Challenges and Risks of Utilising AI Tools for Postgraduate Supervision in Higher Education: Perspectives from South Africa

Peter Babajide Oloba[1] iD

AFFILIATIONS

[1]Department of Languages, Cultural Studies and Applied Linguistics, Faculty of Humanities, University of Johannesburg, Johannesburg, South Africa.

CORRESPONDENCE

Email: olobapeter4u@gmail.com

Abstract: The integration of artificial intelligence (AI) tools into postgraduate supervision in higher education has accelerated globally, offering opportunities to enhance efficiency in research processes and academic mentoring. However, limited empirical evidence exists regarding the risks and challenges of this integration, particularly within Global South contexts such as South Africa. This study investigates the challenges associated with the use of AI tools in postgraduate supervision from a South African perspective. Anchored in a constructivist paradigm, the study adopts a qualitative research design, employing semi-structured interviews with 20 purposively selected participants—10 postgraduate students and 10 supervisors from faculties that are actively integrating AI into supervisory practices. Data were analysed thematically using qualitative content analysis. The findings identify six key challenges: increasing dependence on AI that may erode students' critical thinking and originality; sufficient digital literacy and institutional support; financial and sustainability constraints; the questionable reliability and accuracy of AI-generated outputs; ethical dilemmas and limited cultural contextualisation; and resistance to technological change among supervisors. While acknowledging the potential of AI to enhance research productivity and the quality of supervision, the study cautions against its uncritical adoption, which may compromise academic integrity, creativity, and equity. It recommends institutional strategies, including subsidised AI access, structured training on ethical and critical AI use, the embedding of digital literacy in postgraduate curricula, and the fostering of collaboration with AI developers to ensure culturally relevant systems. A context-sensitive approach is essential to balance the affordances of AI with the preservation of human intellectual agency and critical scholarly engagement in postgraduate supervision.

Keywords: Artificial intelligence, critical thinking, contextual sensitivity, cultural sensitivity, digital literacy, higher education, postgraduate supervision.

1. Introduction

The integration of artificial intelligence (AI) tools into higher education has accelerated significantly in recent years, with an increasing adoption in teaching, learning, and research supervision. AI platforms—such as ChatGPT, Grammarly, Elicit, Jasper AI, QuillBot, Scite, Research Rabbit, Consensus, and various research summarisation tools—provide postgraduate students and supervisors with opportunities to streamline administrative tasks, enhance writing quality, and access information efficiently (Gasaymeh et al., 2024; Raheem et al., 2023). In the domain of postgraduate supervision specifically, AI is increasingly employed to assist with literature reviews, methodology design, data analysis, writing support, and even the drafting of

research proposals. However, despite the growing reliance on AI tools in postgraduate supervision, there remains an insufficient understanding of the challenges and risks that these technologies introduce within higher education systems, particularly in non-Western contexts such as South Africa.

While an expanding body of global scholarship has investigated the ethical implications, biases, and technical limitations of AI in education (Al-Zahrani & Alasmari, 2024; Akgun & Greenhow, 2022), few studies have focused specifically on the postgraduate supervision environment. Even fewer have examined these issues from a Global South perspective, where socio-cultural, linguistic, infrastructural, and epistemological differences profoundly shape the experiences and impacts of AI use (Ade-Ibijola & Okonkwo, 2023; Roche et al., 2023). In African contexts such as South Africa, postgraduate education often involves navigating indigenous knowledge systems, multilingual realities, and community-based research protocols that AI tools, largely developed in Western, high-income settings, are poorly equipped to address. Consequently, there exists a critical gap in understanding how AI-related challenges manifest differently outside of Euro-American academic frameworks. This gap constitutes the central problem underpinning this study: despite the rapid integration of AI tools into postgraduate supervision, there is a lack of empirical evidence concerning the risks and challenges these technologies pose within South African higher education. The existing AI tools may inadvertently undermine critical thinking, compromise originality, misrepresent culturally grounded knowledge, and exacerbate digital inequalities—issues that remain underexplored in current research (Ade-Ibijola & Okonkwo, 2023; Akgun & Greenhow, 2022). In the absence of a clear, context-specific understanding of these risks, South African universities face significant uncertainty in developing appropriate guidelines, training frameworks, and ethical policies to regulate AI use in supervision.

In light of the rapid evolution of AI technologies and their increasingly unpredictable influence on research practices, it is imperative to critically evaluate their implications for postgraduate supervision in South Africa. This study is, therefore, justified by the pressing necessity to generate evidence-based insights that can inform institutional policy, safeguard academic integrity, and guide the ethical and culturally sensitive integration of AI. As supervisors and students increasingly rely on AI in the absence of robust institutional frameworks, there exists a significant risk that uncritical adoption may undermine human intellectual agency, marginalise indigenous knowledge systems, and exacerbate existing inequities. Addressing this gap is essential to ensure that AI functions as an enabler, rather than a threat, to quality, inclusive, and contextually relevant postgraduate education in South Africa. This study aims to address this underexplored area by focusing on the South African higher education context. In doing so, it seeks to contribute to a more nuanced and situated body of knowledge on AI in education, ensuring that technological innovation supports rather than undermines academic quality, ethical research practices, and local relevance in postgraduate education.

1.1 Statement of the problem

The rapid proliferation of AI tools within higher education has begun to reshape postgraduate research and supervision practices, often at a pace that exceeds the development of coherent institutional, ethical, and pedagogical frameworks. In South Africa, postgraduate students and supervisors are increasingly engaging with AI tools to support literature reviews, academic writing, data analysis, and research planning (Al-Zahrani, 2024; Dwivedi et al., 2023). While these technologies promise enhanced efficiency and productivity, their uncritical adoption raises substantive concerns regarding the integrity and quality of postgraduate education. Emerging evidence indicates that excessive reliance on AI tools may undermine critical thinking, scholarly independence, and the formation of academic identity, which are core outcomes of postgraduate training (Gerlich, 2025; Zhai et al., 2024). These risks are particularly pronounced within supervision contexts, where mentorship, intellectual guidance, and epistemic judgement are central, yet increasingly mediated by opaque and poorly understood technological systems.

The problem is further intensified by the South African higher education landscape, which is shaped by persistent socio-economic inequalities, uneven digital infrastructure, multilingual realities, and the importance of indigenous and community-based knowledge systems. AI tools are predominantly developed within Western epistemological and cultural frameworks, limiting their capacity to engage meaningfully with local research contexts, ethical protocols, and culturally embedded forms of knowledge production (Ade-Ibijola & Okonkwo, 2023; Ofosu-Asare, 2024). As a result, the integration of AI risks reinforcing existing inequities, marginalising local and indigenous knowledge, and compromising ethical research practices, particularly where access, affordability, and digital literacy remain uneven (Akgun & Greenhow, 2022; Al-Zahrani, 2024). Despite the accelerating uptake of AI tools in postgraduate supervision, there remains a notable lack of empirical, context-specific evidence to guide institutional policy, supervisory practice, and ethical governance in South African universities. This absence of grounded insight leaves both students and supervisors navigating AI adoption without clear guidance, increasing the likelihood that AI functions as a destabilising rather than enabling force in postgraduate education. Addressing this problem is therefore imperative to ensure that AI supports, rather than erodes, scholarly rigour, ethical integrity, and contextual relevance in postgraduate supervision.

Hence, this study answers the research question: *What are the challenges and risks associated with utilising AI tools for postgraduate supervision in higher education?*

2. Methodology

This study was framed within the constructivist paradigm, which emphasises the understanding of individuals' lived experiences and the subjective meanings they attribute to those experiences (Turin et al., 2024). Constructivism presupposes that reality is socially constructed, thus knowledge emerges through interaction between the researcher and participants (Clarke et al.,

2023). In the context of this study, a constructivist lens was deemed appropriate for exploring how students and supervisors in South Africa perceive the challenges and risks associated with the utilisation of AI tools in postgraduate supervision. A qualitative research approach was adopted to facilitate an in-depth exploration of participants' perspectives (Lim, 2024). Qualitative research is especially suited for investigating complex phenomena where little is known and where understanding participants' experiences, perceptions, and meanings is paramount (Creswell & Creswell, 2017). Given the nascent nature of AI use in higher education supervision, this approach permitted the capture of rich, detailed descriptions. The study employed a phenomenological research design. Phenomenology concentrates on understanding how individuals experience a specific phenomenon by uncovering commonalities in their lived experiences (Gill, 2020). In this instance, the design enabled the researcher to delve into the lived experiences of postgraduate students and supervisors regarding the challenges and risks encountered when utilising AI tools in postgraduate supervision.

The population for this study comprised postgraduate students and supervisors from ten universities in South Africa. In total, twenty participants engaged in the interview process: ten postgraduate students and ten supervisors. This sample size was deliberately selected, as qualitative phenomenological studies prioritise depth of insight over numerical representativeness; smaller samples facilitate the emergence of rich, detailed accounts of lived experiences (Creswell & Poth, 2016). A sample of twenty was deemed sufficiently large to capture diverse perspectives across different universities while remaining manageable for conducting in-depth, interpretive analysis, which is central to phenomenological inquiry (Vagle, 2018). Furthermore, purposive sampling ensured that all selected participants had direct experience with the integration or navigation of AI tools in postgraduate supervision, rendering them information-rich cases capable of providing meaningful data relevant to the study's aim. Purposive sampling was appropriate, as it allowed the researcher to deliberately select participants possessing direct experience with the phenomenon under investigation, thereby ensuring rich, pertinent insights (Campbell et al., 2020). A descriptive analysis of participants' demographic characteristics was compiled to contextualise their perspectives.

The sample included ten postgraduate students and ten supervisors, representing a range of disciplines across humanities, education, science, law, engineering, commerce, arts, design, and architecture faculties. The student group consisted of honours, master's, and doctoral candidates, while supervisors varied in supervisory experience from early-career academics to senior scholars. Participants were drawn from both historically advantaged and historically disadvantaged universities, thereby ensuring diversity in institutional contexts and enriching the phenomenological insights into AI use in postgraduate supervision. This demographic variation strengthened the study by capturing a wide spectrum of lived experiences and contextual influences. Semi-structured interviews served as the primary method of data collection. This approach provided the flexibility to explore predetermined topics while also allowing

participants the freedom to discuss issues they deemed significant (Ruslin et al., 2022). Although interviews were guided by an interview schedule, they permitted probing and follow-up questions to clarify and expand upon participant responses. The data were analysed using qualitative content analysis, which involves a systematic coding and categorisation process to identify patterns, themes, and categories within the data (Zhang & Wildemuth, 2009). Content analysis enabled the researcher to interpret and organise the complex data into meaningful units that addressed the research questions regarding the challenges and risks associated with the use of AI tools for postgraduate supervision.

3. Presentation of Results and Discussion of Findings

In response to the research question regarding the challenges and risks associated with utilising AI tools for postgraduate supervision in higher education, six key themes emerged: dependence on AI tools; inadequate literacy, support, and integration of supervision; cost and sustainability; quality and accuracy of AI output; ethical concerns, including cultural and contextual sensitivity; and resistance to change.

3.1 Dependence on AI tools

An emerging risk associated with the use of AI tools in postgraduate supervision is the increasing dependence on these technologies, which may compromise students' critical thinking, analytical abilities, and independent academic development. Both supervisors and students shared concerns about how over-reliance on AI is affecting the quality of research training and academic skills. A senior humanities supervisor raised concerns about students' excessive reliance on AI for major sections of their academic work: *"Some students now rely on AI to write entire sections of their research. It is worrying because they are not developing the necessary analytical or writing skills* themselves." Similarly, a master's student from the Education Faculty acknowledged the tendency to overuse AI tools, particularly during tasks like literature reviews: *"AI tools help speed things up, especially when doing literature reviews, but I find myself depending on them too much instead of reading and analysing articles on my* own." An early-career supervisor in the sciences noted the superficial quality of some student submissions, attributing it to the overuse of AI tools: *"I have noticed a trend where students submit work that looks polished but lacks depth. It becomes clear in discussions that the critical engagement is missing, possibly because the AI did most of the* work." A doctoral student in engineering also shared personal struggles with diminished writing confidence after habitual use of AI platforms: *"At first, using AI tools was exciting, but now I struggle to write without them. It is like I have lost confidence in my own ability to start from* scratch." Echoing this concern, a senior supervisor from the Commerce Faculty emphasised the need for proper training on appropriate AI use: *"We need to teach students when and how to use AI appropriately. Right now, some are just outsourcing their thinking to* machines." An honours student in the Arts and Design Faculty expressed doubt about the authenticity of their own academic growth: *"Sometimes I wonder if I am learning or just editing what the AI generates. It is convenient, but I am not sure I am growing as a* researcher."

Supervisors particularly noted that some students are relying heavily on AI tools for composing significant sections of their research, resulting in underdeveloped analytical and academic writing skills. This observation aligns with findings by Gerlich (2025), who cautions that while AI technologies provide efficiency, they can also deskill students by automating cognitive processes that are essential for achieving higher education learning outcomes, such as critical reasoning and academic argumentation. Students' own reflections reveal an internal struggle with dependence, particularly during literature reviews, where AI tools expedite the process but at the expense of deep reading and personal analysis. Zhai et al. (2024) similarly argue that an overreliance on AI for information processing can disincentivise deeper cognitive engagement, thereby undermining one of the core purposes of postgraduate education: developing the ability to critically interrogate texts and construct original arguments. The concern regarding the superficial quality of submissions—polished work lacking critical depth—is particularly alarming. Supervisors observed that although AI-generated text may appear well-written, it often conceals the absence of genuine analytical thought, which becomes evident during oral discussions. This echoes the critique by Delcker et al. (2024), who highlight the veneer of competence that AI can create in education, wherein surface-level outputs are mistakenly recognised as indicative of deep learning.

Moreover, students reported a loss of confidence in their own writing abilities after frequent use of AI, feeling unable to initiate academic writing without machine assistance. This aligns with Gerlich (2025), who warns that habitual use of AI can lead to academic learned helplessness, whereby students doubt their own capabilities and become overly reliant on AI-generated content to commence or structure their work. Supervisors emphasised the necessity for explicit training on the appropriate and ethical use of AI, as well as guidance on when and how to integrate AI tools without outsourcing one's intellectual labour. Without clear guidelines, students risk not only stunting their academic growth but also breaching ethical boundaries in their research processes. Additionally, students' doubts about the authenticity of their learning, questioning whether they are genuinely developing as researchers, point to a deeper existential risk: the erosion of academic identity. If postgraduate education fundamentally aims to develop scholars capable of independent thought and the contribution of original knowledge, then unchecked dependence on AI threatens the very core of that mission (Zhai et al., 2024).

3.2 Inadequate literacy, support, and supervisory integration

Participants indicate that one of the challenges associated with the use of AI tools for postgraduate supervision is the lack of adequate digital literacy, institutional support, and the seamless integration of these technologies into existing supervisory practices. One of the Honours students in the faculty of Humanities shared their difficulty in navigating AI platforms without formal guidance: "*I often hear about ChatGPT and other AI platforms, but honestly, I do not know how to use them in a way that supports my research. There is no proper orientation or guidance.*" A master's student in engineering highlighted the knowledge gap that exists among supervisors,

which affects the effective use of AI tools in the supervision process: "*Most supervisors in my department are not familiar with AI tools, so even if we wanted to use them, there is a knowledge gap that affects how we are supervised.*" A doctoral student in Education recounted the lack of structured training and the need for self-directed learning: "*We were introduced to AI tools without any structured training. I had to teach myself through YouTube videos. It is not sustainable or effective.*" Similarly, a senior supervisor in the Faculty of Science described the overwhelming pressure of incorporating AI tools without proper training or institutional preparation: "*I am expected to incorporate AI tools into my supervision, but I have received no formal training. It becomes overwhelming and time-consuming, especially with an already heavy workload.*" Another early-career engineering supervisor pointed to the absence of technical support structures within the institution: "*There is no institutional support or technical team to help us when we get stuck using these tools. You either figure it out yourself or abandon it.*" Expressing frustration over the lack of clear frameworks, another senior supervisor in Arts and Design stated, "AI *tools are supposed to help streamline supervision, but in reality, they have added extra layers of confusion because we don't have a clear framework on how to use them in our workflow.*" A master's student in law also reflected on how supervisors' lack of confidence with AI tools limits collaboration possibilities: "*My supervisor is not very confident using AI tools, so we just stick to emails and WhatsApp. It is frustrating because I know there are better ways to collaborate.*" A supervisor in Humanities stressed the need for better integration of AI tools into traditional supervisory models: "*We need proper integration of AI tools into our existing supervision models. Right now, it feels like we are trying to fit a square peg into a round hole.*"

The findings indicate that both students and supervisors emphasised how a lack of structured training, technical support, and clear implementation frameworks undermines the potential benefits of AI technologies. The issue of digital literacy was strongly highlighted. Several students reported difficulties in navigating AI platforms such as ChatGPT without formal guidance. One participant noted reliance on self-directed learning through informal sources, such as YouTube videos. This reflects broader concerns raised by Filiz et al. (2025), who assert that the successful adoption of AI in education heavily depends on the digital competencies of both students and educators. In the absence of structured training, students are often left to navigate these technologies independently, resulting in inconsistent usage and superficial engagement with AI tools (Zhou et al., 2024). The knowledge gaps among supervisors further exacerbate this challenge. Students observed that supervisors' unfamiliarity with AI tools adversely affects the quality of supervision they receive. This aligns with Al-Zahrani (2024), who contends that if supervisors are not adequately trained or confident in using AI, they are unlikely to encourage or model effective usage for their students. Rather than serving as a collaborative tool to enhance supervision, AI becomes a source of disconnection.

The data also revealed that the absence of clear frameworks for AI integration generates confusion instead of streamlining the supervisory process. Supervisors reported that AI tools often feel like an additional burden rather than an enabler of efficient supervision, as they are introduced without alignment to existing academic workflows. Southworth et al. (2023) argue

that AI tools must be integrated thoughtfully into existing pedagogical models, rather than being deployed as add-ons or afterthoughts. When integration is poorly planned, it increases cognitive overload for both supervisors and students, thereby reducing the effectiveness of supervision. Furthermore, the lack of confidence among supervisors has downstream effects on collaboration. One student noted that, due to their supervisor's discomfort with AI tools, communication reverted to traditional methods such as emails and WhatsApp, which limited opportunities for AI-supported collaboration. As highlighted by Haleem et al. (2017), digital collaboration tools can enhance feedback loops, peer interaction, and mentorship, but only when users possess the necessary confidence and competence in their use. Ultimately, the call for improved integration of AI tools into existing supervisory models reflects the necessity for systematic change. Simply introducing AI tools without re-evaluating supervision models results in a scenario where the technology appears mismatched and ineffective.

3.3 Cost and sustainability

Cost and sustainability emerged as significant challenges associated with the utilisation of AI tools for postgraduate supervision in higher education. An honours student emphasised the unaffordability of AI tools without institutional support: "*Some of these AI tools come with monthly subscription costs that are simply not feasible for students like us. If the university does not cover them, we can't afford to use them consistently.*" Similarly, an early-career supervisor raised concerns about the sustainability of AI platforms once initial funding or trial periods end: "*We were introduced to a promising AI platform for supervision, but the institution couldn't maintain the subscription after the trial period. It raises concerns about how sustainable these tools really are for long-term use.*" A master's student highlighted infrastructural barriers, particularly for those in less-resourced areas: "*AI tools are great in theory, but without stable internet and access to high-performance devices, many of us in rural areas are left behind.*" A mid-senior supervisor also pointed to the financial burden institutions face in maintaining AI tools beyond mere initial access: "*The cost of maintaining and updating AI platforms can be burdensome for our institution. It is not just about access but about sustaining that access over time.*" A doctoral candidate noted how financial challenges at the student level can lead to exclusion, undermining the goal of inclusive supervision: "*I have seen students drop out of AI-supported initiatives simply because they could not keep up with the cost of data or did not have devices compatible with the tools. It defeats the purpose of inclusive supervision.*"

The data presentation underscores cost and sustainability as significant challenges to the effective utilisation of AI tools in postgraduate supervision. Both students and supervisors expressed concerns regarding subscription fees, infrastructure requirements, and the long-term financial burden associated with the integration of AI in higher education contexts. Financial accessibility emerged as a major barrier. Students reported that many AI tools require monthly subscription fees that are unaffordable without institutional support. This finding aligns with Sabiteka et al. (2025), who indicated that affordability is one of the key barriers to the adoption of educational technology, particularly in low- and middle-income contexts. Without subsidised

access, AI tools risk exacerbating inequalities among students, particularly those from disadvantaged backgrounds. As one student noted, unless universities cover these costs, consistent use of AI tools remains unattainable for many. Furthermore, the sustainability of AI initiatives beyond initial trials was another concern. A supervisor pointed out that while institutions often introduce AI platforms through pilot programmes or funded projects, they struggle to maintain subscriptions after the funding period concludes. This mirrors concerns raised by Patel and Ragolane (2024), who argue that universities often underestimate the hidden costs of AI systems, including subscription renewals, system updates, technical support, and capacity-building initiatives.

In addition to financial concerns, infrastructural challenges, such as limited internet connectivity and lack of access to high-performance devices, were raised, particularly by students from rural areas. This highlights the digital divide that continues to limit the effectiveness of educational innovations. As noted by Haleem et al. (2022), equitable access to both hardware and reliable connectivity is foundational for any technology-driven educational reform. AI tools, which often require substantial computing power and continuous internet access, unintentionally disadvantage students in less-resourced contexts. Supervisors also emphasised the institutional financial burden of maintaining AI platforms beyond the initial rollout phase. While initial access may be funded, ongoing expenses such as licensing, technical support, training, and cybersecurity infrastructure create additional pressures on already stretched university budgets. These financial strains can limit the sustainability of AI integration, leading to a cycle of promising pilot projects that fail to scale. The exclusionary effects of financial barriers were highlighted. One supervisor observed that students unable to afford compatible devices or continuous data costs were effectively excluded from AI-enhanced supervision initiatives. This finding resonates with Laufer et al. (2021), who warn that uncritical adoption of educational technologies without attention to access and equity can reinforce existing inequalities rather than bridge them.

3.4 Quality and accuracy of AI output

Another significant challenge associated with the use of AI tools in postgraduate supervision is the questionable quality and accuracy of AI-generated outputs. A senior academic supervisor pointed out that while AI-generated outputs often appear convincing, they can be dangerously inaccurate upon closer inspection: "*AI tools sometimes give results that sound convincing but are completely inaccurate when you double-check the sources. That can mislead a student if they don't verify the information.*" A master's candidate student shared a personal experience of receiving poor-quality assistance when attempting to summarise literature using AI tools: "*I used an AI tool to help with summarising literature, but it gave me outdated and irrelevant studies. I had to redo everything manually.*" An honours candidate student participant raised a serious concern about fabricated references produced by AI platforms: "*The AI-generated references looked perfect until I tried to find them. Many did not even exist. That is dangerous for any academic work.*" Highlighting a different dimension of the issue, an early-

career academic supervisor noted the inability of AI tools to engage with local or indigenous knowledge systems: "*These tools are not context-aware. When dealing with local South African data or indigenous knowledge systems, the AI simply does not know what to do with it.*" A senior academic supervisor warned about the broader risk of students' critical thinking being compromised by blind trust in AI-generated content: "*If we are not careful, AI will start shaping how students interpret information, even if the data is flawed. We need to teach critical evaluation more than ever.*" Similarly, a doctoral candidate described confusion resulting from conflicting advice between AI-generated suggestions and their supervisor's guidance: "*I relied on an AI tool to help check my methodology section, but the suggestions it gave contradicted my supervisor's advice. It created confusion instead of clarity.*"

The data underscore a critical and escalating concern regarding the quality and accuracy of AI-generated outputs in postgraduate supervision. Participants articulated how convincingly written yet factually inaccurate outputs, fabricated references, outdated materials, and a lack of contextual sensitivity pose significant risks to academic integrity and quality. Supervisors expressed that while AI-generated outputs often appear polished, they frequently contain factual inaccuracies that could mislead unsuspecting students. This observation aligns with findings by Dwivedi et al. (2023), who caution that AI systems, such as GPT models, often produce plausible nonsense—text that sounds credible but is factually incorrect or misleading. In a research-intensive environment like postgraduate study, such errors can compromise the validity of scholarly work if students fail to critically verify AI-generated content. Students' experiences of receiving outdated or irrelevant summaries from AI tools reflect broader limitations in how AI accesses and updates information. As Kumar et al. (2024) note, many AI models are trained on static datasets, meaning their knowledge is frozen at a certain point, which leads to the risk of citing obsolete research, a significant issue in fast-evolving academic fields. Fabricated references emerged as a particularly serious problem. Students recounted instances where AI tools generated convincing-looking yet non-existent citations, a phenomenon known as AI hallucination. This not only undermines academic credibility but could also lead to severe breaches of academic honesty policies if students unknowingly utilise such fabricated sources in formal submissions.

Supervisors raised important concerns regarding AI's lack of contextual and localised knowledge, particularly concerning indigenous knowledge systems and region-specific research. AI's inherent biases towards dominant Western knowledge paradigms and its inability to meaningfully engage with local contexts have been well-documented. This limitation poses a critical challenge for postgraduate research in diverse contexts such as South Africa, where local data, indigenous methodologies, and culturally sensitive approaches are essential. Another alarming risk is the potential erosion of critical thinking. If students accept AI outputs uncritically, they may internalise flawed information and reproduce it in their work, ultimately weakening their academic rigour. Zhai et al. (2024) emphasise that as AI becomes increasingly embedded in educational practice, there is an urgent need to cultivate critical digital literacy—

the ability to question, verify, and reflect on machine-generated outputs rather than passively accept them. Students' confusion due to conflicting advice from AI tools versus their supervisors underscores a deeper problem: AI is not yet sophisticated enough to replace the nuanced, context-sensitive, dialogical nature of human supervision. As Akgun and Greenhow (2022) argue, genuine education involves a level of emotional, ethical, and contextual intelligence that current AI simply cannot replicate.

3.5 Ethical concerns, cultural and contextual sensitivity

Ethical concerns, cultural misalignment, and lack of contextual sensitivity emerged as critical risks when utilising AI tools in postgraduate supervision. One of the senior supervisors expressed concern about the cultural limitations embedded within AI technologies: "*Most of these AI platforms are developed in the West, so they do not always align with our cultural or ethical research standards here in South Africa.*" Similarly, a master's candidate voiced anxiety about the potential ethical risks related to confidentiality and sensitive data management: "*I worry that using AI tools could lead to accidental breaches of confidentiality, especially when dealing with sensitive community-based data.*" An early-career supervisor highlighted the inability of AI systems to accurately interpret cultural nuances in qualitative research: "*AI does not understand the cultural nuances behind some of our qualitative interviews. It flattens everything and removes the human element.*" From the honours candidate's perspective, there was also dissatisfaction with the ethical guidance provided by AI tools, which was often too generic and misaligned with local practices: "*I once asked an AI tool to help with ethical considerations, and it gave generic Western norms that did not apply to our local protocols.*" An experienced supervisor cautioned about the risks of students bypassing critical ethical processes due to overreliance on AI assistance: "*There is a risk that students could rely on AI without understanding the ethical approval process, which could lead to serious violations.*" A doctoral candidate emphasised the broader issue of AI outputs ignoring local realities and linguistic diversity: "*AI outputs often ignore the local context and language, which is problematic when our research is grounded in South African communities and realities.*"

One of the key issues identified by the participants is the cultural limitations of AI technologies. This concern is consistent with research by Ofosu-Asare (2024), which highlighted that AI tools trained on datasets predominantly from Western contexts may be culturally insensitive, reinforcing biases that exclude non-Western perspectives. In the South African context, where research often engages with diverse communities, these limitations are even more pronounced, particularly when AI tools are applied to research involving qualitative methodologies, such as interviews with marginalised or indigenous populations. Furthermore, participants expressed anxiety over ethical risks, particularly related to confidentiality and the handling of sensitive data. This aligns with broader discussions in the literature, where scholars such as Paul (2024) argue that AI systems, particularly those involving cloud-based platforms, pose privacy risks due to potential data breaches or improper handling of sensitive information. In postgraduate research,

where data confidentiality is paramount, these concerns are valid and need to be addressed to ensure the ethical integrity of academic work.

Another dimension of the ethical risks discussed is the inability of AI tools to interpret cultural nuances in qualitative research. This limitation highlights the inherent challenges associated with employing AI in fields such as social science and the humanities, where cultural sensitivity is paramount. As De Freitas et al. (2023) suggest, AI tools are not adept at recognising the complexity of human experiences, particularly those rooted in specific cultural or social contexts. In South Africa, where research frequently addresses issues such as inequality, historical trauma, and indigenous knowledge systems, this can result in misinterpretations or oversimplifications of the data.

Concerns regarding the ethical guidance provided by AI tools, which are often too generic and misaligned with local research practices, further underscore the necessity for contextual sensitivity in AI design. This critique aligns with Hagendorff (2022), who argues that AI tools can perpetuate ethical blind spots because they are frequently not designed to account for the ethical standards of varying global contexts. Given that ethical standards can differ significantly across countries and cultures, reliance solely on AI for guidance may lead to unethical practices, particularly in areas involving human subjects or vulnerable communities. Furthermore, the risk of circumventing ethical approval processes due to students' overreliance on AI tools presents another serious concern. Supervisors have warned that students may overlook the necessary steps to obtain ethical clearance for their research if they rely excessively on AI-generated content. This issue has been raised by Osasona et al. (2024), who note that while AI tools can expedite research processes, they should not replace the rigorous, human-driven ethical review processes required for conducting responsible research. Linguistic diversity and local realities emerge as concerns when AI tools fail to account for the local context. This aligns with Zhai et al. (2024), who stress that AI's lack of understanding of diverse linguistic contexts can lead to misinterpretations or the exclusion of non-dominant languages and knowledge systems.

3.6 Resistance to change

Resistance to change was highlighted as a major challenge in the adoption of AI tools for postgraduate supervision. A senior lecturer reflected on the scepticism among academic staff regarding AI integration: "*Many of my colleagues are still sceptical about AI; they see it as a threat rather than a support tool.*" A master's candidate shared a similar perspective, explaining how some supervisors' feelings of insecurity hinder AI adoption: "*Some supervisors feel that using AI undermines their expertise, so they avoid it completely.*" An early-career academic pointed out the broader institutional resistance to innovation: "T*here is a general resistance at the institutional level; they are slow to adopt anything new, especially if it challenges traditional academic practices.*" Echoing these sentiments, a doctoral candidate expressed frustration about how fear of change hampers progress: "*It is frustrating because even though AI could help us work smarter, the fear of change holds everyone back.*"

Participants articulated that scepticism, fear of undermining traditional expertise, and institutional inertia impede the adoption of AI technologies in academic practices. A supervisor's reflection on the scepticism surrounding AI adoption is consistent with existing literature addressing the barriers to technology integration in educational contexts. As noted by Mirbabaie et al. (2020), academic staff frequently perceive new technologies, such as AI, as a threat to their established practices and professional identities. This fear of disruption is particularly pronounced among those who may feel that AI could challenge their expertise or diminish their role in the research process. The sentiment that AI could "undermine expertise" was echoed by both students and supervisors in the data, underscoring the human element of academic work that many educators consider irreplaceable by technology. In particular, Sobaih et al. (2025) argue that such resistance is exacerbated by a lack of familiarity with AI tools and discomfort regarding their perceived inability to replicate the nuanced judgement and insight that human experts contribute to academic work. The apprehension surrounding AI as a tool that could undermine academic authority also aligns with concerns raised by Singun (2025), who discusses educators' reluctance to embrace digital tools due to perceived threats to their autonomy and academic freedom. In many instances, the introduction of AI tools is viewed as an external imposition on well-established academic processes, thereby generating friction between traditional academic norms and technological innovation. The data presentation indicates that supervisors, particularly those entrenched in traditional pedagogical methods, may entirely avoid using AI tools to safeguard their expertise and maintain control over academic supervision.

Institutional resistance, as highlighted by a supervisor, represents another significant challenge. Higher education institutions often exhibit institutional inertia, characterised by a slow adoption of new technologies that heavily relies on established systems and structures. This resistance is further compounded by the bureaucratic and hierarchical nature of many academic institutions, which can stifle innovation. The level of institutional resistance frequently arises from resource constraints, a lack of training opportunities, and uncertainty regarding the potential benefits of AI. As Pramjeeth and Ramgovind (2024) argue, higher education institutions must transition from a reactive to a proactive stance, adopting policies that actively promote the integration of AI into teaching and research practices to surmount this inertia. The fear of change within academic environments, as expressed by students in the data, reflects a broader reluctance to disrupt the status quo. While AI could present opportunities to work more efficiently, as one student noted, the fear of the unknown and the perceived risks associated with integrating AI into academic workflows can lead to a lack of enthusiasm for its adoption. Kraus et al. (2022) assert that technological change is often met with resistance because it necessitates shifts not only in individual behaviours but also in organisational culture and mindset. In this context, both supervisors and students exhibit hesitance towards embracing AI tools due to the uncertainty

surrounding their effectiveness and the potential costs associated with adapting to new technologies.

4. Conclusions

This study examined the challenges and risks associated with the utilisation of artificial intelligence (AI) tools for postgraduate supervision within South African higher education contexts. The findings reveal that, while AI tools offer opportunities for enhancing efficiency and accessibility, their integration into postgraduate supervision is fraught with significant challenges. Principal among these are issues of cost and sustainability, inadequate digital literacy and institutional support, overdependence on AI leading to compromised critical thinking, concerns regarding the quality and accuracy of AI-generated outputs, and ethical and cultural insensitivity of AI systems. The research demonstrates that without careful management, AI tools may exacerbate existing inequalities, undermine academic development, and dilute the quality and integrity of postgraduate education. Furthermore, the study highlights that the successful integration of AI tools necessitates more than mere technical adoption; it requires pedagogical redesign, critical engagement, and a commitment to equity, ethics, and contextual relevance. As South African higher education institutions progress towards the incorporation of digital technologies in supervision, a nuanced, critical, and supportive approach will be essential to mitigate risks and foster meaningful educational outcomes.

4.1 Recommendations

Universities are urged to assume proactive responsibility for providing subsidised access to AI tools, reliable hardware, and stable internet connectivity for postgraduate students and supervisors, with particular emphasis on individuals from disadvantaged backgrounds. This can be realised through targeted funding schemes, partnerships with technology companies, and internal budget allocations prioritising digital equity. Beyond improving access, the social implications of this recommendation encompass the reduction of digital inequality, the fostering of equitable participation in research, and the prevention of the exacerbation of socio-economic divides within higher education. Practically, sustained institutional investment guarantees that access to AI does not conclude with short-term pilot projects but instead becomes a permanent feature of the research ecosystem. Implementing this recommendation is essential for creating an inclusive research environment where all students can meaningfully leverage AI tools to enhance the quality, efficiency, and innovation of their academic work.

It is recommended that postgraduate schools, in collaboration with digital learning departments and ethics committees, design and deliver formal, ongoing training programmes aimed at developing both the technical competence and critical engagement capacities of supervisors and students regarding the use of AI. These programmes should extend beyond operational training to include modules on ethical standards, verification of AI outputs, responsible referencing, and the maintenance of academic integrity. Clear institutional guidelines and policies should also

regulate the appropriate use of AI tools in postgraduate research. The social implications of this recommendation encompass the fostering of a culture of ethical scholarship, the prevention of academic misconduct, and the promotion of responsible digital citizenship within the academic community. Practically, such training equips researchers with the skills to navigate AI-driven environments confidently, reduces inappropriate reliance on AI, and strengthens the supervisory relationship by ensuring a shared understanding of AI's role and limitations.

Universities should integrate critical digital literacy into postgraduate curricula, led by curriculum developers, supervisors, and academic staff, to enable students to critically evaluate, question, and verify AI-generated information. Institutions should also advocate for the development and adoption of culturally sensitive AI systems that incorporate diverse, local, and indigenous knowledge frameworks, potentially through collaborations with AI developers and researchers in the humanities and social sciences. Establishing regular monitoring and evaluation systems, supported by quality assurance units, will assist institutions in tracking AI's impact on supervision, identifying unintended consequences, and refining guidelines over time. The social implications of this recommendation include the promotion of culturally grounded knowledge production, the safeguarding of linguistic and epistemic diversity, and the assurance that AI technologies reflect the lived realities of South African communities. Practically, this approach ensures that AI serves as a tool for empowerment, enhancing research relevance, improving decision-making, and strengthening academic inclusivity, rather than perpetuating bias or homogenising knowledge.

4.2 Limitations of study

While the phenomenological, constructivist design produced rich, contextually grounded insights into postgraduate students' and supervisors' experiences with AI tools, several limitations should be acknowledged. The purposive sample of twenty participants (ten students and ten supervisors) from selected South African universities limits the study's generalisability beyond the sampled institutions and disciplines. Additionally, the findings reflect participants' self-reported perceptions rather than observed practices. The reliance on semi-structured interviews introduced the possibility of social desirability and recall biases. Furthermore, the sole use of qualitative content analysis without methodological triangulation (e.g., document analysis or observation) may have constrained the breadth of evidence. Researcher positionality within a constructivist lens could also have influenced data interpretation. Given the rapid evolution of AI technologies, the findings are time-bound and may not fully capture subsequent developments or platform changes. These limitations are discussed in relation to the study design and sample details in the manuscript.

The study suggests further enquiries on:

- How higher education institutions could design contextually responsive policies and support frameworks for the ethical, equitable, and sustainable use of AI tools in postgraduate supervision.
- Postgraduate students' and supervisors' experiences of developing critical AI literacy skills and how these skills impact research quality and academic identity formation in higher education.

5. Declarations

Funding: This research did not receive any external funding.
Conflicts of Interest: The author declares no conflict of interest.
Use of Artificial Intelligence: The current work was created without the assistance of artificial intelligence technologies, as confirmed by the author.

References

Ade-Ibijola, A., & Okonkwo, C. (2023). Artificial intelligence in Africa: Emerging challenges. In *Responsible AI in Africa: Challenges and opportunities* (pp. 101–117). Springer International Publishing.

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K–12 settings. *AI and Ethics, 2*(3), 431–440. https://doi.org/10.1007/s43681-021-00096-7

Al-Zahrani, A. M. (2024). Unveiling the shadows: Beyond the hype of AI in education. *Heliyon, 10*(9), e30696. https://doi.org/10.1016/j.heliyon.2024.e30696

Al-Zahrani, A. M., & Alasmari, T. M. (2024). Exploring the impact of artificial intelligence on higher education: The dynamics of ethical, social, and educational implications. *Humanities and Social Sciences Communications, 11*(1), 1–12. https://doi.org/10.1057/s41599-024-03432-4

Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., ... & Walker, K. (2020). Purposive sampling: Complex or simple? Research case examples. *Journal of Research in Nursing, 25*(8), 652–661. https://doi.org/10.1177/1744987120927206

Clarke, A., Healy, K., Lynch, D., & Featherstone, G. (2023). The use of a constructivist grounded theory method is a good fit for social work research. *International Journal of Qualitative Methods, 22*, 1–10. https://doi.org/10.1177/16094069231186257

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches.* Sage Publications.

De Freitas, J., Agarwal, S., Schmitt, B., & Haslam, N. (2023). Psychological factors underlying attitudes toward AI tools. *Nature Human Behaviour, 7*(11), 1845–1854. https://doi.org/10.1038/s41562-023-01734-2

Delcker, J., Heil, J., Ifenthaler, D., Seufert, S., & Spirgi, L. (2024). First-year students' AI competence as a predictor for intended and de facto use of AI tools for supporting learning

processes in higher education. *International Journal of Educational Technology in Higher Education, 21*(1), 1–13. https://doi.org/10.1186/s41239-024-00452-7

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges, and implications of generative conversational AI for research, practice, and policy. *International Journal of Information Management, 71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Filiz, O., Kaya, M. H., & Adiguzel, T. (2025). Teachers and AI: Understanding the factors influencing AI integration in K–12 education. *Education and Information Technologies.* Advance online publication. https://doi.org/10.1007/s10639-025-13463-2

Gasaymeh, A. M. M., Beirat, M. A., & Abu Qbeita, A. A. A. (2024). University students' insights into generative artificial intelligence (AI) writing tools. *Education Sciences, 14*(10), 1–18. https://doi.org/10.3390/educsci14101062

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), 1–28. https://doi.org/10.3390/soc15010006

Gill, M. J. (2020). Phenomenology as a qualitative methodology. In *Qualitative analysis: Eight approaches* (pp. 73–94).

Hagendorff, T. (2022). Blind spots in AI ethics. *AI and Ethics, 2*(4), 851–867. https://doi.org/10.1007/s43681-021-00122-8

Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers, 3*, 275–285. https://doi.org/10.1016/j.susoc.2022.05.004

Kraus, S., Durst, S., Ferreira, J. J., Veiga, P., Kailer, N., & Weinmann, A. (2022). Digital transformation in business and management research: An overview of the current status quo. *International Journal of Information Management, 63*, 102466. https://doi.org/10.1016/j.ijinfomgt.2021.102466

Kumar, V., Ashraf, A. R., & Nadeem, W. (2024). AI-powered marketing: What, where, and how? *International Journal of Information Management, 77*, 102783. https://doi.org/10.1016/j.ijinfomgt.2024.102783

Laufer, M., Leiser, A., Deacon, B., Perrin de Brichambaut, P., Fecher, B., Kobsda, C., & Hesse, F. (2021). Digital higher education: A divider or bridge builder? Leadership perspectives on edtech in a COVID-19 reality. *International Journal of Educational Technology in Higher Education, 18*(1), 1–17. https://doi.org/10.1186/s41239-021-00287-6

Lim, W. M. (2024). What is qualitative research? An overview and guidelines. *Australasian Marketing Journal.* Advance online publication. https://doi.org/10.1177/14413582241264622

Mirbabaie, M., Brünker, F., Möllmann, N. R., & Stieglitz, S. (2022). The rise of artificial intelligence—Understanding the AI identity threat at the workplace. *Electronic Markets, 32*(2), 1–27. https://doi.org/10.1007/s12525-021-00496-x

Ofosu-Asare, Y. (2024). Cognitive imperialism in artificial intelligence: Counteracting bias with indigenous epistemologies. *AI & Society*. Advance online publication. https://doi.org/10.1007/s00146-024-02065-0

Osasona, F., Amoo, O. O., Atadoga, A., Abrahams, T. O., Farayola, O. A., & Ayinla, B. S. (2024). Reviewing the ethical implications of AI in decision-making processes. *International Journal of Management & Entrepreneurship Research, 6*(2), 322–335.

Patel, S., & Ragolane, M. (2024). The implementation of artificial intelligence in South African higher education institutions: Opportunities and challenges. *Technium Education and Humanities, 9*, 51–65.

Paul, J. (2024). *Privacy and data security concerns in AI.* https://www.researchgate.net/publication/385781993_Privacy_and_data_security_concerns_in_AI

Pramjeeth, S., & Ramgovind, P. (2024). Generative artificial intelligence (AI) tools in higher education: A moral compass for the future? *African Journal of Inter/Multidisciplinary Studies, 6*(1), 1–13. https://doi.org/10.51415/ajims.v6i1.1560

Raheem, B. R., Anjum, F., & Ghafar, Z. N. (2023). Exploring the profound impact of artificial intelligence applications (QuillBot, Grammarly, and ChatGPT) on English academic writing: A systematic review. *International Journal of Integrative Research, 1*(10), 599–622. https://doi.org/10.59890/ijir.v1i10.366

Roche, C., Wall, P. J., & Lewis, D. (2023). Ethics and diversity in artificial intelligence policies, strategies, and initiatives. *AI and Ethics, 3*(4), 1095–1115. https://doi.org/10.1007/s43681-022-00218-9

Ruslin, R., Mashuri, S., Rasak, M. S. A., Alhabsyi, F., & Syam, H. (2022). Semi-structured interview: A methodological reflection on the development of a qualitative research instrument in educational studies. *IOSR Journal of Research & Method in Education, 12*(1), 22–29. https://doi.org/10.9790/7388-1201052229

Sabiteka, M., Yu, X., & Sun, C. (2025). Toward sustainable education: A contextualised model for educational technology adoption in developing countries. *Sustainability, 17*(8), 1–44. https://doi.org/10.3390/su17083592

Singun, A. J. (2025). Unveiling the barriers to digital transformation in higher education institutions: A systematic literature review. *Discover Education, 4*(1), 1–41. https://doi.org/10.1007/s44217-025-00430-9

Sobaih, A. E. E., Chaibi, A., Brini, R., & Abdelghani Ibrahim, T. M. (2025). Unlocking patient resistance to AI in healthcare: A psychological exploration. *European Journal of Investigation in Health, Psychology and Education, 15*(1), 1–27. https://doi.org/10.3390/ejihpe15010006

Southworth, J., Migliaccio, K., Glover, J., Glover, J. N., Reed, D., McCarty, C., ... & Thomas, A. (2023). Developing a model for AI across the curriculum: Transforming the higher education landscape via innovation in AI literacy. *Computers and Education: Artificial Intelligence, 4*, 100127. https://doi.org/10.1016/j.caeai.2023.100127

Turin, T. C., Raihan, M., & Chowdhury, N. (2024). Paradigms of approaches to research. *Bangabandhu Sheikh Mujib Medical University Journal, 17*(2), e73973. https://doi.org/10.3329/bsmmuj.v17i2.73973

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*(1), 28. https://doi.org/10.1186/s40561-024-00316-7

Zhang, Y., & Wildemuth, B. M. (2009). Qualitative analysis of content. *Applications of social research methods to questions in information and library science*, *308*(319), 1-12.

Zhang, Y., & Wildemuth, B. M. (2009). Qualitative analysis of content. In B. M. Wildemuth (Ed.), *Applications of social research methods to questions in information and library science* (pp. 308–319). Libraries Unlimited.

Zhou, X., Zhang, J., & Chan, C. (2024). Unveiling students' experiences and perceptions of artificial intelligence usage in higher education. *Journal of University Teaching and Learning Practice, 21*(6), 126–145. https://qmro.qmul.ac.uk/xmlui/handle/123456789/96258