explicitly covered by the rules.

They provide specific, operational guidance rather than vague exhortations to 'use AI responsibly.' Researchers benefit from concrete examples of permitted and prohibited uses, explicit disclosure requirements, and clear guidance on what disciplinary proceedings would follow from violations.

They are regularly updated. Given the pace of AI development, policies published in 2023 are already substantially outdated. Leading institutions have committed to annual or semi-annual policy reviews, with standing committees that include graduate student representatives alongside faculty and administrators.

They address training and support needs. Policies that simply announce rules without providing the educational infrastructure to implement them place an unfair burden on researchers, particularly those from educational backgrounds where AI norms were never taught.

Table 3. Illustrative institutional AI policy orientations for postgraduate research.

University / System	Policy Orientation	Distinctive Feature
University of Cambridge	Regulated transparency	Course-specific AI guidance; mandatory disclosure forms
Stanford University	Disciplinary autonomy	Department-level policies; honour code adapted for AI
University of Melbourne	Regulated transparency	AI use declaration for all thesis submissions
MIT	Disciplinary autonomy + openness	Faculty encouraged to integrate AI literacy into training
University of Toronto	Regulated transparency	Graduate AI integrity handbook; supervisor guidelines
ETH Zurich	Disciplinary autonomy	Research group-level policies; strong emphasis on method transparency
University of Cape Town	Developing framework	Equity-focused; acknowledges infrastructure disparities
University of Tokyo	Modified prohibitionism	Discipline-specific; strong in STEM, stricter in humanities

6.3. Cross-Institutional Coordination

One emerging response to policy fragmentation is the development of cross-institutional frameworks. The Russell Group (UK), the Group of Eight (Australia), and the Association of American Universities have all issued collective statements on AI and academic integrity that provide frameworks within which member institutions develop specific policies. United Nations Educational, Scientific and Cultural Organization (UNESCO)'s 2023 Guidance for Generative AI in Education and Research [16, 22] represents an important supra-national effort to establish normative consensus, though its non-binding character limits its practical impact.

Disciplinary associations have also begun to develop field-specific guidance. The American Psychological Association (APA), the Modern Language Association (MLA), and the American Historical Association (AHA) have all issued AI use guidelines for researchers and authors in their respective fields. These disciplinary standards are particularly influential in postgraduate research because they shape the norms of the scholarly communities into which doctoral students are being initiated.

7. AI Detection Technologies and Their Limitations

7.1. The Rise of AI Detection

The emergence of AI-generated text as an academic integrity concern has driven rapid growth in AI detection technologies. Tools such as Turnitin's AI Writing Detection, GPTZero, Originality. AI, and Copyleaks AI Detector claim to distinguish AI-generated text from human writing by analysing statistical patterns, particularly the 'perplexity' and 'burstiness' of text, which refer to the predictability of word sequences and the variability of sentence length, respectively. AI-generated text tends to be more uniform and predictable in these metrics than human writing [23].

Several major institutions, including Turnitin's 10,000+ client universities, have integrated AI detection into their plagiarism-checking workflows, producing AI likelihood scores alongside traditional plagiarism percentages in submitted work. This integration has had significant practical consequences: students and researchers whose work returns high AI likelihood scores may face integrity investigations, even in the absence of other evidence of misconduct.

7.2. Accuracy and Error Rates

The reliability of AI detection tools is a matter of serious scientific and legal concern. Published validation studies have found substantially higher false-positive rates, that is, misidentifying human-authored text as AI-generated, than tool developers typically acknowledge in their marketing materials. A widely cited 2023 study by Weber-Wulff and colleagues [24] tested seven AI detection tools against a ground-truth corpus and found false-positive rates ranging from 1.7 percent to 23.4 percent for human-authored text.

False positive rates are not uniformly distributed. Research consistently finds that AI detectors are significantly more likely to misidentify the writing of non-native English speakers as AI-generated, because non-native writing, which tends to use simpler vocabulary, shorter sentences, and more conventional grammatical structures, shares statistical features with AI output. A 2024 study published in the International Journal of Educational Technology [13] found that detector accuracy for native English writers was approximately 85 percent, compared to just 61 percent for non-native English writers.

This differential accuracy has profound equity implications. Applying AI detection tools without accounting for their discriminatory impact on non-native speakers risks unjustly stigmatising and penalising international postgraduate students, who are already navigating multiple disadvantages, on the basis of unreliable evidence.

7.3. The Arms Race Problem

AI detection also faces a fundamental adversarial dynamics problem. As detection algorithms improve, so do techniques for evading them, including paraphrasing tools designed to introduce human-like variability into AI text, hybrid writing strategies that intersperse human and AI content, and instruction techniques that prompt AI models to produce more 'human-sounding' output. This creates a technological arms race in which detection tools are perpetually reactive rather than definitive.

More fundamentally, AI detection cannot address the integrity problem at its core. Even if a detection tool could identify AI-generated text with perfect accuracy, it would not tell us anything about the researcher's actual intellectual engagement. A researcher who reads widely, thinks deeply, and then uses AI to help polish their prose may have contributed more genuine intellectual work than one who writes every word themselves but does so superficially. Relying on AI detection as the primary integrity assurance mechanism mistakes the symptom for the disease.

7.4. Procedural Fairness Concerns

The use of AI detection in academic misconduct proceedings raises serious procedural fairness concerns. Several UK and US universities have faced legal challenges from students who received academic penalties based primarily on AI detection scores, without corroborating evidence and without adequate opportunity to refute the algorithmic finding. The fundamental problem is that AI detection scores are probabilistic assessments, not determinations of fact. In any context where the consequence of a positive finding is significant, such as suspension, degree revocation, or reputational damage, the procedural standards must be commensurate with those stakes.

Best practice guidance increasingly holds that AI detection scores should be treated as intelligence for further investigation rather than as evidence of misconduct, and that misconduct findings should require corroborating evidence beyond a detection score alone. This represents a more legally defensible and educationally appropriate use of these tools.

8. Cross-Cultural Dimensions and Equity Concerns

8.1. The International Postgraduate Community

The global postgraduate research community is profoundly international. At universities in the United Kingdom, Australia, Canada, and the United States, international students typically account for between thirty and sixty percent of doctoral enrolments. These students bring diverse educational backgrounds, prior experiences with academic integrity expectations, and linguistic proficiencies, and they navigate postgraduate programmes under conditions of significant additional pressure.

Academic integrity norms are not culturally universal. Research in the sociology of education has documented systematic differences in how educational traditions across Asia, Africa, the Middle East, and Eastern Europe conceptualise plagiarism, citation practice, and the appropriate use of authoritative sources [3]. In some traditions, the reproduction of authoritative text is not merely permitted but valued as a mark of respect and learning. In others, the boundary between collaboration and collusion is drawn differently than in Anglo-American academic culture. These differences do not reflect lower ethical standards; they reflect genuinely different cultural epistemologies.

Postgraduate students from these backgrounds who have

not received explicit instruction in Anglo-American academic integrity norms may use AI tools in ways that seem natural within their prior educational experience but that constitute serious violations under their new institution's policies. Holding such students to the same standards as those who have been socialized into those norms over many years, without providing adequate education and support, is neither fair nor effective as an integrity strategy.

8.2. Language Equity and AI

The language dimension of AI use in postgraduate research is among the most sensitive equity issues in contemporary higher education. The global academic enterprise operates overwhelmingly in English. Researchers whose native language is not English must produce thesis chapters, journal articles, and conference presentations in a language in which they may have sophisticated ideas but imperfect expression. The cognitive burden of this translation, not merely linguistic but conceptual, as academic registers vary significantly between languages, is substantial and poorly appreciated by many supervisors and examiners.

AI language tools offer real benefits for non-native English-speaking researchers. An AI that can help a researcher express a complex idea in idiomatic English, one that the researcher has developed and understands fully, may be enabling genuine academic contribution that would otherwise be obscured by linguistic limitation. In this respect, AI writing assistance for non-native speakers may function as a form of reasonable accommodation, analogous to the extended time provided to students with dyslexia [14].

However, there is a critical difference between using AI to express one's own ideas more clearly and using AI to generate ideas one does not have. The former is a legitimate accessibility accommodation; the latter is the same integrity violation regardless of the researcher's linguistic background. The challenge for institutions and supervisors is to develop policies sensitive enough to accommodate the former while still prohibiting the latter, a distinction that requires nuanced judgement and trust-based supervisory relationships.

8.3. Infrastructure Disparities

A further equity dimension concerns differential access to AI tools across the global research community. While major LLMs are nominally available globally, access is constrained by internet connectivity (often unreliable in lower-income countries), language coverage (tools perform significantly better in English than in most other languages), cost (premium versions of tools like ChatGPT Plus, Copilot Pro, or Claude Pro are unaffordable for researchers in lower-income contexts), and regulatory restrictions (several major AI tools are geographically restricted).

This differential access risks creating a new dimension of the existing North-South research inequality: researchers at well-resourced institutions in high-income countries gain substantial productivity advantages from AI tools that are inaccessible or less effective for peers in lower-resourced settings. Academic integrity frameworks that assume equal AI access and apply equal restrictions may inadvertently deepen these inequalities [26].

9. The Role of Supervisors and Graduate Schools

9.1. Supervisory Relationships in the AI Era

The supervisor-student relationship is the most consequential single factor in postgraduate researchers' AI use. Supervisors shape their students' understanding of what is permissible, model appropriate scholarly behaviour, and provide the primary context within which integrity norms are communicated and reinforced. Yet surveys consistently show that supervisors are themselves uncertain, inconsistent, and often inadequately trained in AI-related integrity issues.

A 2024 survey of doctoral supervisors at ten UK universities found that 67 percent reported uncertainty about what AI use they should permit or encourage in their students' work, 54 percent had not discussed AI use explicitly with their supervisees, and only 23 percent felt well-equipped to advise students on AI integrity questions [9]. This supervisory knowledge gap creates precisely the conditions, absence of clear guidance, inconsistent enforcement, and implicit rather than explicit norms, in which integrity violations are most likely to occur.

The role of the supervisor in the AI era requires expansion beyond traditional functions. In addition to providing expert guidance on research design and intellectual development, supervisors must now discuss AI use explicitly and early in the supervisory relationship; model appropriate disclosure practices; help students develop critical AI literacy, the ability to evaluate, interrogate, and use AI tools responsibly; and be sufficiently conversant with AI capabilities and limitations to recognise when a student's work may have been unduly AI-assisted.

9.2. Structural Graduate School Responsibilities

Individual supervisors cannot bear the full responsibility for AI integrity education. Graduate schools have structural responsibilities that must be addressed at the institutional level. These include developing and disseminating clear, discipline-sensitive AI use policies; providing mandatory orientation on AI integrity for all new postgraduate students; offering continuing professional development for supervisors on AI tools and their integrity implications; establishing clear procedures for AI-related misconduct investigations; and creating safe channels through which students can seek guidance on AI use

questions without fear of self-incrimination.

Several graduate schools have developed exemplary resources in this area. The University of Melbourne's 'Responsible Use of Generative AI in Research' guide, MIT's 'AI and Research Integrity' faculty resource, and the University of Toronto's 'Graduate Student AI Integrity Handbook' represent different but effective models. Common to all effective resources is an educational rather than punitive tone: they are designed to help researchers navigate genuinely difficult questions, not primarily to warn them against misbehaviour.

9.3. Rethinking Assessment and Supervision Practices

The AI era creates an opportunity, and arguably a necessity, for graduate schools to rethink their assessment and supervision practices in ways that are more robust to AI assistance and more educationally valuable. Traditional thesis and dissertation formats, which require the submission of a final text product without ongoing evaluation of the research process, are particularly vulnerable to AI substitution. Assessment approaches that evaluate process alongside product, including staged submissions, supervisory dialogues, viva examinations designed to probe understanding in depth, and research journals, are simultaneously more resistant to AI fraud and more educationally meaningful.

The oral examination, or viva voce, has renewed importance in this context. A researcher who has genuinely engaged with their material over several years will be able to discuss, defend, and extend their ideas in ways that a researcher who has outsourced substantial intellectual work to AI cannot replicate. Strengthening viva examination practices, increasing their rigour, standardising their format, and ensuring examiners are trained to identify superficial understanding, may be among the most effective integrity interventions available to graduate schools.

10. A Proposed Ethical Framework for AI Use in Postgraduate Research

10.1. Principles of the Framework

The preceding analysis suggests that what is needed is not a simple set of rules but a principled ethical framework that provides guidance for the full range of AI use decisions that postgraduate researchers face. The framework proposed here is organised around five principles: transparency, epistemic responsibility, proportionality, disciplinary sensitivity, and developmental purpose (adapted from [7]).

Transparency requires that all significant AI contributions to research be disclosed in a clear, specific, and standardised manner. 'Significant' is operationalised as any AI use that goes beyond conventional grammar and spell-checking, including literature discovery, text drafting, argument structuring, data analysis, and visualisation. Disclosure should specify the tool used, the research task for which it was used, and the nature of the researcher's critical review and modification of AI outputs.

Epistemic responsibility holds that researchers must take genuine ownership of all claims made in their work. This means they must understand every claim they make, be able to defend it under examination, have verified its accuracy through independent means, and be willing to be held accountable for it. AI outputs that the researcher has not understood, verified, and claimed as their own intellectual responsibility are not appropriate for submission, regardless of their apparent quality.

Proportionality requires that the degree of AI assistance be proportionate to the nature and purpose of the research task. Tasks that are primarily intellectual contributions, such as the development of a theoretical argument, the interpretation of empirical findings, or the evaluation of evidence, should primarily reflect the researcher's own intellectual effort. Tasks that are primarily technical or mechanical, such as reformatting data, checking grammar, or generating initial literature lists, may appropriately involve more substantial AI assistance.

Disciplinary sensitivity acknowledges that appropriate AI use varies legitimately across research fields. In computational research, AI coding tools are standard and expected; in close-reading literary analysis, they are peripheral and potentially integrity-compromising. Frameworks must be developed and applied in ways that are sensitive to these disciplinary norms, which are themselves in evolution.

Developmental purpose reflects the defining feature of postgraduate education: the development of the researcher's own capacities. Any use of AI that systematically replaces the intellectual processes through which these capacities are developed, including critical reading, analytical synthesis, sustained argumentation, and methodological design, is inimical to the purpose of postgraduate study, regardless of its effect on the quality of the final text product.

10.2. Tiered Use Categories

Based on these principles, the framework organises AI use into three tiers, each with associated disclosure and oversight requirements:

Tier 1 - Standard Use (Declare, No Special Permission Required): Grammar, spell, and style checking; standard reference management; AI-powered literature search tools used for discovery (not synthesis); straightforward translation assistance; code debugging. These uses should be briefly noted in the Methods section or a standardised AI use statement.

Tier 2 - Enhanced Use (Declare and Describe; Supervisory Discussion Recommended): AI-assisted literature synthesis (with independent source verification); AI-supported argument structuring (with critical development by researcher); AI-assisted qualitative coding (with researcher review and re-

vision); AI statistical method suggestions (with researcher independent understanding); AI-enhanced writing of non-core sections. These uses require more detailed disclosure and are recommended subjects for explicit supervisory discussion.

Tier 3 - Restricted Use (Institutional/Supervisory Pre-Approval Required; Full Disclosure Mandatory): Any AI generation of core argumentative or analytical text; AI-assisted research design for complex empirical studies; AI-generated figures or visualisations presented as research outputs. These uses require prior approval and the most detailed disclosure.

The Five Principles of the AI Integrity Framework

1) TRANSPARENCY - Disclose all significant AI use, specifically and standardly.
2) EPISTEMIC RESPONSIBILITY - Own and be able to defend every claim.
3) PROPORTIONALITY - Calibrate AI use to the nature of the task.
4) DISCIPLINARY SENSITIVITY - Apply field-appropriate standards.
5) DEVELOPMENTAL PURPOSE - Protect the learning process from AI substitution.

10.3. Disclosure Standards

A standardised disclosure format reduces the ambiguity and inconsistency that currently characterises AI disclosure practice. The following disclosure template is proposed for adaptation by institutions:

'AI tools were used in this research as follows: [Tool name(s)] was/were used for [specific purpose(s)] during the [stage(s)] of the research. All AI-generated outputs were critically reviewed, verified against primary sources, and substantially modified by the author. The intellectual arguments, interpretations, and conclusions in this work are the author's own. [Specify any instances where AI assistance was more substantial and how those contributions were evaluated.] '

This format provides the specificity needed for meaningful transparency while remaining concise enough for practical implementation. It places the burden of disclosure on the researcher, where it belongs, while affirming rather than undermining their intellectual ownership of the work.

10.4. Implementation Pathways

For this framework to be effective, it must be embedded in three levels of institutional practice. At the programme level, AI integrity education should be a core component of all postgraduate induction programmes, with follow-up sessions as researchers move from coursework to thesis phases. At the supervisory level, AI use should be a standing agenda item in supervisory meetings, with explicit agreements about permitted uses and disclosure requirements recorded in supervision contracts or equivalent documents. At the institutional level, policies should be reviewed annually, assessment practices should be adapted to strengthen process evaluation, and viva examination standards should be strengthened to ensure depth of understanding is rigorously tested.

11. Conclusion

The relationship between artificial intelligence and academic integrity in postgraduate research is one of the defining challenges in contemporary higher education. It is a challenge that cannot be met by prohibition alone, by technology alone, or by any single institution acting independently. It requires a sustained, collaborative, and intellectually serious engagement with questions that are, at their core, questions about the nature and purpose of advanced scholarship.

The postgraduate research enterprise rests on a set of commitments that AI, however powerful its tools, cannot fulfil: the commitment to genuine intellectual curiosity, to the patient development of expertise, to honest engagement with evidence that contradicts one's hypotheses, and to the cultivation of a scholarly identity that is grounded in what one has genuinely understood and contributed. These commitments are not merely procedural requirements imposed by institutions; they are the conditions under which knowledge can be trusted and built upon.

AI tools, used with integrity, can serve these commitments. They can help researchers navigate vast literatures more efficiently, express complex ideas more clearly, and overcome the artificial barriers that linguistic inequality imposes on global scholarship. Used without integrity, as a substitute for intellectual engagement rather than an instrument of it, they undermine the very foundations of the enterprise.

The framework proposed in this article, grounded in transparency, epistemic responsibility, proportionality, disciplinary sensitivity, and developmental purpose, provides a starting point for institutions, graduate schools, supervisors, and researchers navigating this terrain. But frameworks are only as effective as the cultures that implement them. Building the culture of principled AI use in postgraduate research requires sustained investment in education, dialogue, and trust, between researchers and supervisors, between graduate schools and faculties, and between the academic community and the broader public that ultimately depends on the integrity of the knowledge that universities produce.

The challenge of AI and academic integrity is not, ultimately, a technological challenge. It is a challenge of purpose: to maintain clarity about what postgraduate research is for, and to build the institutional and interpersonal conditions in which that purpose can be authentically pursued in an age of extraordinary machine intelligence.

"We do not need to fear AI tools. We need to fear the erosion of the intellectual commitments that make academic research worth doing. Clarity about those commitments is our best protection."

Abbreviation

AI	Artificial Intelligence
APA	American Psychological Association
EU	European Union
GPT	Generative Pre-trained Transformer
ICAI	International Center for Academic Integrity
ICMJE	International Committee of Medical Journal Editors
LLM	Large Language Model
MLA	Modern Language Association
OECD	Organisation for Economic Co-operation and Development
STEM	Science, Technology, Engineering, and Mathematics
UK	United Kingdom
UNESCO	United Nations Educational, Scientific and Cultural Organization
US	United States

Author Contributions

Mohammed Zeinu Hassen: Conceptualization, Data curation, Methodology, Writing – original draft, Writing – review & editing

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Anderson, L., & Rainie, L. (2023). Emerging frameworks for AI use in higher education: Institutional responses and student perspectives. Journal of Higher Education Policy, 44(3), 211-238.

[2] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, 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, 610-623. https://doi.org/10.1145/3442188.3445922

[3] Bretag, T. (Ed.). (2016). Handbook of academic integrity. Springer. https://doi.org/10.1007/978-981-287-098-8

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

[5] Eaton, S. E. (2023). Academic integrity in the age of artificial intelligence. In T. Bretag (Ed.), Research and practice in higher education (pp. 44-67). Springer.

[6] European Commission. (2024). Artificial intelligence act: A European approach to excellence and trust. Publications Office of the European Union.

[7] Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

[8] Fyfe, P. (2023). How to cheat on your final paper: Assigning AI for undergraduate writing. AI and Society, 38, 1395-1405.

[9] Gorichanaz, T. (2024). Student perspectives on AI tools and academic integrity: A mixed-methods study in doctoral education. Computers and Education: Artificial Intelligence, 6, 100185.

[10] International Center for Academic Integrity. (2023). The fundamental values of academic integrity (4th ed.). Clemson University.

[11] ICMJE. (2023). Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. International Committee of Medical Journal Editors.

[12] Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F. & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

[13] Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. Patterns, 4(7), 100779.

[14] Lodge, J. M., & Corrin, L. (2024). Generative AI in higher education: Pedagogical principles for equitable implementation. British Journal of Educational Technology, 55(2), 451-468.

[15] Nature. (2023). Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. Nature, 613, 612.

[16] OECD. (2024). Trustworthy AI in education: Perspectives from OECD countries. OECD Publishing.

[17] OpenAI. (2023). GPT-4 technical report. arXiv preprint arXiv: 2303.08774.

[18] Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching and Learning Practice, 20(2), 1-24.

[19] Prather, J., Reeves, B. N., Denny, P., Becker, B. A., Leinonen, J., Luxton-Reilly, A.,. & Weimer, M. (2023). 'It's weird that it knows what I want': Usability and interactions with GitHub Copilot. ACM Transactions on Computer-Human Interaction, 31(1), 1-31.

[20] Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? Journal of Applied Learning and Teaching, 6(1), 342-363.

[21] Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove. Nature, 613, 620-621.

[22] UNESCO. (2023). Guidance for generative AI in education and research. UNESCO Publishing.

[23] Uzun, A. M. (2023). Exploring the landscape of AI detector tools: Accuracy, limitations, and implications for research integrity. Computers in Human Behavior Reports, 12, 100369.

[24] Weber-Wulff, D., Anohina-Naumeca, A., Bjelopavlic, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O.,. & Waddington, L. (2023). Testing of detection tools for AI-generated text. International Journal for Educational Integrity, 19(1), 26.

[25] Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S.,. & Gabriel, I. (2022). Taxonomy of risks posed by language models. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 214-229.

[26] Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education, where are the educators? International Journal of Educational Technology in Higher Education, 16(1), 39.

Biography

Mohammed Zeinu Hassen is a Professor of Philosophy in the Department of Social Sciences at Addis Ababa Science and Technology University, Ethiopia. His current work focuses on the intersection of artificial intelligence, ethics, and higher education. He has published extensively on AI ethics, addressing topics such as algorithmic governance, the impact of AI on students' critical thinking, and equitable frameworks for AI integration. His research aims to develop principled, educationally grounded approaches for responsible AI use in postgraduate research.