Cybrarians Journals Issue 78, April 2026

Research

https://doi.org/10.70000/cj.2026.78.708

Empowering ethical innovation: designing a responsible computing workshop for emerging economies

Brenda van Wyk [1], **Marlene Holmner** [1], **Lesego Makhafola** [1], **Anika Meyer** [1], **Martie Mearns** [1]

[1] University of Pretoria, Department of Information Science, South Africa

brenda.vanwyk@up.ac.za https://orcid.org/0000-0003-3898-7042
marlene.holmner@up.ac.za https://orcid.org/0000-0003-1021-7601
lesego.makhafola@up.ac.za https://orcid.org/0000-0002-2983-1129
anika.meyer@up.ac.za https://orcid.org/0000-0001-5934-384X
martie.mearns@up.ac.za https://orcid.org/0000-0001-7362-3047

Article info

History
Received: 28.12.2025
Accepted: 01.03.2026
Published: 29.04.2026

Keywords
Responsible computing, artificial intelligence, generative AI, fourth industrial revolution, information ethics

Abstract

This paper reports on a multi-method collaborative autoethnographic (CAE) research project aimed at improving the understanding of the status quo of artificial intelligence (AI) and AI ethics in a university situated in an emerging African country. The rapid and disruptive technological developments of the Fourth Industrial Revolution (4IR), which include recent AI developments, have resulted in an ethics and skills gap among African researchers and tertiary students. In addressing this problem, the University of Pretoria embarked on a collaborative research project that includes reaching out to other African universities to plan and facilitate training workshops for academics and postgraduate students to enhance the effective use of AI. The multi-phased study primarily collected qualitative data through an initial scoping literature review to explore extant research on AI ethics globally and, specifically, in African Universities. Based on identified gaps, further confirmed by the scoping review results, a training workshop series was designed to offer practical and theoretical support towards fostering the responsible use of AI in research. Furthermore, the workshop was designed to equip and encourage researchers and students to research and test AI-powered tools and platforms in teaching, learning, and research. The workshop's design was informed by the scaffolded nature of the Zone of Proximal Development frame (ZPD). Pre– and post-workshop observations and surveys were used to add to the knowledge gained from the scoping review. The post-workshop survey was used to gauge the success and usefulness of the workshops, as reflected in the expectations of workshop delegates from the pre-workshop surveys. Data from the scoping review, the reflexive data from the CAE process, and workshop surveys were triangulated and thematically analyzed. Findings are that the workshop was informative and timely, and more interventions of this nature are needed to further entrench the real-life experience of ethical AI use in an academic environment.

Introduction

Since its advent in mid-2000, the Fourth Industrial Revolution (4IR) has been irreversibly restructuring our lived experiences and realities. This change includes the augmentation of disruptive technologies such as artificial intelligence (AI), the Internet of Things (IoT), blockchain, and big data analytics, to name but a few. These technologies rapidly redefine how economies operate and how education systems, healthcare services, and governance structures function. Floridi (2013) refers to these pervasive and rapid developments as an AI Spring, and in later works (Floridi, 2019, 2021) warns that these innovations are not yet tapering down, with added novel ethical dilemmas surfacing. Specifically, generative AI (GenAI) underpinned these changes and is increasingly appearing in commonplace applications. Not only has the disproportionate pervasiveness of these new technologies caused a widening digital exclusion in emerging economies, but it is also compounding ethical dilemmas such as ethics dumping and digital extractivism, further adding to already existing social injustices (Floridi, 2019; Ricaurte, 2022). For emerging economies, like South Africa, incorporating AI into higher educational teaching and learning is an uncontested necessity, and researching its impact on society at large is paramount.

It stands to reason that African economies, which include the role that universities play, must join forces and collaborate to strengthen collective capacity and skills and foster digital inclusion. Joint efforts could work towards a position of relevance and preparedness for the challenges in a global digital economy. Thus, new generations must be empowered with both the technical and ethical skills to face these challenges. This resulted in the University of Pretoria embarking on a research project to further explore training, development, and collaboration opportunities with other African Universities, including the University of Makerere.

Background

O'Neill et al. (2024) state that Africa's geo-statistics must consider situating the continent within the 4IR. With a growing population of 1.5 billion in 2025 and a significant youth bulge, as Onyeiwu (2024) explained, African countries face many challenges, such as unemployment, violence, and unstable politics. However, many authors are positive and see massive potential for technological evolution (O'Neill et al, 2024). Global inequalities remain entrenched and pervasive in Africa, and the historically based power dynamics perpetuated AI development and LLM training models. The dominant Imperial Core versus the exploited Imperial Periphery has offered a conceptual shortcut for the geographically inaccurate and oversimplified Global North or South. The dominance of GenAI developments and LLM training data from the Imperial Core results in digital exclusion of many African countries and other historically exploited economies of the Imperial Periphery (O'Neill et al., 2024). Literature reports an uptake of technological, digital, and AI education, especially in countries such as South Africa, Kenya, and Nigeria, but

many countries are lagging in closing skills gaps (Fundi et al., 2024). Authors such as Baidoo-Anu et al (2023) report an enthusiasm among younger populations to learn and apply AI technologies. However, a call for more action and interventions is imminent to ensure ethical, informed, and responsible use of technologies.

AI and ethics

Rooted in Normative Ethics, the applied ethics of AI-use addresses the moral and responsible use of AI tools and technologies. Principles such as beneficence, non-maleficence, autonomy, justice, and accountability are part and parcel of AI ethics (Van Wyk and Holmner, 2024). Authors report a prevalence of ethical dilemmas that include bias and discrimination, privacy and surveillance, accountability gaps, and the social impacts of automation (Floridi et al., 2018; Gosh, 2025; Mittelstadt et al., 2016). Ethical guardrails ensure AI promotes human well-being, equity, and transparency. Authors such as Floridi (2019) and Van Wyk and Holmner (2024) agree that more effort is needed towards enhancing the awareness and effectiveness of AI ethics to mitigate potential harm. Floridi (2019) stresses that global efforts to advocate AI ethics are critically important to ensure an informed information society. This is furthermore supported by the IEEE report, titled 'Ethically Aligned Design' [The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, (IEEE, 2019)], which states that the primary ethical considerations that should guide AI design are: implementing the most fundamental principles of human rights, emphasising the most significant advantage for humanity and the natural environment, and minimising risks and adverse effects as Autonomous and Intelligent Systems (A/IS) develop as socio-technical systems.

AI literacies towards ai competencies, fluencies and dexterities

Introducing AI, GenAI, and AI-powered academic tools and platforms requires careful consideration of engagement and risk awareness, regardless of the prevailing dominant Imperial Core epistemologies. Ng et al (2021) stress that training interventions must ensure that AI literacy is clearly understood, and they should not focus on using AI. They state that AI applications' underlying AI concepts and concerns are required to become a responsible citizen (Ng et al, 2021). AI literacy is an emerging field and is closely linked to ethics. Maimela and Mbonde (2025) state that AI systems may display bias in datasets and are reportedly implemented without the requisite cultural sensitivity. All projects and interventions must consider cognitive bias when AI is incorporated in training and teaching. Responsible interventions should not perpetuate social injustices. Responsible computing emphasises sustainability, ethical design, and social responsiveness. It advocates for technologies that automate efficiency and promote equity, transparency, and accountability. This approach has been operationalised through initiatives such as the Responsible Computing Workshop, a collaborative effort between the Departments

of Information Science and Philosophy at the University of Pretoria and the College of Computing and Information Sciences at Makerere University.

South Africa, as an upper-middle-income country, has a well-developed higher education system, and as such has a moral and ethical responsibility to lead in AI innovation on the African continent, teaching in global knowledge production and technological innovation. Despite historical inequalities and lingering disparities (see Maimela and Mbonde, 2025) that may still have implications for the equitable rollout of AI education, collaboration in the development and ethics across institutions and borders is critically important.

Over and above, ethics must be integrated into South African higher education; it must therefore be understood not only as a response to technological change, but as a proactive strategy for shaping a future in which digital innovation serves the broader goals of social justice and equitable human development. It is an endeavour that demands interdisciplinary collaboration and research, policy alignment, and grassroots engagement with communities and educators. In corroboration, the AI Policy Education Framework emphasises that institutions must go beyond reactive policies and develop strategic participatory models for AI literacy that reflect both global best practices and local needs (Chan, 2023). South Africa's role in leading here depends on our ability to invest in teacher and lecturer training, updating educational content and practical exposure to ensure equitable access to digital infrastructure across its deeply divided education system. It is evident that the development of AI literacies and competencies has become a critical skill to navigate a moral and political shift. With the right educational strategies, grounded in responsible innovation and ethical awareness, South Africa has the potential to lead in shaping an AI ecosystem that is not only technologically advanced but also socially just and inclusive. In order to do this, collaborations must be forged to equip individuals with the knowledge, critical perspective, and ethical grounding to engage with AI in empowering, rather than extractive ways. For Africa, and South Africa in particular, the stakes are high. Without widespread AI literacy and AI ethics, the continent risks being a passive recipient and consumer of imported technologies rather than elevating to a co-creator of equitable, sustainable digital futures. Without Africa's voice in AI and LLM training, the Imperial Core biases will deepen these historical power dynamics, further pushing Africa into the Imperial Periphery.

The workshop reported in this study focuses on developing educational frameworks that embed ethical principles. These principles cover data justice, bias mitigation, and sustainability, which must be reflected in the very design and implementation of AI content in curricula. Such efforts are essential for equipping educators and students with the tools to navigate AI's moral and societal implications in contexts where technology is not neutral, but deeply embedded in histories of inequality and exclusion.

Theoretical frames and models informing this study

The study draws on established theoretical frameworks and models that provide both a conceptual foundation and practical direction to guide the researchers in planning, designing, and developing an intervention in the form of a training workshop. These include Lewin's Theory of Action Research and Vygotsky's Zone of Proximal Development (ZPD) (Lewin, 1946; Van Wyk et al, 2020; Vygotsky, 1929). Both of these learning theories provided a sound foundation to help formulate research questions, which aided in reaching a conclusion and making recommendations. Emerging frameworks were identified after the completion of the study based on the patterns and insights that emerged from the analysed data. It is important to note that in this study, the theoretical perspectives taken were used to guide the development of the training intervention rather than to test or validate existing theories.

Lewin's theory emphasises a cyclic process of planning, acting, observing, and reflecting, aimed at solving social problems through participatory collaboration and reflective inquiry (Burnes, 2020). Action research integrates qualitative approaches to deepen understanding of the lived experience and social contexts, aiming to generate knowledge and effect meaningful change through collaboration and learning interactions with stakeholders. Whereas Vygotsky's ZPD is a foundational concept in educational psychology that describes the difference between what a student or subject can do independently with existing knowledge and what they can achieve after training, where knowledgeable persons offer guidance and collaborative learning. Typically, such a knowledgeable person will plan the workshop and may act as facilitator and mentor. The ZPD refers to the "zone" or area where learning is most effective, because it is just beyond the student's current competence or skills. However, appropriate skills can be achieved with the correct support or intervention (Van Wyk et al., 2020). One could argue that a workshop designed alongside the principles of proximal development should expose participants to new skills and knowledge that they have not fully mastered, or improve skills that are not fully achieved. In this case, a workshop was planned to enhance responsible and ethical computational use in participants' studies, supported by Vygotsky's ZPD and Lewin's Action Research model to ensure both learner-centeredness and responsiveness according to the participants' needs.

Research questions

The following research questions were formulated in preparation for the responsible computing workshop and for the write-up of this paper:

1. What gaps are identified in AI ethics awareness, literacy, and responsible use of AI among staff and postgraduate students in an African university?

2. How can the identified gaps be addressed by training interventions (e.g. workshops) to enhance the responsible computing practices of staff and postgraduate students in emerging economies?
3. To what extent does participation in a responsible computing workshop improve staff and postgraduate students' knowledge of AI ethics, literacies and responsible computing?

Research design and methodology

An inductive research study typically does not start with a theory to prove. Instead, inductive research begins with data collection and observation that may result in the formation of a theory (Bucher, 2021). Collaborative autoethnography (CAE), complemented by a scoping literature review, was employed as the methodological approach for the study. CAE originates from the approaches of autobiography and ethnography (Ellis, Adams and Bochner, 2011). According to Mouton (2022), ethnographic studies are usually qualitative in nature and aim to provide an in-depth description embedded in the lived worlds of the participants or actors. This approach allows for a deep emergence in self-experience and reflection. As the study's actors, the researchers are all part of the Department of Information Science at the University of Pretoria. As a qualitative method, the CAE was suitable for this study as it facilitated the gaining of insights and a rigorous understanding of lived experiences in a longitudinal study (as posited by Roy and Uekusa, 2020). The roles of the collaborating researchers varied throughout the project, depending on the phase of the project. Roles that can be differentiated are Researcher One handled the communication and outreach with African Research Universities, Researcher two, three and four researched the topic, while Researcher One and Five conducted quality control. Researchers met at regular intervals to discuss the progress of the project. The study followed a two-phased process. The first phase focused on completing a scoping review to establish the topics that would need to be included in a training workshop focusing on responsible computing, addressing ethics and integrity in teaching and learning, and identifying the latest AI and GenAI content. The second phase employed CAE to collaboratively explore the researchers' reflections, dialogues and experiences pertaining to the requirements, objectives and desired outcomes of a training workshop for enhancing responsible computing. In this phase, the researchers also used the ZDP framework to monitor and evaluate the success of the training interventions.

The first phase of the research project – conducting a scoping review

As part of the process for building the curriculum for the workshop, a scoping review was conducted as a basis to structure content and actions. The PRISMA extension for scoping reviews (PRISMA-ScR) was used to map evidence on the topics to be covered (Page et al., 2021). The use of the PRISMA-ScR also ensured a high standard of reporting and enhanced the reproducibility of results acquired. Items included from the PRISMA-ScR are: eligibility criteria, search strategy,

selection of sources of evidence, data charting, data items, critical appraisal of individual sources of evidence, and synthesis of results.

Eligibility criteria for inclusion in the scoping review

The criteria used when searching for relevant literature include: the research focus, type of published works to be included, format, databases to be included, language of publication, and peer review. Scopus, Web of Science, and ProQuest were the databases used to search for literature. For ProQuest, only the ERIC, Library and Information Science Abstracts (LISA), and Library Science databases were included. Studies included were published between 2022 - 2025, written in the English language, available in full-text, have undergone peer review, are electronic journal articles, and focus on higher education institutions, specifically universities and colleges. Studies that did not meet the inclusion criteria stated, were excluded from consideration. Exclusion criteria were indicated to be K-12 schools, high schools and technical vocational education and training (TVET) colleges to remain aligned with the specific tertiary educational context intended for this study. Other exclusion criteria were pre-2022 research, research published in languages other than English that were not peer reviewed or made only abstracts available.

Search strategy

A comprehensive search was conducted on the selected databases to find relevant results. The keywords used were responsible computing, AI literacy, AI ethics, generative AI, research intelligence, research data management, research integrity, and responsible conduct of research. Boolean operators were used to construct a search string to ensure that the most relevant results could be retrieved. Where possible, synonyms were added.

The search string was used to search in the title, abstract, and keywords, where the database included the feature. (“generative AI”) AND (AI literacy) AND [(AI ethics) OR (“research integrity”) OR (“responsible research”) OR (“responsible conduct of research”)] AND (research intelligence) AND (Research management). Table 1 represents the results obtained from the database searches.

Table 1: Represents the results of the articles included in the study

Database	Initial search results	Records limited to inclusion criteria	Duplicates removed
Proquest	227	48	8
Scopus	3	1	0
Web of Science Core Collection	1722	536	1

Total results	**1952**	**575**	**9**

Knowledge Synthesis of Qualitative Data from the Scoping Review

Covidence, a specialised tool for systematic evidence synthesis, was used in the first phase of this study to assist in the collaborative screening of eligible studies for the scoping review. This allowed the researchers to extract data, assess study quality, and manage review workflows effectively. From the searches on the selected indexing databases, 585 studies were imported for screening. Ten duplicates were removed, and 575 titles and abstracts were screened, removing 160 irrelevant studies (Figure 1). Three reviewers were involved in this process, and a Cohen's Kappa value of 0.35584 and 0.53917 was achieved, resulting in a moderate intercoder reliability score. In-depth screening for constructs of AI literacy, AI and research ethics, academic misconduct and research integrity followed. With a total of 415 articles remaining to be screened for full text, 272 were excluded, resulting in 143 papers selected for the review.

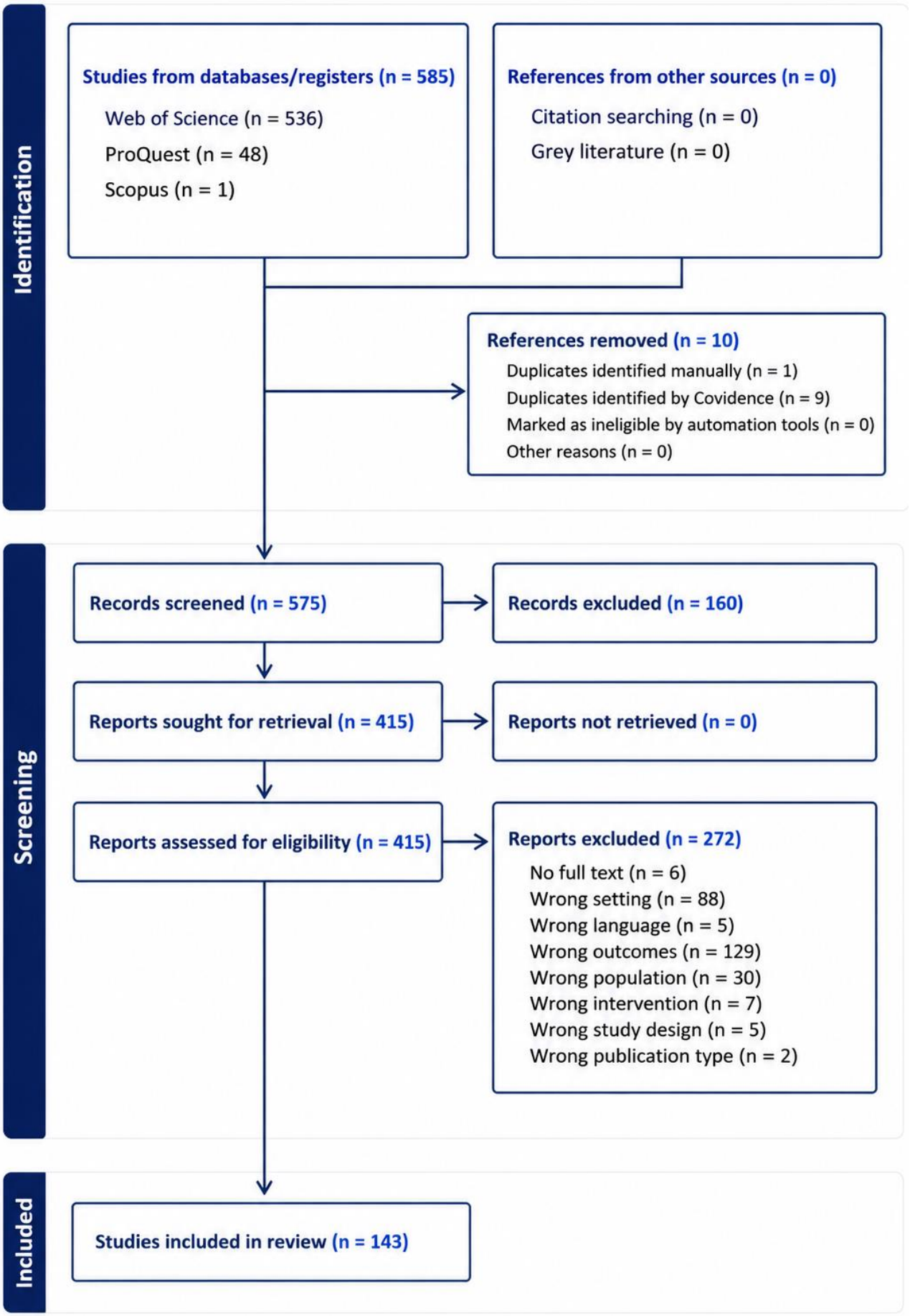
Identification
Studies from databases/registers (n = 585)
Web of Science (n = 536)
ProQuest (n = 48)
Scopus (n = 1)
References from other sources (n = 0)
Citation searching (n = 0)
Grey literature (n = 0)
References removed (n = 10)
Duplicates identified manually (n = 1)
Duplicates identified by Covidence (n = 9)
Marked as ineligible by automation tools (n = 0)
Other reasons (n = 0)
Screening
Records screened (n = 575)
Records excluded (n = 160)
Reports sought for retrieval (n = 415)
Reports not retrieved (n = 0)
Reports assessed for eligibility (n = 415)
Reports excluded (n = 272)
No full text (n = 6)
Wrong setting (n = 88)
Wrong language (n = 5)
Wrong outcomes (n = 129)
Wrong population (n = 30)
Wrong intervention (n = 7)
Wrong study design (n = 5)
Wrong publication type (n = 2)
Included
Studies included in review (n = 143)

Figure 1: PRISMA flow diagram for the scoping review

Inter-coder reliability on full-text screening was at a fair level, with Cohen's Kappa values of 0.24037 and 0.30679. McHugh (2012) reports a score as low as 0.41 to be acceptable. A score below this level requires acknowledgement, which is a limitation of this research.

The main clusters identified were research integrity, research misconduct, plagiarism, culture, and academic ethics. Other terms linked to the studies included artificial intelligence, tools, and higher education. Research integrity has a central position as a popular topic, bridging issues of misconduct and ethics with emerging concerns about AI tools' use in education and research, especially when it comes to ghost-writing and authorship attribution.

Themes Identified

Though there is a proliferation in publications on research integrity and responsible conduct of research, there is a marked paucity of research done by researchers at African universities, particularly on responsible computing and the value of AI ethics training interventions. After thematic analysis, the following themes could be identified:

1. Research integrity
2. AI and academic misconduct
3. Using AI within the parameters of research ethics
4. AI literacy and quality of research
5. Trustworthy research
6. AI use and plagiarism
7. AI guidelines in academic writing
8. AI and ghost-writing
9. AI policy and guardrails for researchers and postgraduate students

The scoping review depicts the areas of accountability and transparency in using AI-empowered tools and platforms in higher education institutions. Whereas phase two focused on postgraduate students, the scoping review shed light on the responsibility of tertiary decision makers' policy and practice to create the supporting environment in terms of developing policies, upholding standards, and ensuring recurriculation for relevant course content. It highlighted that ethics and a range of literacies cover how knowledge is created, interpreted, ideally to ensure responsibility and accountability in high-technology tertiary ecosystems and learning environments.

Armed with a better understanding of researched trends, identified gaps, and ethical dilemmas faced, the research project was ready to move to the next phase.

Phase two of the study – designing a training intervention

This paper focuses on an aspect of a larger AI project of the Department of Information Science. The aim and objective of the larger study is to research AI ethics and literacies. Phase two of the paper reflects on the development of the content and the methods of delivering the workshop. It was necessary to establish whether the knowledge transferred and skills developed during the workshop would result in greater responsible computing practices of the delegates. The workshop was offered as a way to introduce and encourage the concept of responsible computing, covering the entire research cycle, AI tools, and responsible and ethical research practices that can be applied at each stage of conducting research. The workshop was meant as an introduction that can ultimately be internalised, and new competencies can be acquired for independent application in further research and studies. By measuring participants' post-training performance, the study seeks to assess both the immediate success of the instructional strategies used and the extent to which the workshop fostered meaningful learning and skill development juxtaposed against expressed expectations and skills reported in the pre-workshop survey.

Collaborative Autoethnography Reflections to Inform Workshop Design and Content

Over 10 months, the five researchers engaged in CAE to gather lived experiences, engage in mutual sense-making, and use collective reflection to identify the requirements, desired outcomes, and delivery method of a training workshop for enhancing responsible computing. These narratives and reflections were built on the broader themes and gaps identified in the scoping literature review. Utilising Analytical Induction (AI) iteratively, the researchers refined emerging insights from the CAE and the scoping literature review, ensuring that the workshop design and content were both experientially authentic and grounded in established research evidence.

The main source of error in ethnographies, including CAE, is a researcher's potential bias or the lack of research rigour (Mouton, 2022). As a result, reflexivity is applied from a methodological perspective as a process of critical self-reflection on the researchers' biases, theoretical predispositions and preferences (Schwandt, 2007b). Fetterman (2020) states that reflexivity propels the researcher into a state of continual analysis throughout the research process. The benefit of conducting a CAE is highlighted by Miyahara and Fukao (2022), which results in increased awareness and reflexivity and a better understanding of the researchers' own practice and beliefs. Such collective exploration resulted in deeper learning about self and others, when

challenging their own beliefs, elaborating on ideas and experiences through self-inquiry and shedding light on possible blind spots, towards more engaged research.

Applying the ZDP Framework to Identify Learning and Training Opportunities

After reflecting on the nature of the workshop, the researchers followed the ZDP Framework (Figure 2) to plan, monitor and assess success, and identify learning and training opportunities. Figure 2 illustrates the scaffolding zones where learning occurs, namely: *Zone of Actual Development* (blue circle) - tasks within the student's actual competence; *Zone of Proximal Development* (solid green line circle) - tasks achievable with support; and *Beyond Current Reach* (dotted blue line circle) - tasks beyond their current capacity. Using learning strategies and theories, findings of the scoping review, and reflections from CAE, the researchers designed a training workshop, spanning three days of in-person sessions. The workshop sought to be interactive, engaging and collaborative. An online pre-workshop survey was sent to all participants to gauge expectations and training needs.

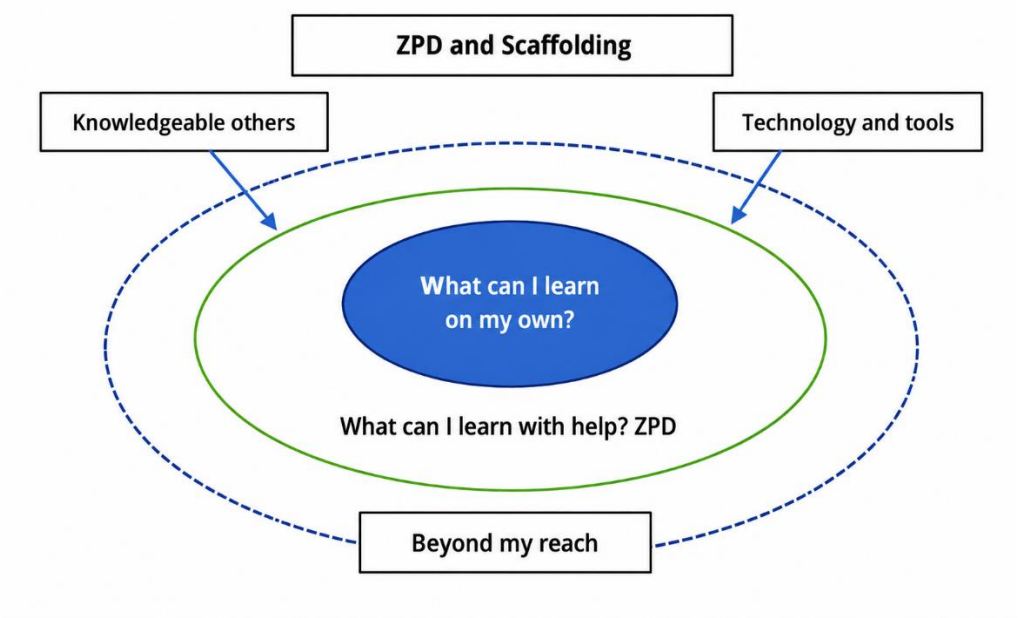

Figure 2. Adapted from Vygotsky's Zone of Proximal Development (1978) in Ong and Annamalai (2024)

The purpose of the online pre-workshop was to engage potential participants. Both open-ended questions and multiple-choice questions were asked to gauge their current roles, existing knowledge, further needs and expectations. There were ten questions and 20 responses. For the post-workshop survey a further ten questions were asked on the participants experiences and 18 responses were recorded. This pre-workshop survey noted what participants already knew, thus

noting the *Zone of Actual Development*. The researchers were more interested in identifying the *Zone of Proximal Development*, which constituted what can be learnt with guidance in the form of workshop sessions and practical activities. The scaffolded approach of the ZPD framework highlights the following opportunities for learning and training interventions:

1. Introduction (Known Zone to Zone of Actual Development) - Quick recap of existing knowledge;
2. Demonstration (Within the ZPD) - Facilitator models the new skill;
3. Guided Practice (Within the ZPD) - Students attempt with scaffolding and peer collaboration;
4. Independent Task (Beyond ZPD, skills now mastered) - Apply skill solo;
5. Reflection and Feedback (Across all Zones) - Discuss learning process, challenges, and next steps.

Based on the findings of the scoping review, the researchers' collective autoethnographic reflections, using the ZDP framework to identify learning and training opportunities, as well as the pre-workshop survey, the workshop was designed to cover the following learning areas, as explained in Table 2:

Table 2: Workshop Training Areas and Content Covered

Training Areas	**Content**
1. Ethics and AI Ethics and GenAI Ethics in the Information Age	PAPAS Model, AI and LLM Bias
2. Practical sessions - Designing Images, Writing Text and Crafting Conversations	Claude, ChatGPT, Deepseek, Scite, Gemini, Copilot, MetaAI, DALE-3, ResearchRabbit, NotebookLM, Humanizer
3. An Overview of the Philosophical Origins of Information Ethics and the Practice of AI Literacy in Higher Education	Philosophy of Information (PI), Social justice, Digital colonialism and extraction
4. Re-evaluating AI Literacy Challenges in the Global South and Developing Economies	Social cognitive justice, Asilomar Principles, Metaliteracy,
5. Positioning AI Literacy Facilitation and Programmes in HEI	AI literacy Frames, Metaliteracy Frames
6. Research Data Management and FAIR Principles	RDM, Figshare, FAIR, Openness

7. Responsible Research Practices	Research Ethics, Belmont Principles and AI
8. Using AI-powered Research Tools Responsibly and Effectively	Julias-AI, data analysis, data visualisation
9. How AI can be used to derive Research Intelligence for Competitive Advantage to Inform Policy and Practice	RIMS, Institutional Repositories, research impact factors, University rankings, Research visibility
10. Ethical use of AI in Research Writing, avoiding AI-ghostwriting, Plagiarism and Academic Misconduct	Turnitin and plagiarism detection shortcomings, quality of paraphrasing tools, paper mills,

Workshop feedback

The pre- and post-workshop feedback of the programme was reflected to establish the expectations of the delegates and to assist the facilitators in tailoring the responsible computing workshop to their needs. Feedback during the workshop was requested at the conclusion of each day on all sessions presented. Feedback was received to evaluate the programme's delivery and content, which aligned with delegates' expectations. For the pre-workshop portion, feedback was received from 20 delegates, and the post-workshop feedback included reflections from 18 delegates.

Delegates were requested to provide feedback on their current position or role at the higher education institution, specifically whether the delegate is a student or staff member, their data analysis experience and background, as well as their past experiences and expertise using GenAI tools. Their understanding of AI Ethics, Responsible Computing and Research, as well as research data management, and their knowledge of Findable, Accessible, Interoperable, and Reusable (FAIR) principles were determined. Feedback also included what their expectations and learning goals are for the workshop. On the question of how skilled they feel in ethics, all participants responded that they lack knowledge and understanding.

In the online pre-workshop survey, participants were requested to state their prior knowledge of AI use. Only three delegates responded to having used AI regularly for personal and work projects, while 14 respondents tried the basic tasks, such as asking a simple question to a GenAI tool. Three respondents indicated total unfamiliarity with using AI tools. Those who have heard of AI or tried basic AI tools mostly identified ChatGPT as the tool that they have used. The responses on the comfort level they feel using AI are completely split, with half of the respondents indicating not being comfortable with using AI and the other half expressing somewhat comfortable or very comfortable. Respondents were asked to describe the tasks they currently

perform or wish to perform in the near future using GenAI. Most of the responses indicated using GenAI to draft documents such as literature reviews, research proposals and emails, or using it to paraphrase or edit their written work. Considering the responses to drafting literature reviews and research proposals, the most significant ethical challenges that respondents mentioned were plagiarism, accountability and transparency, data protection, and violation of privacy. In terms of skills readiness with responsible research, FAIR data, data governance policies, data quality and data integrity, respondents in almost all cases indicated that they either heard about the term but do not know details, or they have a basic understanding thereof. From these responses, it was clear that the level of exposure to these concepts was minimal, with the group of delegates expected to attend the workshop. All respondents expressed a desire to gaining hands-on experience with DALL·E, ChatGPT, Scite.ai, and other AI tools; to understand how to integrate AI tools into their research workflows; to learn best practices for data governance and FAIR compliance; to be able to explain AI ethics to colleagues; to networking with peers to form future collaborations; and to understand and be able to implement and/or explain responsible computing to others. The challenges that respondents expressed they have resorted mostly revolved around AI ethics and learning how to avoid plagiarism when using AI or knowing how much AI is acceptable to include in research.

In the post-workshop feedback, delegates were requested to reflect on their experiences during the Responsible Computing Workshop. While totally anonymous, their role as a student or staff member could be matched with their responses to the extent to which each of the sessions met their expectations. Open-ended feedback was included to allow narratives on their experiences. In the post-workshop survey, participants rated the quality of the workshop from good to excellent. On a question of which aspects and sessions they found valuable, participants replied as follows:

1. Most participants replied that both theory and practical sessions on AI ethics and responsible use of AI were informative;

2. Participants also valued the practical applications of tools;

3. Data analysis using AI was popular with postgraduate students currently doing research.

On a question, what participants would like to see in future workshop responses varied. Some replied that they would like more information on AI literacies and how to detect the unethical use of AI. The following testimonials were offered:

> *"It has been an eye-opening experience"*
>
> *"The presenters were excellent – we enjoyed the practicals [sic]"*
>
> *"This helped me to better understand the use of AI in academia"*

> *"This workshop has opened my mind and passion for responsible use of AI, tools and that it is not a crime to use them as long as you acknowledge that you have used them"*

Recommendations for future workshops included more hands-on, practical sessions in facilitated learning, with opportunities to try out various tools. This suggestion, however, is hampered by poor data connectivity and slow data speeds in the venue used. Venues such as computer labs were also unavailable, and delegates either did not bring their own devices or did not have devices they could bring to the venue. This lack of infrastructure and the availability of devices created a significant barrier to implementing such recommendations.

Delegates also requested more days to be added to workshops to allow for the processing of the information and for more time to elaborate and discuss the content. They also requested more workshops per year, which were offered more frequently than only once a year. During one of the collaboration sessions, the discussion about institutional policies revealed that either specific policies about the use of AI are lacking in the institution, or awareness of these policies results in delegates not knowing the institutional stance on the use of AI. During a student session, a participant expressed the concern that we have come to teach the students how to use AI, and we have equipped their lecturers with knowledge and skills to catch the students out if they do use AI. The concept of responsible use of AI has not yet been penetrated, and either AI use needs to remain clandestine, or lecturers should remain ignorant of students' indiscriminate use of AI. The technological utopia versus dystopia appears to be ingrained in the perception of the student versus the lecturer. It remains a perspective to address through ethical human choices and the development of responsible practices.

Further analysis

The researchers were able to analyse all collated data to gauge the workshop's success. In making sense of the findings, Mishra's (2025) categorisation of AI users was useful. This recent development by Mishra (2025) is a framework that looks at four categories or different types of knowledge positions and information behaviour of users' experience with AI. His framework is hugely valuable for planning future AI training interventions, as it sets up four scenarios to understand the challenges people encounter when using AI tools in Figure 3.

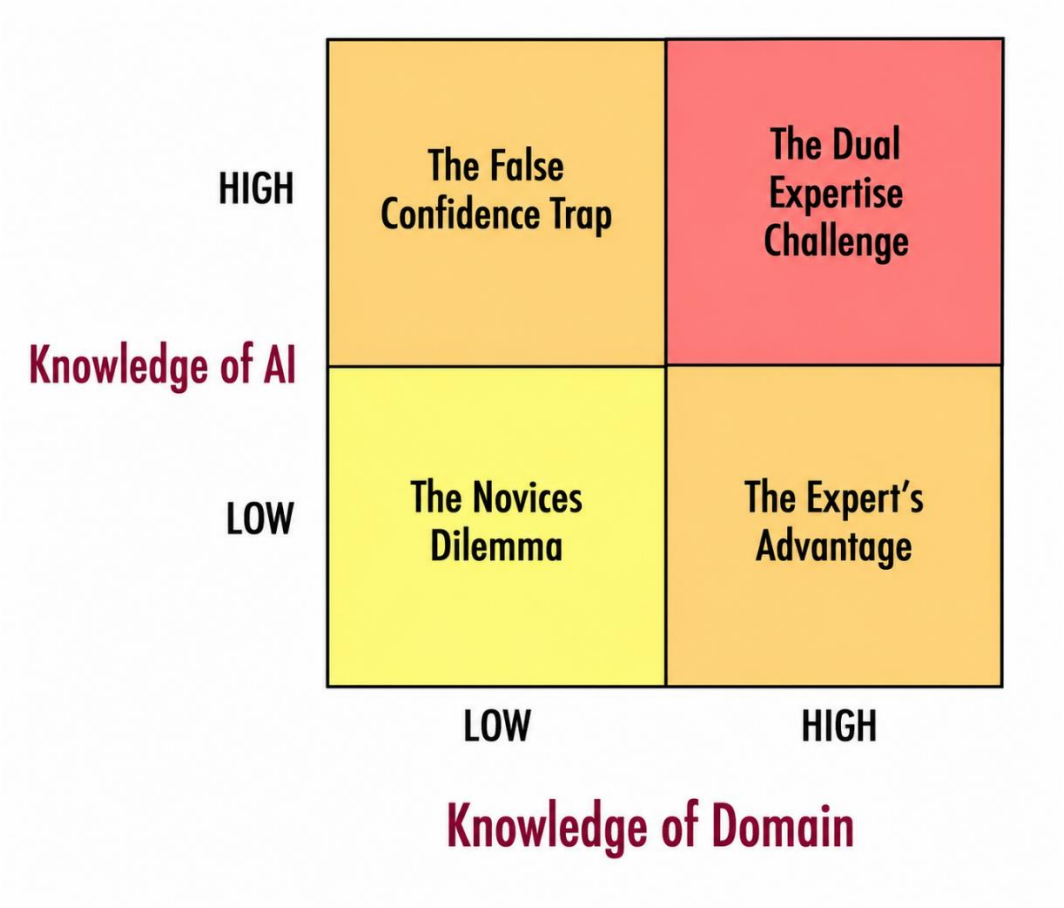

Figure 3: The Interaction between Domain Expertise and AI Knowledge (Mishra, 2025)

1. The Novice's Dilemma

This concerns students who do not have AI expertise and do not have disciplinary knowledge. They cannot evaluate the AI's outputs for accuracy and do not know when AI leads them astray because they do not understand what is within the discipline or subject area.

2. The Expert Advantage

The expert's advantage refers to those with disciplinary expertise and subject knowledge, but may not be AI literate. This is a better position because learning how to use AI is easier than building disciplinary expertise. It means you can start supporting your students even before you have full AI literacy.

3. The False Confidence Trap

The false confidence trap involves those highly knowledgeable in AI but not in the discipline or subject where it is being applied. They are overconfident in dealing with AI tools but do not have the literacies or competence to know when AI is hallucinating, supplying superficial information

or information that is out of scope. Without discipline, knowledge users cannot create prompts that could supply deeper insights.

4. The Dual Expertise Challenge

The dual expertise challenge is the scenario that all learning activities or interventions aim to achieve. People who are highly knowledgeable in both AI and their discipline are the most engaged when working with AI because they spend a lot of time thinking about not just which tool they are using, but what it is giving them. Having the discernment of discipline knowledge, AI prompts can be engineered intuitively and creatively to optimise accuracy and depth of responses.

This categorisation applies to both postgraduate students as well as supervisors and lecturers. The observations stemming from this study are that participants fall within the first two groups, namely novice dilemmas and expert advantage. More participants fall within the novice dilemma group. Only two participants, who are postgraduate students and are also teaching, appeared to fall in the expert advantage group. The researchers did not experience the false confidence gap category during the workshop, but researchers could identify with this grouping based on prior teaching experience. Whereas the ZDP assisted the researchers in making sense of the planning, designing and execution of the workshop to ascertain if the objective was achieved, Mishra's (2025) new frame is providing insights for future interventions.

Conclusion

The CAE study followed a phased process, collecting data from the scoping review and workshop surveys, and most importantly, from observations and group deliberations. The training workshop was especially revealing, and much was learnt from the workshop in surfacing the gaps in AI ethics awareness and literacy among staff and postgraduate students. There is a correlation between the scoping review findings and the feedback from the workshop. Students demonstrated eagerness to adopt AI and GenAI tools, but also admitted uncertainty about ethical and effective use, at times showing overreliance and blind trust. Educators, by contrast, expressed hesitancy and, in some cases, actively discouraged AI use. These findings highlight a clear divide in readiness between staff and students, and underscore the value of workshops in exposing and beginning to address these gaps.

The design of the training intervention proved effective in transferring AI ethics knowledge. Its phased structure, combined with reflexivity, observation, hands-on activities and group deliberation, allowed participants to confront ethical concerns in real-life scenarios. More importantly, the study was guided by Mishra's (2025) new framework, which provides a foundation for shaping future interventions. This suggests that training grounded in both ethics and literacy, and which is scaffolded by established (Vygotsky's ZDP framework) and emerging

frameworks (Mishra, 2025), can help to prepare educators and students in emerging economies for responsible computing practices. During workshop participation and engagement, participants' understanding of AI ethics and responsible computing improved, and this is evidenced in greater recognition of ethical uncertainty, acknowledgement of risks such as overreliance, and openness to reflecting on future practice. While these shifts do not resolve all the challenges, they show that carefully designed training interventions can meaningfully enhance awareness and preparedness. More broadly, the findings emphasise that educators should move beyond traditional teaching and research practices and adopt future-oriented approaches. Prior knowledge is no longer sufficient when addressing AI, given its accelerated development and its impact on society. AI is not merely a set of tools and platforms but is reshaping norms, values, and educational practices. Moreover, HEIs need to clarify their institutional and departmental policy stance on AI use as a matter of guiding educators as they support students in their investigations into AI use and research. This requires collaboration and support from a management and governance perspective.

This study's value lies in both its methodological richness and the epistemological contribution to an under-researched and still misunderstood developing area of research. The layered reflexivity offered a greater understanding of the issue being studied, and created the platform to understand the inter-relations between disruptive technologies, literacies, teaching and learning, and expectations in real-life scenarios.

Recommendations

The findings of this study align with the concerns mentioned by Mishra (2025) and co-researchers (Mishra, Oster and Henriksen, 2024) in recognising the vulnerabilities of students' ill-informed and undirected use of AI. Closing this literacy and skills gap not just for students but also for governance and educators, particularly in emerging ecosystems, calls for in-depth research and further discourse and collaborations.

Training interventions in universities in emerging economies should be participatory, context-sensitive, and sustainable. They should build individual skills and embed capacity into institutional frameworks, ensuring long-term impact. As both the participating universities in this study are from emerging economies, further collaboration, scaffolding and participative efforts are being planned, and more research of this nature is required.

References

Adelman, C. (1993). Kurt Lewin and the Origins of Action Research, *Educational Action Research*, Vol. 1, Issue 1, pp. 7-24, https://doi.org/10.1080/0965079930010102

Allen, M. (2017). Intercoder reliability techniques: Fleiss system. *In The Sage Encyclopedia of Communication Research Methods*, Vol. 4, Issue 10, pp. 739-740. Thousand Oaks, CA, USA: SAGE Publications, Inc., https://doi.org/10.4135/9781483381411.n257

Baidoo-Anu, D., Asamoah, D., Amoako, I. and Mahama, I. (2024). Exploring student perspectives on generative artificial intelligence in higher education learning. *Discovery Education,* Vol. 3, Issue 98, pp. 98, https://doi.org/10.1007/s44217-024-00173-z

Biagini, G. (2025). Towards an AI-Literate future: a systematic literature review exploring education, ethics, and applications. *International Journal of Artificial Intelligence in Education, pp. 1-51,* https://doi.org/10.1007/s40593-025-00466-w

Bucher, J. (2021). Inductive Reasoning. *The Encyclopedia of Research Methods in Criminology and Criminal Justice*, Vol. 1, pp. 200-204.

Burnes, B. (2020). The origins of Lewin's three-step model of change. *The Journal of Applied Behavioral Science*, Vol. 56, Issue 1, pp. 32-59.

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, Vol. 20, Issue 1, pp. 38.

Cooper, R. and Lilyea, B. V. (2022). I'm Interested in Autoethnography, but How Do I Do It?. *The Qualitative Report*, Vol. 27, Issue 1, pp. 197-208. https://doi.org/10.46743/2160-3715/2022.5288

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT, USA: Yale University Press.

Dlamini, R. and Ndzinisa, N. (2025). Towards a critical discourse on artificial intelligence and its misalignment in sub-Saharan Africa: Through an equality, equity, and decoloniality lens. *Journal of Education (University of KwaZulu-Natal)*, Vol. 98, pp. 42-61.

Edwards, J. (2021). Ethical autoethnography: Is it Possible? *International Journal of Qualitative Methods*, Vol. 20, Issue 3. DOI: 10.1177/1609406921995306

Ellis C., Adams T. E. and Bochner A. P. (2011). Autoethnography: an overview. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, Vol. 12, Issue 1, p.10 http://nbn-resolving.de/urn:nbn:de:0114-fqs1101108

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality. First Edition*. Oxford, UK: Oxford University Press.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. and Vayena, E. (2018). AI4People—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, Vol. 28, Issue 4, pp. 689–707. https://doi.org/10.1007/s11023-018-9482-5

Floridi, L. (2019). What the Near Future of Artificial Intelligence Could Be. *Philosophy and Technology*, Vol. 32, Issue 1, pp. 1-15. https://doi. org/10.1007/s13347-019-00345-y

Floridi, L. (2020). AI and its new winter: from myths to realities. *Philosophy and Technology,* Vol. 33, Issue 1, pp. 1-3. https://doi.org/10.1007/s13347-020-00396-6

Fundi, M., Sanusi, I. T., Oyelere, S. S. and Ayere, M. (2024). Advancing AI education: assessing Kenyan in-service teachers' preparedness for integrating artificial intelligence in competence-based curriculum. *Computers in Human Behavior Reports,* Vol. 14, pp. 100412. https://doi.org/10.1016/j.chbr.2024.100412

IEEE. (2019). The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. In: *IEEE Standards Association*. https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

Laupichler, S., Paulus, A., Rösch, A. G. and Pinkwart, N. (2022). What is AI literacy? Towards a definition of AI literacy for non-technical higher education students. In *Proceedings of the 13th International Conference on Computer Supported Education (CSEDU 2022)* Vol. 2, pp. 55–66. SCITEPRESS. https://doi.org/10.5220/0011064000003179

Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues*, Vol. 2, Issue 4, pp. 34-46.

Long, D. and Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). https://doi.org/10.1145/3313831.3376727

Maimela, C. and Mbonde, P. (2025). Artificial Intelligence (AI) in South African Universities: Curriculum Transformation and Decolonisation–Aid or Obstacle?. *Frontiers in Sociology*, Vol. 10, pp. 1543471. https://doi.org/10.3389/fsoc.2025.1543471

McHugh M.L. (2012). Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. Vol. 22, Issue 3, pp. 276-82. PMID: 23092060; PMCID: PMC3900052.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society*, Vol. 3, Issue 2, pp. 1–21. https://doi.org/10.1177/2053951716679679

Mishra, P. (2025). *The GenAI and Expertise Paradox: Why It Makes Expert Work More Important But Harder*. Blog: Punya Mishra, Thursday, February 13, 2025.

Mishra, P., Oster, N. and Henriksen, D. (2024). Generative AI, Teacher Knowledge and Educational Research: Bridging Short- and Long-Term Perspectives. *TechTrends,* Vol. 68, pp. 205–210 https://doi.org/10.1007/s11528-024-00938-1

Miyahara, M. and Fukao, A. (2022). Exploring the use of collaborative autoethnography as a tool for facilitating the development of researcher reflexivity. *System*, Vol. 105, pp. 102751. https://doi.org/https://doi.org/10.1016/j.system.2022.102751

Mouton, J. (2022). *How to succeed in your Masters and Doctoral Studies: A South African guide and resource book*. Second Edition. Pretoria: Van Schaik.

Ng, D. T. K., Leung, J. K. L., Chu, K. W. S. and Qiao, M. S. (2021). AI literacy: Definition, teaching, evaluation and ethical issues. *Proceedings of the Association for Information Science and Technology*, Vol. 58, Issue 1, pp. 504-509.

O'Neill, J., Marivate, V., Glover, B., Karanu, W., Tadesse, G.A., Gyekye, A., Makena, A., Rosslyn-Smith, W., Grollnek, M., Wayua, C. and Baguma, R. (2024). *AI and the Future of Work in Africa White Paper*. arXiv preprint arXiv:2411.10091.

Ong, Q.K.L. and Annamalai, N. (2024). Technological pedagogical content knowledge for twenty-first century learning skills: The game changer for teachers of industrial revolution 5.0. *Education and Information Technologies*, Vol. 29, Issue 2, pp.1939-1980. https://doi.org/10.1007/s10639-023-11852-z

Onyeiwu, S. (2024). Population, Youth Bulge and Economic Development. In *Emerging Issues in Contemporary African Economies*, Cham, Switzerland: Palgrave Macmillan, pp. 123-141.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E. and Chou, R. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Review,* Vol. 10, pp. 89. https://doi.org/10.1186/s13643-021-01626-4

Poerwandari, E.K. (2021). Minimizing bias and maximizing the potential strengths of autoethnography as a narrative research method. *Japanese Psychological Research*, Vol. 63, pp. 310-323. https://doi.org/10.1111/jpr.12320

Ricaurte, P. (2022). Ethics for the majority world: AI and the question of violence at scale. *Media, Culture and Society*, Vol. 44, Issue 4, pp. 726-745. https://doi.org/10.1177/01634437221099612

Roy, R. and Uekusa, S. (2020). Collaborative autoethnography: 'self-reflection" as a timely alternative research approach during the global pandemic. *Qualitative Research Journal*, Vol. 20, Issue 4, pp. 382-392. DOI 10.1108/QRJ-06-2020-0054

Schwab, K. (2016). The Fourth Industrial Revolution. Crown Business.

Schwandt, T. A. (2007a). Analytic induction. In *The SAGE dictionary of qualitative inquiry*. Third Edition, Thousand Oaks, CA, USA: SAGE Publications, Inc. https://doi.org/10.4135/97814129862681.n6

Schwandt, T. A. (2007b). Reflexivity. In *The SAGE dictionary of qualitative inquiry*. Third Edition, Thousand Oaks, CA, USA: SAGE Publications, Inc. https://doi.org/10.4135/97814129862681.n298

Van Wyk, B and Holmner, M.A. (2024). AI and the Ethics of Tomorrow. *In Proceedings of the International Conference on AI Research*, Vol 4, Issue 1. https://doi.org/10.34190/icair.5.1.3017

Van Wyk, B., Mooney, G., Duma, M. and Faloye, S. (2020). Emergency remote learning in the times of COVID: A higher education innovation strategy. *In Proceedings of the European Conference on e-Learning*, Berlin, Germany (pp. 28-30).

Vygotski, L.S. (1929). II. The problem of the cultural development of the child. *The Pedagogical Seminary and Journal of Genetic Psychology*, Vol. 36, Issue 3, pp.415-434.