Research Article

American Journal of Science Education Research

Ethical Artificial Intelligence in Education: A Southern African Framework Decoding Algorithmic Through Governance Lens

Mfanelo Ntsobi[1], Bongani June Mwale[2*], Kholekile Ntsobi[3]

[1]Sci-Bono Discovery Centre, Corner of Miriam Makeba and Helen Joseph (formerly Newtown |, Helen Joseph St), Johannesburg, South Africa
[2]Sci-Bono Discovery Centre, Corner of Miriam Makeba and Helen Joseph (formerly Newtown |, Helen Joseph St), Johannesburg, South Africa
[3]School of Business, DaVinci Institute for Technology Management, 16 Park Ave, Johannesburg, South Africa

[*]**Corresponding author**: Bongani June Mwale

Abstract
This paper explores the social and ethical considerations of artificial intelligence (AI) as it integrates into education and society. It examines challenges such as data privacy, algorithmic bias, AI trustworthiness, and human agency. The literature highlights context, human agency and the importance of diverse stakeholder involvement in AI governance, AI literacy, responsible education, and strategies for ethical assessment and mitigation. A literature review of recent articles and policy documents informs this study, focusing on AI's evolving role in education. The research develops an input-process-output (IPO) framework to address ethical concerns at each stage of AI development. The IPO model outlines the ethical implications for input, process, and output phases. Section one addresses AI's social implications. Section two examines AI educational policy using United Nations Education, Scientific and Culture Organisation's (UNESCO's) guidelines as a benchmark for member states. Section three explores ethical considerations in AI development and usage. The final section presents the AI IPO Ethical Framework, detailing ethical responsibilities at each stage. The study underscores the role of policymakers, researchers, and higher education institutions in shaping AI's ethical trajectory. It emphasizes responsible AI implementation, ensuring that AI systems are developed and deployed with ethical considerations in mind. The proposed framework serves as a guiding tool for assessing ethical risks and ensuring responsible AI integration in education. By fostering AI literacy and ethical awareness, this study contributes to ongoing discussions on AI ethics, advocating for transparent, fair, and accountable AI practices. It ultimately aims to support the ethical advancement of AI in education and governance.

Introduction

Artificial Intelligence (AI) as a technology that gives rise to ethical questions throughout its development and use. In other words, there is necessity for ethical principles that determine and evaluate its proper use. For example, [**Error! Reference source not found.**] situate the notion of "proper use" in the ambit of technology or utilitarian ethics and highlight four governing principles: "proper use of technology, responsibility, fairness, and cost." To institutionalize and regulate these AI principles, among others that we will discuss, requires collaboration at policy, industry and societal levels, among governments, technology companies, educators, ethicists, and other stakeholders. To define AI precisely, as per the United Nations Educational, Scientific and Cultural Organisation [**Error! Reference source not found.Error! Reference source not found.**]:"AI systems are information-processing technologies that integrate models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in material and virtual environments." Additionally, these systems "are designed to operate with varying degrees of autonomy by means of knowledge modelling and representation and by exploiting data and calculating correlations." Moreover, they have "the capacity to process data and information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning or control." The ethical issues surrounding the design and use of artificial intelligence (AI) have become more important as it becomes more common in schools, government, and society as a whole. AI has the potential to improve decision-making, make services more accessible, and make learning more personal [38]. However, it also poses difficult ethical issues about data protection, algorithmic fairness, transparency, and responsibility. The Input–Process–Output (IPO) Ethics Framework is a complete paradigm for finding and dealing with ethical hazards at every stage of the AI lifecycle in response to these problems.

Significance of AI Ethics

The ethical issues surrounding the design and use of artificial intelligence (AI) have become more important as it becomes more common in schools, government, and society as a whole. AI has the potential to improve decision-making, make services more accessible, and make learning more personal. However, it also poses difficult ethical issues about data protection, algorithmic fairness, transparency, and responsibility [38]. The Input–Process–Output (IPO) Ethics Framework is a complete paradigm for finding and dealing with ethical hazards at every stage of the AI lifecycle in response to these problems.

Artificial Intelligence Input-Process-Output Model

The aforementioned UNESCO definition of Artificial Intelligence indicates that AI operates through a systematic framework of inputs, processes, and outputs. [**Error! Reference source not found.**]; programming [**Error! Reference source not found.**]; accounting and financial management [**Error! Reference source not found.**]; and business solutions [**Error! Reference source not found.**], among many other sectors. This paper proposes that the IPO model could also be used to embed ethical considerations in the development and use of AI. In other words, ethics would not only be considered as regulations or principles prior to an AI activity, or as post-activity compliance evaluation. Instead, ethics compliance would be tracked at every stage of AI development and usage. The figure below illustrates how AI operates, using an IPO approach, and it is the point of departure for the IPO ethics framework that we will suggest at the end of this paper.

INPUT (Data Collection & Preprocessing)	PROCESS (Data Processing & Model Execution)	OUTPUT (AI-Generated Results & Actions)
• Raw Data (Text, Images, Audio, Video, Sensor Data, etc.) • Predefined Rules or Knowledge Base • Human Feedback or Annotations • Real-time Inputs from Users or Environment	• Data Cleaning & Feature Extraction • Model Selection & Training (Machine Learning, Deep Learning) • Pattern Recognition & Decision-Making • Optimization & Fine-Tuning	• Predictions, Recommendations, or Classifications • Generated Text, Images, or Speech • Automated Actions or Responses (e.g., Chatbots, Robotics) • Insights or Reports for Decision-Making

Source: Author computed

Table 1: *Artificial Intelligence Input-Process-Output Model*

In addition to the ethical questions of AI development and use, is the underlying foundation of responsible AI education, in both formal and informal settings, for digital industry practitioners and the public at large. AI literacy usually focuses on teaching programming and coding starting from the K-12 schooling leaves. However, the public who are concerned about the implications of automated decision-making systems are often excluded in this AI literacy drive.

The use of AI in automated decision systems (ADS) is increasing rapidly, breeding job anxiety and job security concerns among both citizens and policymakers because of the ethical and legal challenges these technologies present [**Error! Reference source not found.**]. In light of this gap, [**Error! Reference source not found.**] suggest a "stakeholder-first' approach that considers the target audience and their context when rolling out AI literacy initiatives. In this way, AI education engages a diverse audience and addresses the societal challenges posed by AI systems. Thus, this paper focuses on the social and ethical considerations of AI, on the one hand, and their implications for education, on the other.

Social Considerations of AI

The United Nations 17 Sustainable Development Goals are tailored to advance various areas of society, including the goal of quality education. Thus, policymakers, scientists, educators, communities and civil society at large are required to contribute their expertise and experience to facilitate the required development. Those who are specifically involved in education, in various capacities, have the opportunity to contribute to teaching and learning at a policy level by firstly helping educators and students to understand the socio-political context within which education takes place. This can be achieved by combining STEM or applied sciences, social science and humanities subjects that afford both technical and human knowledge. The teaching and learning of artificial intelligence are no different. Technical proficiencies have to be gained simultaneously with interpersonal competencies, and socio-cultural values and expectations should facilitate ethical and effective integration. In this way, the social considerations of AI in education and research are factored in, and AI literacy and awareness is a crucial starting point. Important concepts in this regard include responsible AI, AI education, ethical AI, AI literacy, AI fairness, and AI accountability [**Error! Reference source not found.**].

Notably, many of these socio-ethical considerations are interconnected and require a holistic approach to address them effectively. However, AI integration gives rise to the intricate interplay between "AI usage, purposes, difficulties, and their impact on attitudes, perceptions, and future implications" that have ethical, social, and educational dimensions [**Error! Reference source not found.**]. By recognizing how AI is utilized, understanding its objectives, and addressing related challenges, educational stakeholders can maximize its benefits while ensuring its responsible and effective use in teaching and learning environments. Appropriate AI integration in education embraces a human-centered approach, where AI tools enhance, rather than replace, educators [**Error! Reference source not found.**]. This approach fosters collaboration, where for example, AI can be used for administrative tasks, while educators focus on student engagement and mentorship [**Error! Reference source not found.**].

Holmes et al. [**Error! Reference source not found.**] highlight a globally significant gap in the field of Artificial Intelligence in Education (AIED), the paucity of research, guidance, policy development, or regulation specifically addressing the ethical challenges posed by AI in education. They offered the following background for their observation:

The AIED conferences in 2018 and 2019 hosted a workshop called "Ethics of AIED. Who Cares?" (Holmes et al., 2019a,19) Although the workshop discussions were engaging and productive, the small number of attendees suggested a disappointingly low level of interest in the topic from the broader AIED community. While other AI communities are increasingly attending to ethical considerations around the design and deployment of AI-based technologies, ethical dimensions of AIED do not yet appear to be a central area of focus for many in the AIED community [**Error! Reference source not found.**].

Interestingly, since the workshops mentioned in the passage above, AIED ethics has received increasing attention, where the majority of national AI policy strategies do include dedicated sections on AI ethics more broadly [**Error! Reference source not found.**, **Error! Reference source not found.**]. This widespread emphasis on AI ethics may be attributed to the growing scholarly and public interest in ethical concerns surrounding AI over the past decade.

In the process of AI development, both Large Language Models (LLMs) and human annotators are respectively instrumental in shaping AIED. The LLM design influences the capabilities and potential risks of AIED, whereas the ethical perspectives and values of annotators influence whether the model serves society positively or negatively [**Error! Reference source not found.**]. Annotators are particularly responsible for biases related to culture, gender, or race that may be introduced when categorizing and labelling educational content, which go on to being embedded within AIED platforms. For example, an AI-driven history education tool that is trained on biased annotation data may distort or misrepresentation historical events or figures, which ultimately affects students' factual knowledge, understanding and perspectives [**Error! Reference source not found.**]. Conversely to ethical historical representation, a future ethical concern about AI is that it is creating a new world that our children will grow up in – a fundamentally different world. This raises further existential questions about human experience, mutual-and-self-understanding, interaction, engagement, autonomy, agency and dignity [**Error! Reference source not found.**]. In this respect, Stefania Giannini, the UNESCO Assistant Director-General for Education, cautions:

This study not only examine current developments in these technologies but also envision their implications 20 or 30 years hence. How do we reconcile the necessity of preparing youth for a human-machine world with preserving the integrity of the human intellect while we delegate certain cognitive tasks? The study cannot permit experimentation on an entire generation. Digital innovations should be engineered to safeguard human agency (Giannini, 2023:3).

AI and Educational Policy
By 2021, over 30 nations have implemented national artificial intelligence (AI) policy frameworks that delineate the prospective impacts of AI across diverse policy domains, including education, while also considering its social and ethical ramifications. [**Error! Reference source not found.**]. The author conducted a thematic analysis of 24 national AI policy strategies, examining the role of education in global AI policymaking and found that discussions on AIED are largely absent from these policies. Instead, the focus is predominantly on using education as a tool to develop an AI-ready workforce and to train more AI specialists [**Error! Reference source not found.**]. While AI ethics was a recurring theme in these policy documents studied, the specific ethical considerations of AIED receive little attention. Schiff [**Error! Reference source not found.**] subsequently argues that effective policymaking must be closely aligned with ethical considerations and he applies a framework of five AI ethics principles to explore how policymakers can better integrate AIED's ethical and societal impacts. Schiff divides the five principles into two domains, namely: Education for AI and AI for Education (or AIED), respectively:

Education for AI (i.e., training)
- **Training AI Experts:** discussion of developing future AI practitioners, such as computer scientists and engineers.
- **Preparing the Workforce for AI:** discussion of education and training efforts to help workers adapt to labour disruption due to AI.
- **Public AI Literacy:** discussion of the need to educate the broader public about AI.

AI for Education (i.e., AIED)
- **Teaching and Learning:** discussion of AI-based teaching and learning tools such as intelligent tutoring systems, pedagogical agents, and predictive assessments.
- **Administrative Tools:** discussion of AI used to support administration in educational systems, for example, to make admission, promotion, or graduation decisions [**Error! Reference source not found.**].

A major task for all countries and their education departments is to develop comprehensive AI-ready policies and curriculums. This is not only a process of developing policy and programme documents, but it also entails forging relevant partnerships with other sectors, as well as raising and allocating funds to the entire process. This process and procedural integration of AI ethics informs our proposed IPO Ethical Framework that we will suggest and outline later. Even though concerns such as bias, privacy, and transparency are widely discussed in AI-related policy documents [**Error! Reference source not found.**], these social and ethical considerations are rarely linked directly to AIED. Instead, a pro-innovation perspective dominates, which asserts that intelligent educational systems will ensure "inclusive, renewed, and adapted training" for students and teachers [**Error! Reference source not found.**]. Overall, despite growing attention to AIED, there remains a significant gap in addressing both its transformative potential and ethical challenges. Thus, collaborations between government policymakers, educational institutions, researchers, educators, industry and communities are crucial.

The Role of UNESCO in AI Educational Policy
Artificial Intelligence (AI) offers transformative potential for education, enhancing teaching, learning, and progress toward SDG 4. However, UNESCO supports Member States in leveraging AI to achieve the Education 2030 for rapid advancement which is significant to risks that outstrip current policies (UNESCO, 2019). For UNESCO to achieve is mandate it must consider human-centred approach to AI, this will include ethical considerations in the use of AI. AI's role in addressing current inequalities incorporates access to knowledge, research and the diversity (Tawil & Miao, 2024). The cultural expressions to ensure AI does not widen the technological divides within and between countries is enable through cooperation and adoption by all countries to address issues of access to AI, technological revolution, innovation and knowledge (Dinker, 2024) .

The United Nations Educational, Scientific and Cultural Organization (UNESCO) has published a comprehensive framework and guidelines to assist countries in supporting students and teachers in understanding both the potential and the risks of AI. These resources emphasize ethical considerations, aiming to promote responsible and inclusive use of AI in education (Mutawa & Sruthi, 2025). The UNESCO global ethical themes as adopted in Beijing Consensus include:

1. **Cultural diversity**: AI should respect and promote cultural diversity, ensuring that technologies are inclusive and do not reinforce cultural biases or marginalize minority voices
2. **Education:** AI must support equitable, quality education, fostering critical thinking, digital literacy, and ethical awareness among learners and educators.
3. **Scientific knowledge:** The development and use of AI should advance open science, promote access to scientific knowledge, and encourage transparency and accountability in research.
4. **Communication and information:** AI systems must uphold freedom of expression, access to reliable information, and media pluralism while combating misinformation and digital exclusion
5. **Peace:** AI should be used to foster dialogue, understanding, and conflict prevention, contributing to peaceful and inclusive societies.
6. **Sustainability:** AI must align with environmental sustainability goals, minimizing ecological impacts and supporting climate action and sustainable development
7. **Gender:** AI should promote gender equality, challenge stereotypes and ensuring that systems are free from gender bias and inclusive of all identities.

Artificial Intelligence (AI) holds colossal promise for advancing human development, particularly in education, science, and communication (Marques-Cobeta, 2024) . AI deployment must be guided by strong ethical principles to ensure fairness, inclusivity, and respect for human rights. (Sargiotis, 2024). The Beijing Consensus on Artificial Intelligence and Education, adopted under UNESCO's guidance, emphasizes the importance of aligning AI with core ethical values such as transparency, accountability, equity, and sustainability. It calls on countries to develop policies and practices that harness AI's potential while safeguarding against its risk ensuring that AI serves as a tool for the common good and contributes to achieving the Education 2030 Agenda (Yadav et al ;2024).

More recently in August 2024, UNESCO introduced guidelines on AI competence in education, emphasizing the importance of ethical considerations. The guidelines stress that educators must be equipped with knowledge of fundamental ethical principles, regulations, institutional frameworks, and moral guidelines related to AI. Yan & Liu [**Error! Reference source not found.**] note that the "fragmented nature of existing research makes it challenging to develop a comprehensive ethical framework" and so in their technical analysis of the AI development process, they found significant similarities between AI development and human education. We also argue that the generalizable IPO framework works the same in both traditional analogue education and digitally enhanced pedagogy. Thus, Cukurova and Miao [**Error! Reference source not found.**], developed the *AI Competency Framework for Teachers* [**Error! Reference source not found.**] for UNESCO that suggests principles to be adhered to be adhered to in the rapidly evolving field of AI ethics, and its growing significance in the education sector

In light of the in-progress revolution of education in the age of AI, companies that provide digital educational services are mushrooming, meaning that they too should be subject to AI ethics. In tandem to the growth of the AI sector, urgent regulations are required for the protection of children given the disruptions that accompany the innovations. To this end, UNESCO published the first-ever Guidance for Generative AI in Education and Research (September 2023), which complements the Recommendation on the Ethics of Artificial Intelligence [**Error! Reference source not found.**] and A Guidance for Policy-Makers on AI and Education [**Error! Reference source not found.**].

Ethical Considerations of AI
Ethics integrate both positive and negative rights, that together enable and restrict in order to promote beneficial social goals, as well as mitigating risks and harms. This requires a delicate balancing act, and in the realm of education this is done by integrating ethics into AIED policy. Policy and law serve as essential tools for translating ethical values into societal practice, which ensures the protection of rights, expansion of opportunities, and upholding societal principles and values [**Error! Reference source not found.**]. Not only are ethics devoted to preventing harm in education, such as addressing discrimination and bias, but they also involve actively fostering inclusivity and enhancing opportunities for students and educators [**Error! Reference source not found.**].

The goal of AI ethics should be to nurture an ongoing learning journey that is agile and adapts to the evolving landscape of AI, and not to merely to be a restrictive mechanism. The integration of AI ethics in various sectors, is not just about the imposition of strict rules, but is an ongoing, dynamic learning process that enhances users' understanding and responsible engagement with AI technology. Dabbagh et al [**Error! Reference source not found.**], therefore argue that "AI ethics should be mandatory for schoolchildren" to foster a continuous educational journey that evolves alongside advancements in AI. This approach embeds AI ethics as a fundamental part of technology education. Students are encouraged to actively engage with, question, and critically assess AI systems. Moreover, they develop a deeper understanding of how AI functions, the ethical challenges it presents, its societal implications and how to apply this knowledge in real-world situations. The end-goal of AI ethics education is empowerment that cultivates an informed, ethically conscious, and technologically skilled generation [**Error! Reference source not found.**].

An ethical evaluation of AI helps to formulate principles that provide orientation regarding the teleology of what ought to be done, morally. According to [**Error! Reference source not found.**], ethical questions surrounding AI can be explored through multiple approaches. First, he argues that they can be examined by weighing the potential benefits and promises of AI against the possible risks and harmful consequences associated with its design and use (see [**Error! Reference source not found.**]). Second, he asserts that this is seen in established fields like bioethics and medical ethics, where ethical discussions develop a set of duties and responsibilities which inform principles that guide moral decision-making [**Error! Reference source not found.**]. Third, he maintains that ethical evaluation can also involve defining ideals and a positive vision for AI, setting goals for its responsible use or even aiming to create AI systems that embody ethical virtues [**Error! Reference source not found.**].

[**Error! Reference source not found.**] adds that these three approaches are not mutually exclusive and are often combined in ethical debates. For example, in discussions about the need to understand, explain, and control AI, different philosophical

ethical frameworks can be applied, and in this regard, he argues that:

*Ethical reflections and arguments in scholarly publications as well as in policy documents and tech industry guidelines [4] often proceed as a discussion of (a) risks and opportunities, or of (b) rules and principles, or of (c) visions and ideals of AI, mirroring the three different normative theories that shape the tradition of Western moral philosophy: consequentialism, deontology and virtue ethics [**Error! Reference source not found.**].*

In simple terms, a consequentialist approach focuses on the consequences of a particular action or process, and whether it promotes the overall good and mitigates bad outcomes. In contrast, a deontological approach is concerned with determining the difference between right and wrong course of action, prior to acting. Virtue ethics, on the other hand, speak to the moral responsibilities of AI actors such as developers, practitioners, educators and users. Often these approaches are used in combination, in order to promote the good, the right, and the moral, collectively.

Ethical considerations are crucial when integrating AI into everyday environments, particularly regarding its impact on individuals and society [**Error! Reference source not found.**, **Error! Reference source not found.**]. Dignum, [**Error! Reference source not found.**], underscores the need to uphold societal values, assess moral and ethical implications, and ensure transparency in AI decision-making processes. Seo et al [**Error! Reference source not found.**] similarly highlight concrete concerns related to accountability, agency, surveillance, and potential privacy violations associated with AI systems. Regarding d**ata privacy,** AI systems depend on vast datasets that often include sensitive personal information, raising significant far-reaching concerns. To mitigate these risks, institutions must establish strong data governance policies, incorporating measures such as advanced encryption, secure storage protocols, and compliance with privacy regulations. In order to mitigate these challenges, human-centred values help to encourage the responsible and ethical development and deployment of AI technologies, irrespective of context [**Error! Reference source not found.**, **Error! Reference source not found.**].

Institutions must have robust data governance policies. A thorough examination of the ethical implications of AI in higher education is necessary to pinpoint the primary difficulties that must be tackled for responsible AI implementation. These challenges—encompassing data privacy, algorithmic bias, and the digital divide—necessitate strong governance and deliberate responses [**Error! Reference source not found.**]. Regarding algorithmic bias, one overarching concern is the trustworthiness of AI in education in terms of the potential bias of algorithms [**Error! Reference source not found.**]. In the South African context, McNulty (2024) notes some concerns regarding the role of AI in the classroom and biased decision-making due to algorithms that are trained to be biased, and that assess students unfairly and inaccurately. Thus, there is a need for transparent and auditable use of AI systems that can safeguard against any potential misuse of student data.

Pertaining to the digital divide, it is common knowledge that access to IT and AI in the past has been a privilege enjoyed by those with elite education or who originate from higher socio-economic population groups. Thus, the drive to scale up AI

education access is aimed at addressing this digital divide. Moreover, Butler-Adam (2018:1) emphasizes that "to succeed as a member of society, and as an employee, in the era of the Fourth Industrial Revolution, numeracy, literacy and an understanding of how the world operates are all essential." The world is increasingly becoming automated and without foundational knowledge of AI, many will be left behind. Given the complexity of AI, the value of transparency facilitates the process of tracking AI usage throughout any given process and accounting for the ethical and social considerations throughout the chain of development and usage. Buitem [**Error! Reference source not found.**] also emphasizes the importance of increasing transparency in input data, algorithm testing, and decision-making models to improve the interpretability of machine learning algorithms. Our IPO Ethical Framework that will be shared in the following section attempts to provide such transparency.

The study concurs with [**Error! Reference source not found.**], who argue that it is crucial to enhance the comprehension of stakeholders of a new educational AI system by making all information and decision processes available to those affected. As such, [**Error! Reference source not found.**] highlight the importance of addressing key ethical concerns – such as accountability, explainability, fairness, interpretability, and safety – when integrating AI into educational technology. Additionally, [**Error! Reference source not found.**] stress that it is paramount for stakeholders to fully understand new educational AI systems, and for this, transparency is essential and requires that all relevant information and decision-making processes should be accessible to those impacted by these technologies.

Finally, in the context of formal education, it is important to safeguard the interests and information security of students, parents, teachers, and schools through data privacy and security. To protect this privacy, it is essential to focus on high-risk AIED systems that could compromise data security [**Error! Reference source not found.**]). With AI tools like ChatGPT becoming widely used in student assignments and exams, universities have increasingly adopted AI-powered plagiarism detection tools such as Turnitin and ZeroGPT to maintain academic integrity and uphold educational standards [**Error! Reference source not found.**].

Costa et al. (2024) undertook an extensive and illuminating study of the challenges, benefits and recommendations for the use of GenAI in academic writing, with ChatGPT as their case study. Even their focus was the South African context, they conducted a literature review of the phenomenon in 22 countries and were able to synthesize findings that have global applicability. In summary, their findings and recommendations were as follows:

The findings reveal that while ChatGPT can significantly enhance productivity, improve language support for non-native English speakers, and provide immediate feedback, it also poses risks related to plagiarism, accuracy, and over-reliance on AI. Ethical concerns are paramount, with the need for transparent disclosure of AI use, proper attribution, and policies to prevent misuse. To address these issues, the study proposes the OTHA (Openness, Transparency, Honesty, and Accountability) Framework. This framework aims to guide the ethical integration of ChatGPT by fostering a collaborative approach, ensuring transparency in AI use, promoting ethical training, and establishing accountability mechanisms. The study recommends

comprehensive policies, targeted training, equitable access to AI tools, and regular monitoring and evaluation to ensure responsible AI use in academic settings. By implementing these recommendations, educational institutions can leverage the benefits of ChatGPT while maintaining academic integrity and promoting inclusive educational practices (Costa, et al., 2024:3).

The study agrees with the potential advantages of ChatGPT as described by the authors and their suggested OTHA Framework, which guides the integration of the digital tool while considering the ethical ramifications of its use. They concluded that OTHA fosters "collaboration, ensuring transparency, promoting ethical use, and maintaining accountability, institutions can leverage the benefits of AI while mitigating its risks" and also "enhances educational quality but also upholds the integrity and inclusivity of academic work, aligning with the broader goals of higher education" (Costa, et al., 2024:34).

However, we highlight a fundamental difference between their framework and the AI IPO Framework that we propose. We note that the OTHA Framework demonstrates a categorical imperative, similar to philosopher Immanuel Kant's seminal categorical imperative – the universal principle that we should seek to act towards others as would want them to act towards us. Our AI IPO Framework is instead undergirded by a pragmatic imperative – one that is procedural and focused on the practical and operational application of specific social and ethical considerations for each key step of AI development and use. They key steps we have identified as illustrated in Figure 1 are:
1. INPUT: Data Collection & Preprocessing
2. PROCESS: Data Processing & Model Execution
3. OUTPUT: AI-Generated Results & Actions

Thus, when one utilizes GenAI such as ChatGPT for example, the AI IPO Framework would require the user to account for the entire IPO chain of AI data custody and application. Although this paper focuses on education and AIED specifically, we recognize that research is an extension of the educational process, particularly in higher education and professional research institutes. In this regard, we touch on the ethical considerations in digital qualitative research as succinctly outlined by Ntsobi et al. (2024):
The realities of utilising digital tools and technologies in qualitative research are not merely a mirror or extension of physical realities as they create new contexts, accompanied by ethical choices and

dilemmas. On the one hand, they increase the intimacy of the research process by reducing the distance between the researcher and the participant. On the other hand, they create an open-ended process that cannot be fully predicted prior to implementation [...]. As such, these technologies create assumptions regarding identity, transparency, consent, and the delineation between private and public spaces that may be viewed differently by the researcher and participant [1]. To ensure ethical standards are met, it is essential to respect protocol, consult participants adequately, obtain informed consent, and maintain transparent processes and procedures. However, over and above, the confidentiality and sensitivity involved in using these technologies are crucial [11]. Ethical principles pertain to the use of software in data analysis, the organisation of importing of data, the interpretation, description, codification, and questioning of the data, as well as the exporting of the data [2]. These are new implications and considerations that accompany digital tools and spaces (Ntsobi, et al., 2024:20).

To ensure the responsible integration of AI in higher education, it is essential to establish ethical guidelines and regulatory frameworks. AI should be developed and implemented in ways that uphold student autonomy and agency, while also preventing the deepening of societal inequalities [**Error! Reference source not found.**]. The ethical collection and use of student data should be a priority, with transparency and accountability guiding AI applications. Significantly, AI should complement rather than replace human interaction and support in educational settings. Therefore, the need for ethical guidelines and regulatory measures in education is both urgent and imperative.

Input, Process, Output (IPO) AI Ethics Framework

The Input–Process–Output (IPO) Ethics AI Governance Framework is a conceptual paradigm that helps policy makers and scholars think about how to build and use AI systems in an ethical way, especially in schools and communities. The framework breaks the AI lifecycle into three parts: Input, Process, and Output. The framework categorises the AI lifespan into three interrelated stages: Input, Process, and Output, each encompassing unique ethical implications. Each phase of the life cycle has a distinct emphasis: the input phase primarily concentrates on data collection and preprocessing, the process phase addresses data processing and model execution, and the output phase yields AI-generated results and actions, while the IPO Ethics Framework introduces ethical considerations for each stage:

INPUT (Data Collection & Preprocessing)	PROCESS (Data Processing & Model Execution)	OUTPUT (AI-Generated Results & Actions)
Raw Data (Text, Images, Audio, Video, Sensor Data, etc.) Predefined Rules or Knowledge Base Human Feedback or Annotations Real-time Inputs from Users or Environment	Data Cleaning & Feature Extraction Model Selection & Training (Machine Learning, Deep Learning) Pattern Recognition & Decision-Making Optimization & Fine-Tuning	Predictions, Recommendations, or Classifications Generated Text, Images, or Speech Automated Actions or Responses (e.g., Chatbots, Robotics) Insights or Reports for Decision-Making
ETHICAL CONSIDERATIONS	ETHICAL CONSIDERATIONS	ETHICAL CONSIDERATIONS
AI policy, regulations, governance Compliance with relevant data protection regulations Informed consent Privacy protections Data minimization (Collecting only necessary data) Transparency about data usage Avoiding bias in data selection Non-biased annotation data Accountability for data handling Bias (Lack of Diversity in Training Data)	"Stakeholder-first" model Fairness and Non-Discrimination Accountability, Responsibility for machine learning Ensuring data accuracy Transparency in input data Algorithmic Transparency, Interpretability and Explainability Data Security Discrimination (Amplification of Existing Biases) Moral decision-making duties and responsibilities (Ethical Theories)	Accountability and Responsibility (Who is Responsible for AI Actions?) Misinformation and Disinformation (Deepfakes and Synthetic Media, AI-Generated Propaganda and Fake News) Intellectual Property and Copyright (Ownership of AI-Generated Content, Copyright Infringement) Social and Economic Impact (e.g. Human-machine relations, Trust in Information, Digital Divide, Job Displacement)

Source: Author (Ntsobi, MP 2025[1]

Table 2: *Artificial Intelligence Input-Process-Output Ethics Framework*

This IPO Ethics Framework underscores the fact that ethical AI is a continuous process that extends throughout the AI lifecycle, rather than a one-time concern. In order to guarantee that AI systems are designed and utilised in a manner that promotes equity, enhances human dignity, and fosters trust in technology, it is imperative that developers, educators, regulators, and civil society collaborate. This paradigm offers a practical approach to responsible AI governance, particularly in Southern Africa, where AI adoption is intertwined with social transformation, educational reform, and policy innovation.

Summary of the AI IPO Ethics Framework
The Artificial Intelligence Input-Process-Output (IPO) Ethics Framework (Table 2) offers a systematic method for integrating ethical issues throughout the whole lifecycle of AI systems, encompassing data gathering, processing, and final outputs. The framework guarantees that ethical considerations, including transparency, fairness, accountability, and privacy, are addressed at every phase of AI development and implementation. The framework consists of three main stages: Input, Process, and Output, each accompanied by relevant ethical issues.

The **Artificial Intelligence (AI) Input-Process-Output Ethics Framework** provides a comprehensive ethical roadmap for AI development. It ensures that AI systems align with principles of **fairness, transparency, accountability, and data protection** across all stages of their lifecycle. By embedding ethical considerations into AI policies, data governance, model development, and output evaluation, this framework helps mitigate risks and enhances AI's positive contributions to society. Implementing such a structured ethical approach is crucial for responsible AI integration, particularly in high-impact sectors such as education, healthcare, and finance.

Conclusion
The teaching, learning, and education landscape has been transformed and revolutionized in no small way due to artificial intelligence (AI). With AI technologies becoming increasingly sophisticated, they bring both opportunities and challenges, which give rise to pressing ethical and social considerations. This paper examined some of the key considerations and explored how AI can be implemented responsibly while ensuring transparency, accountability and inclusivity. We analysed selected recent literature, policy guidelines, and ethical concepts to map out the socio-ethical terrain of AI. Subsequently, we developed an input-process-output (IPO) ethical model to provide a structured approach for embedding ethics at every stage of AI development and use in education.

The **introduction** established the premise of this study, by emphasizing that AI technology should be governed by ethical principles to ensure positive societal outcomes. The importance of collaboration among stakeholders, including policymakers, educators, researchers, and industry professionals was highlighted, to ensure that AI systems align with ethical, regulatory and governance standards. The introduction also introduced a general artificial intelligence input-process-output (IPO) model (Figure 1), which demonstrates the three stages of AI development and usage, as a point of departure for the AI IPO Ethical Framework we later develop. The first section on **social considerations of AI** explored the broader implications of AI's integration into education. AI literacy was identified as the means to ensure that students, educators, and the public are capacitated to engage with AI responsibly. Given the prevalent potential for algorithmic bias and misinformation in AI-driven decision-making systems, the danger of reinforcing existing inequalities was highlighted. The discussion further emphasized the importance of a holistic approach to AI education that

---

[1] This table (Figure 2) was also developed with the use of ChatGPT and is a demonstration of the utility of Generative AI (GenAI) provided that the process is transparent and that the user takes accountability for the end-product produced. The ChatGPT prompt used was a request for an outline of the ethical considerations for each stage of the input-process-output (IPO) outlined in Figure 1. Thus, an additional row was added to Figure 1 that substantiates on the ethical considerations that are specific to the Data Collection & Preprocessing, Data Processing & Model Execution, and AI-Generated Results & Actions of the artificial intelligence IPO approach.

integrates technical knowledge with ethical, social and human considerations.

AI educational policy was explored in the second section, and study **outlined the role** of institutions such as UNESCO in setting global standards, benchmarks and recommendations for responsible AI use in education. The analysis of UNESCO's "Beijing Consensus on Artificial Intelligence in Education" and the "Recommendation on the Ethics of Artificial Intelligence," among other UNESCO guideline documents helped to highlight key policy considerations for AI governance in education. The interdisciplinary cooperation between governments, educational institutions, and technology developers was also emphasised, given that many national AI policies do not have specific ethical guidelines for AI in education.

Section three on the **ethical considerations of AI** provided an expansive exploration of key ethical challenges in education, such as data privacy, bias in algorithms, digital divide, transparency, accountability, and the potential for AI to disrupt traditional educational methodologies, pedagogy, and approaches. The importance of ethical compliance at every stage of AI development and usage was engaged, ranging from data collection to decision-making, and implementation. The discussion also highlighted the more prominent philosophical ethical theories (consequentialism, deontology, and virtue ethics) that are usually applied in the domain of AI ethics and policymaking. Moreover, the intersection of the autonomy and agency of learners in the context of AI was explored, to emphasize that AI should primarily support human decision-making, and not replace it.

The suggested AI **IPO Ethical Framework** was structured and illustrated (Figure 2) as a model for embedding ethics in AI systems at each stage of their development and usage. This framework ensures that ethical considerations are not merely an afterthought but are integrated into the design, execution, and implementation of AI systems. The **input** stage offers ethical oversight in data collection, for ensuring non-bias, and compliance with data privacy laws. The second step, which is the **process** stage focuses on algorithmic transparency, interpretability, and explainability, ensuring that AI systems operate fairly and in an accountable manner. The third and final **output** stage evaluates the societal impacts of AI-generated results, by addressing misinformation, intellectual property rights, and the consequences of decision-making. This framework recommends a proactive approach to ethical AI governance, ensuring that AI technologies align with fundamental human values, at each stage of IPO.

Recommendation

This paper shows that there is general consensus that AI has already and will continue to revolutionize education by enhancing learning experiences, improving administrative processes, and expanding access to knowledge. However, these benefits must be balanced with strong ethical governance to prevent abuse of a technology that offers immense utility. The ethical and social challenges presented by AI in education require continuous engagement from all stakeholders to ensure responsible AI use. Even when multinational organisations such as UNESCO provide valuable policy recommendations, it is still the responsibility of national governments and educational institutions to take a more proactive role in developing ethical AI frameworks that are tailored to their specific contexts. Moving forward, it is essential that researchers, policymakers,

educators, and industry leaders continue to engage in dialogue and action to ensure that AI is integrated into education in a manner that upholds human dignity, agency, and equity.

Declarations

Ethics Approval
Not applicable

Consent for Participation and Publication
Not applicable

Availability of data and materials
Not applicable

Competing interests
None

Authors' contributions
Authors contributed to the content and quality as follows: Mfanelo Ntsobi contributed 50%, Bongani June Mwale contributed 20% in alignment, editing, and quality, while Kholekile Ntsobi contributed 30% to the context and quality.

Clinical Trial Registration
Not applicable

References

1. Al-Zahrani, AM., & Alasmari, TM. 2024. Exploring the impact of artificial intelligence on higher education: The dynamics of ethical, social, and educational implications. Humanities and Social Sciences Communications, 11:912. https://doi.org/10.1057/s41599-024-03432-4
2. Al-Zahrani, AM. 2024b. Unveiling the shadows: Beyond the hype of AI in education. Heliyon 10(9): e30696. https://doi.org/10.1016/j.heliyon.2024.e30696
3. André Renz, GV. 2021. Reinvigorating the discourse on human-centered artificial intelligence in educational technologies. Technology Innovation Management Review, 11(5):5–16. https://doi.org/10.22215/timreview/1438
4. Barocas, S., Hardt, M., & Narayanan, A. 2019. Fairness and machine learning: Limitations and opportunities. Available at: https://fairmlbook.org/ (Last accessed 28 February 2025).
5. Bates, T., Cobo, C., Mariño, O., & Wheeler, S. 2020. Can artificial intelligence transform higher education? International Journal of Educational Technology in Higher Education 17(1):42. https://doi.org/10.1186/s41239-020-00218-x
6. Buitem, MC. 2019. Towards intelligent regulation of artificial intelligence. European Journal of Risk Regulation, 10(1), 41–59.
7. Butler-Adam, J. (2018). The Fourth Industrial Revolution and education. South African Journal of Science, 114(5/6). https://doi.org/10.17159/sajs.2018/a0271
8. Chaudhry, MA., Cukurova, M., & Luckin, R. 2022. A transparency index framework for AI in education. In: International Conference on Artificial Intelligence in Education. Springer: 195–198. https://doi.org/10.35542/osf.io/bstcf.

9. Chen, C. 2023. AI will transform teaching and learning. Let's Get it Right. Human-Centered Artificial Intelligence. https://hai.stanford.edu/news/ai-will-transform-teaching-and-learning-lets-get-it-right

10. Coeckelbergh, M. (2020). AI Ethics. MIT Press.

11. Costa K., et al. 2024. "Challenges, Benefits and Recommendations for Using Generative Artificial Intelligence in Academic Writing - A Case of ChatGPT". Medicon Engineering Themes 7(4): 03-38.

12. Cukurova, M., & Miao, F. 2024. AI competency framework for teachers. UNESCO. https://doi.org/10.54675/ZJTE2084

13. Dabbagh, H., Earp, BD., Porsdam, S., et al. 2025. AI ethics should be mandatory for schoolchildren. AI and Ethics, 5:87–92. https://doi.org/10.1007/s43681-024-00462-1

14. Dabis, A., & Csáki, C. 2024. AI and ethics: Investigating the first policy responses of higher education institutions to the challenge of generative. AI Humanities and Social Sciences Communications, 11:1006. https://doi.org/10.1057/s41599-024-03526-z

15. Dignum, V. 2017. Responsible artificial intelligence: Designing AI for human values. ITU Journal: ICT Discoveries, 1. https://dspace.daffodilvarsity.edu.bd:8080/handle/123456789/2181

16. Doncieux, S., Chatila, R., Straube, S., & Kirchner, F. 2022. Human-centered AI and robotics. AI Perspectives, 4(1):1. https://doi.org/10.1186/s42467-021-00014-x

17. European Commission. 2019. Ethical guidelines for trustworthy AI. Brussels, Belgium: European Commission, High-Level Expert Group on Artificial Intelligence (AI HLEG). https://ec.europa.eu/newsroom/dae/document.cfm?docid=58477

18. Figaredo, D., & Stoyanovich, J. 2023. Responsible AI literacy: A stakeholder-first approach. Big Data & Society, 10(2). https://doi.org/10.1177/20539517231219958

19. Giannini, S. 2023. Education in the age of artificial intelligence (Editorial). The UNESCO Courier October-December 2023:3. Available online: https://courier.unesco.org/en/articles/education-age-artificial-intelligence

20. Hasan, AH., Hilmi, MF. 2022. Input process output (IPO) AI chatbot as personal learning assistant for programming coursework. Malaysian Journal of Distance Education, 22(1). https://ejournal.usm.my/mjde/article/view/4038

21. Heilinger, JC. 2022. The ethics of AI ethics. A constructive critique. Philosophy and Technology, 35(3):1-20. https://doi.org/10.1007/s13347-022-00557-9

22. Holmes, W., Bektik, D., Whitelock, D., & Woolf, B.P. 2018. Ethics in AIED: Who cares? In Artificial Intelligence in Education: 19th International Conference, AIED 2018, London, UK, June 27–30, 2018, Proceedings, Part II. https://aied2018.utscic.edu.au/workshops-and-tutorials/

23. Holmes, W., Bektik, D., Di Gennaro, M., Woolf, B. P., & Luckin, R. (2019a). Ethics in AIED: Who cares? In S. Isotani, E. Millán, A. Ogan, P. Hastings, & R. Luckin (Eds.), Artificial Intelligence in Education (Vol. 11625, pp. 424–425). Cham, Switzerland: Springer International Publishing AG. https://doi.org/10.1007/978-3-030-23204-7

24. Jobin, A., Ienca, M., & Vayena, E. 2019. The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

25. Khan, S., Mazhar, T., Shahzad, T., et al. 2025. Harnessing AI for sustainable higher education: Ethical considerations, operational efficiency, and future directions. Discover Sustainability, 6(23). https://doi.org/10.1007/s43621-025-00809-6

26. Khosravi, H., Shum, S. B., Chen, G., et al. 2022. Explainable Artificial Intelligence in education. Computers and Education: Artificial Intelligence, 3, Article 100074. https://doi.org/10.1016/j.caeai.2022.100074

27. Khvatova, T., Appio, FP., Ray, S., & Schiavone, F. 2024. Exploring the role of AI in B2B customer journey management: Towards an IPO model. IEEE Transactions On Engineering Management, 71:13852-13866. https://doi.org/10.1109/TEM.2023.3284532

28. Manel, HA., Sania, W., Fadhillah, N., Mahmud, A. 2023. Implementasi Artificial Intelligence dalam Sistem Informasi Akuntansi dan Manajemen. Jurnal Akuntansi Bisnis Dan Ekonomi, 9(2): 3460-3467. https://journal.widyatama.ac.id/index.php/jabe/article/view/1181

29. McNulty, N. (2024). AI in South African Education and Parental Involvement. Medium. Available from: AI in South African Education and Parental Involvement | by Niall McNulty | Medium

30. Niemi, H. (2021). AI in learning: Preparing grounds for future learning. Journal of Pacific Rim Psychology, 15. Available from: https://doi.org/10.1177/18344909211038105. (Last accessed 28 February 2025)

31. Ntsobi, MP., Costa, AP., Kasperiuniene, J., Brandão, C., Ribeiro, J. 2024. "Digital Tools and Techniques in Qualitative Research: Digital Skills and Research Optimisation." In: Ribeiro, J., Brandão, C., Ntsobi, M., Kasperiuniene, J., Costa, A.P. (eds) Computer Supported Qualitative Research. WCQR 2024. Lecture Notes in Networks and Systems, Vol 1061. Springer, Cham.

32. Schiff, D., Ayesh, A., Musikanski, L., & Havens, J. C. 2020a. IEEE 7010: A new standard for assessing the well-being implications of artificial intelligence. In 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2746–53. https://doi.org/10.1109/SMC42975.2020.9283454.

33. Schiff, D., Biddle, J., Borenstein, J., & Laas, K. 2020b. What's next for AI ethics, policy, and governance? A global overview. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 153 58. AIES '20. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3375627.3375804

34. Schiff, D., Borenstein, J., Laas, K., & Biddle, J. 2021. AI ethics in the public, private, and NGO sectors: A review of a global document collection. IEEE Transactions on Technology and Society, 2, 31–42. https://doi.org/10.1109/TTS.2021.3052127.

35. Schiff, D. 2022. Education for AI, not AI for education: The role of education and ethics in national AI policy strategies. International Journal of Artificial Intelligence in Education 32(3):527–563. https://doi.org/10.1007/s40593-021-00270-2

36. Seo, K., Tang, J., Roll, I., et al. 2021. The impact of artificial intelligence on learner–instructor interaction in online learning. International Journal of Educational Technology in Higher Education 18(1):54. https://doi.org/10. 1186/s41239-021-00292-9

37. Khvatova, T., Appio, FP., Subhasis Ray, S., & Schiavone, F. 2024. Exploring the Role of AI in B2B Customer Journey Management: Towards an IPO Model. IEEE Transactions, 71

38. Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2022). Ethical principles for artificial intelligence in education. Education and Information Technologies, 28(28). https://link.springer.com/article/10.1007/s10639-022-11316-w.

39. Trocin, C., Hovland, IV., Mikalef, P., & Dremel, C. 2021. How artificial intelligence affords digital innovation: A cross-case analysis of Scandinavian companies. Technological Forecasting & Social Change, 173. https://doi.org/10.1016/j.techfore.2021.121081

40. UNESCO. 2019. Beijing Consensus on Artificial Intelligence and Education. International Conference on Artificial Intelligence and Education 'Planning education in the AI era: Lead the leap' 16 – 18 May 2019 Beijing, People's Republic of China Available from: https://unesdoc.unesco.org/ark:/48223/pf0000368303.

41. UNESCO. 2021a. Draft text of the recommendation on the ethics of artificial intelligence. Available From: https://unesdoc.unesco.org/ark:/48223/pf0000377897.

42. UNESCO. 2021b. Artificial intelligence and education. Guidance for policymakers. The United Nations Educational, Scientific and Cultural Organization, 1-50. https://doi.org/10.54675/PCSP7350

43. UNESCO. 2024. Draft AI competency frameworks for teachers and for school students. Available from: https://www.unesco.org/sites/default/files/medias/fichiers/2024/04/UNESCO-Draft-AI-competency-frameworks-for-teachers-and-school-students.pdf.

44. Vallor, S. 2016. Technology and the virtues: A philosophical guide to a future worth wanting. Oxford University Press.

45. Yan, Y., Liu, H. 2024. Ethical framework for AI education based on large language models. Education and Information Technologies. https://doi.org/10.1007/s10639-024-13241-6.