Kagisano

Number 15

Artificial Intelligence and Higher Education in South Africa

2024

CHE
COUNCIL ON HIGHER EDUCATION

Published by the Council on Higher Education (CHE)

The CHE is an independent statutory body established in terms of the provisions of the Higher Education Act No. 101 of 1997, as amended. It advises the Minister responsible for higher education and training, and is the national authority for external quality assurance and promotion in higher education. In terms of the National Qualifications Framework Act No. 67 of 2008, as amended, the CHE is the Quality Council for higher education responsible for, among others, the development, further development and management of the Higher Education Qualifications Sub-Framework (HEQSF).

Disclaimer

Views expressed in this publication are those of the authors in their individual capacities. They are not necessarily the views of the CHE.

Peer review

Each article in this publication has been critically read by two external peer reviewers, both of whom made recommendations for improvement to the authors. An editorial panel provided editorial oversight over the quality of the manuscripts accepted for publication.

ISSN: 978-1-0672559-2-3

Publication date: August 2024

Contact details

Council on Higher Education

1 Quintin Brand Street

Persequor Technopark

Brummeria

Pretoria

South Africa

Tel: 012 349 3850

Email: research@che.ac.za

www.che.ac.za

COUNCIL ON HIGHER EDUCATION

Kagisano

Number 15

Artificial Intelligence and Higher Education in South Africa

2024

Foreword

Artificial intelligence (AI) is defined as the ability and development of information technology-based computer systems or other machines to complete the tasks that would ordinarily require human intelligence and logical deduction.[1] AI is steadily permeating all spheres of life and sectors of human endeavours. From the Global Positioning System (GPS), which has become a common device to many people, to translation and voice recognition applications, as well as the applications that the banking sector employs to secure bank accounts and monitor transaction requests, the potential for AI seems limitless. Higher education is not immune to the rapidly increasing influence of AI. More universities are adopting the use of algorithms for marketing to prospective students, estimating class sizes, planning curricula, and allocating resources such as financial aid and facilities.[2] Another area of higher education that has seen remarkable growth in the use of AI is student support. Universities are utilising machine learning to assist students to automatically schedule their course load. Other AI tools recommend to students courses, majors, and career paths.

In the core area of teaching and learning, virtual AI teaching assistants are able to answer frequently asked questions without the help of human beings; intelligent tutoring systems are able to use cognitive science and AI technologies to provide personalised tutoring in real time; and Smart Education uses AI technology to make education more effective, efficient, flexible and comfortable; to mention a few examples. [3] It is believed that these developments have the potential to make learning more inclusive and accessible for those with language barriers or disabilities because they would be able to use AI to access learning in a format that is personalised and formatted according to their individual requirements – for example, the same content could be provided as audio, video or text, all communicated at the right level.[4] Even the area of research and publication has not been spared the influence of AI as manuscript writing support services using AI technology have become increasingly available in recent years,[5] the latest being the ChatGPT. A recent study showed that ChatGPT could be used to write a finance paper that could be accepted for publication in a reputable academic journal.[6]

Notwithstanding the ever-increasing adoption and utilisation of AI in higher education, there are some questions about AI that still need to be answered. A fundamental question is on whether AI-based academic writing support undermines the originality and the contribution of researchers. Similarly, there are questions of academic and/or research integrity. These questions point to the need for

[1] Ma, Y., and Siau, K. L. 2018. Artificial Intelligence Impacts on Higher Education. *MWAIS 2018 Proceedings*. 42. http://aisel.aisnet.org/mwais2018/42

[2] Zeide, E. 2019. Artificial Intelligence in Higher Education: Applications, Promise and Perils, and Ethical Questions. *Educause Review*. https://er.educause.edu/articles/2019/8/artificial-intelligence-in-higher-education-applications-promise-and-perils-and-ethical-questions

[3] Jain, S. and Jain, R. 2019. The role of artificial intelligence in higher education – an empirical investigation. *International Journal of Research and Analytical Reviews* 6(2): 144z – 150z.

[4] Webb, M. 2022. What is next for AI in higher education? *Times Higher Education*. https://www.timeshighereducation.com/campus/whats-next-ai-higher-education

[5] Nakazawa, E., Udagawa, M. and Akabayashi, A. 2022. Does the use of AI to create academic papers undermine research originality? AI 2022, 3, 702–706. https://doi.org/10.3390/ai3030040

[6] Lucey, B. and Dowling, M. 2023. ChatGPT: Our study shows AI can produce academic papers good enough for journals – just as some ban it. *The Conversation*. https://theconversation.com/chatgpt-our-study-shows-ai-can-produce-academic-papers-good-enough-for-journals-just-as-some-ban-it-197762

conversations among all concerned parties, including academics, researchers, research funding agencies, students and publishers of scholarly journals, among others.

It is this realisation that there is a need for conversations on AI in higher education which prompted the Council on Higher Education (CHE) to convene a research colloquium to provide a national platform for a discourse on developments in AI, uptake of AI in higher education, benefits of AI to higher education with specific focus to South Africa, the threats and risks that AI poses to higher education, and possible strategies that could be put in place to mitigate the threats and risks. This was in line with one of CHE's core functions, which is to provide platforms for intellectual engagement on matters of interest to the higher education sector. The colloquium took place in May 2023.

One recommendation from the colloquium was that the presentations that formed the basis for discussion should be developed into full papers for publication in the internal research journal of the CHE – *Kagisano*. The presenters at the colloquium subsequently undertook further research to enrich their material and develop the presentations into full academic papers that are published in this volume of *Kagisano*. However, not all of the papers herein were developed from presentations that were delivered and discussed at the colloquium. Some were prepared in response to a call for manuscripts that was issued to the higher education sector to afford opportunity to some academics, researchers and students who could not be at the colloquium, but who had good ideas to share with the rest of the higher education community on the subject of AI and higher education, to do so by writing articles to be considered for publication in *Kagisano*. The result is this volume of *Kagisano* that carries articles that cover various perspectives about AI and higher education.

A common thread that runs through all articles in this volume is that AI, similar to any other technological developments, has been developed and continues to be developed to help human beings to be resource efficient while being more productive. This is true in any sphere of life - be it in agriculture, banking, healthcare, manufacturing, or transport, to name a few - and the same is true in higher education. The papers document how AI is being used to enhance teaching and learning, to improve the quality of student support services, to increase the efficiency levels in the provision of administrative support services, to increase the productivity of researchers, and generally, to make higher education activities and programmes more manageable and less onerous. There is no gainsaying the fact that higher education has much to gain from adopting and utilising AI tools.

With no exception, all papers published in this volume of *Kagisano*, contend that as a human tool, the value of AI to human beings depends on how it is used. When a tool of any kind is used whimsically without following rules or protocols of its proper use, it could cause harm to the user and other human beings. As a tool, AI is no exception to this fact of life. When it is employed in any human activity without following rules or protocols, it may pose risks to human beings and compromise the values that human beings hold dear. The papers therefore submit that it is important to be vigilant when using AI; and they call for the authorities in higher education institutions, specifically, to develop and implement policy frameworks that articulate instructions, guidelines and protocols for the judicious use of AI by academics, researchers, students, professional support staff, and the higher education community at large. Several papers point out that a few higher education institutions in South Africa are at the early stages of developing such policy frameworks, whereas the majority are yet to start off on this journey, which leaves academics, researchers, students, professional support staff to utilise AI as they wish, with the attendant risks of compromising the integrity of teaching and learning, research, and engaged scholarship. Three papers in this volume observe that most higher education institutions await a national policy on the adoption and utilisation of AI in higher education to be developed and promulgated, so that they can take the cue from it in the process of developing institutional policy frameworks and guidelines. This makes is critically essential that such a national policy be developed, finalised and gazetted sooner rather than later.

Academics, researchers, students, professional support in higher education, and other higher education stakeholders are encouraged to read the interesting papers in this volume of *Kagisano*. As they read the papers, if they wish to provide feedback, they should feel free to send such feedback by email to research@che.ac.za.

The CHE sincerely appreciates the efforts of the authors, peer reviewers, and members of the editorial panel for bringing the project of producing this volume to fruition. The coordination role played by the staff in the Research, Monitoring and Advice (RMA) Directorate at the CHE, in this project, is also acknowledged with much gratitude.

Dr Whitfield J Green
Chief Executive Officer

August 2024

Table of Contents

Artificial Intelligence as the Latest Tool of *Homo habilis*…. Will it be the Last Tool?

D. Mashao and C. Aigbavboa
University of Johannesburg
dmashao@uj.ac.za

Abstract

The evolution of *Homo habilis* has been intricately linked with the development and use of tools, marking a significant milestone in progress. Using tools is considered a defining characteristic of human beings, and it has played a crucial role in their ability to adapt to, survive, and thrive in various environments. The modern day has witnessed the human creation of more sophisticated tools, such as Artificial Intelligence (AI), which has become increasingly prevalent across various domains, revolutionising industries and impacting daily life in the era of the Fourth Industrial Revolution. Based on literature, this paper explores the symbiotic relationship between humanity and its tools, from the rudimentary implements of *Homo habilis* to the sophisticated intricacies of AI. This symbiosis prompts a reflection on the responsible stewardship of AI, recognising it not just as a tool but as a reflection of humankind's values, intentions, and ethical considerations. Furthermore, the capacity to navigate the AI age hinges on a commitment to continuous learning, adaptation, and keen awareness of the ethical dimensions of technological advancement. Finally, the question "Will AI be the last tool?" prompts introspection and action. It invites critical thinkers to shape a future where the symbiosis between human beings and technology transcends the confines of mere tools, ushering in an era where innovation, ethical considerations, and the heart of *Homo habilis* converge to define the next epoch of human evolution.

Keywords: Artificial intelligence, human-tool symbiosis, ethical considerations, continuous learning and adaptation, evolutionary progress

Introduction

The earliest known member of the genus to which modern human beings belong is thought to have been *Homo habilis* (Leakey *et al*., 1964). The *Homo habilis* was characterised by its more human-like teeth and hands, which showed more excellent dexterity and capability for the use of tools (Leakey *et al*., 1971). The stone tools associated with *Homo habilis*, known as the Oldowan industry, are among the earliest known in the archaeological record. They consisted primarily of simple, unifacially flaked cores and sharp-edged flakes. These tools were likely used for hunting animals and processing plant materials (Leakey *et al*., 1971). Thus, using tools is considered a defining characteristic of human beings, and it has played a crucial role in the ability to adapt to, survive, and thrive in various environments. Subsequent advancements led to the Acheulean tool industry around 1.7 million years ago, characterised by hand axes and more refined bifacial tools. This technological leap marked an important stage in the cognitive development of early *Homo* species (Morgan *et al*., 2015).

Further innovations continued with the Middle Stone Age (around 300,000 to 30,000 years ago) and the Upper Palaeolithic (around 40,000 to 10,000 years ago), which witnessed the creation of even more sophisticated tools, including blades, bone implements, and composite tools. These innovations facilitated hunting, gathering, and eventually agriculture, enabling Homo sapiens to expand across the globe (Klein, 2009). The use of these tools not only allowed for better resource exploitation but also influenced the development of humankind's cognitive abilities, such as problem-solving, planning, and spatial reasoning (Wynn and Coolidge, 2016). This cognitive enhancement, in turn, has played a significant role in shaping the evolutionary trajectory of human beings.

Furthermore, the modern day has witnessed human creation of more sophisticated tools, such as Artificial Intelligence (AI). The use of AI tools by humans being has become increasingly prevalent across various domains, revolutionising industries and impacting daily life in the era of the Fourth Industrial Revolution. AI refers to developing computer systems that can perform tasks that require human intelligence. These tools rely on algorithms, data processing, and machine learning to analyse vast information and make predictions or decisions. According to Da Xu *et al*. (2021), AI teaches computers to mimic human intelligence and do tasks like learning, judgment, and decision-making. Duan and Xu (2012) opined that it uses knowledge as its object, learns about knowledge, examines and investigates knowledge representation techniques, and applies these techniques to simulate human intellectual activity. Hence, AI has been fundamental to societal progress and has produced ground-breaking improvements in labour productivity, labour cost reduction, human resource structure optimisation, and the creation of new employment needs.

AI is employed in healthcare for the interpretation of medical imaging, personalised treatment plans, and drug discovery. For example, the use of AI in radiology has shown significant advancements in detecting and diagnosing diseases (Esteva *et al*., 2017). In finance, AI is used for fraud detection, risk assessment, and algorithmic trading. Credit scoring models incorporating AI techniques have improved accuracy in evaluating creditworthiness (Hand & Henley, 1997) of individuals and organisations. In marketing and e-commerce, AI-driven recommendation systems enhance the experience of users by providing personalised product suggestions based on past behaviour and preferences. Companies like Amazon and Netflix utilise AI algorithms to boost customer engagement (Koren *et al*., 2009). Also, AI is fundamental for object recognition, decision-making, and navigation tasks in autonomous vehicles. Self-driving cars rely on AI systems to interpret sensor data and to make real-time driving decisions (LeCun *et al*., 2015). These tools that humankind has created have always extended human ability in some ways. They mimic human body parts and do things the human body could not do. Considering human beings are an intelligent species, this literature-based paper seeks to answer the question on whether there will be any more tools that humankind will need after the perfection of AI.

Methodology

This paper focuses on AI as the latest tool of the *Homo habilis* and its potential to be the last tool. It is based on the review of relevant literature sourced from Scopus database. Olawumi *et al*. (2017) noted that the Scopus database is regarded as one of the most important databases for scientific research. In terms of abstracts and citations, it is also the largest (Nobre and Tavares, 2017; Olawumi *et al*., 2017; Guz and Rushchitsky, 2009). The extensive coverage, record overlap with Web of Science, and regular review paper usage make Scopus a reputable resource (Vieira and Gomes, 2009; Chadegani *et al*., 2013; Olawumi *et al*., 2017). It also provides wider coverage than other databases (Hosseini *et al*., 2018). The search for pertinent published works concentrated on journal articles, book chapters, and conference proceedings. 'Homo habilis' and 'tools' were the terms used for the search. The search result produced a total of 71 articles, which had their abstracts examined. After reading, 71 papers were adopted for use in drafting this paper based on their contents. Figure 1 presents the framework adopted for the study.

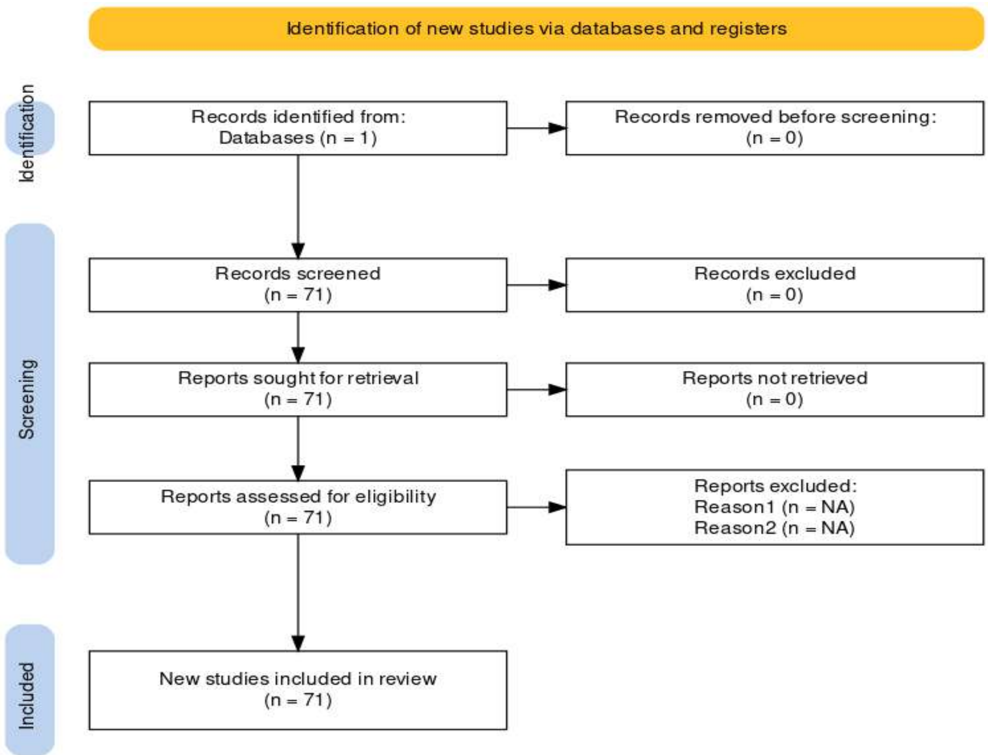

Figure 1: Methodology Framework (Haddaway *et al*. (2022)

The evolution of tools in human history

Tools have played a pivotal role in the evolution of the human species, enabling survival, adaptation, and cultural development. The emergence of the use of tools is a defining characteristic of early hominins, with *Homo habilis* considered one of the earliest tool-makers. Their stone tools, known as

Oldowan tools, date back approximately 2.6 million years (Toth and Schick, 2009). These rudimentary implements, consisting of chipped stones with sharp edges, were likely used for basic tasks such as hunting, cutting, and scraping. The Acheulean toolkit associated with *Homo erectus* and dating from around 1.7 million years ago, marked a significant advancement in tool-making. Acheulean tools were characterised by hand axes and cleavers, showing more refinement and versatility than the Oldowan tools (Wynn, 2009). These tools facilitated more complex activities, including hunting, woodworking, and processing of various materials. The transition to anatomically modern *Homo sapiens* approximately 300,000 years ago brought about further advancements in tool technology. This period saw the development of sophisticated tools made from a broader range of materials, including bone, antler, and, later, metals (Dibble, 1987). Notably, around 40,000 to 10,000 years ago, the Upper Palaeolithic period witnessed the creation of finely crafted tools such as blades, spears, and harpoons, indicating increased specialisation and skill (Shea, 2013). Approximately 10,000 years ago, the Neolithic Revolution marked a pivotal shift in tool use as human beings transitioned from nomadic hunting and gathering to settled agriculture. This period saw the emergence of grindstone tools, such as grinding stones and ploughs, which were crucial for agriculture and early civilisations (Bar-Yosef, 2002). The subsequent eras brought about further technological revolutions, including the Bronze Age and Iron Age, which introduced metal tools and weapons, revolutionising agriculture, warfare, and craftsmanship (Harding, 2000).

It is therefore evident that the early tools were developed to solve specific problems that human beings faced. This also marked the journey towards AI, which can be seen as a continuous effort to develop tools that emulate the cognitive abilities of human beings. Furthermore, the First Industrial Revolution (18^{th}-19^{th} century) dramatically transformed tool-making with the advent of machinery and mass production techniques (Joel, 2016). This period reshaped various industries and significantly impacted human societies. The introduction of electrical energy in the 1870s marked the beginning of the Second Industrial Revolution and the creation of the key mass production system. These revolutions hinged on how much more human capability could be accomplished (Alaloul *et al*., 2020). Also, development in electronics in the 1970s brought about the Third Industrial Revolution. The term "Digital Revolution" describes how technology has advanced from analogue electrical and mechanical devices to the current state of digital technology. Today, in the Information Age (Fourth Industrial Revolution), tools have evolved beyond physical implements. Digital technologies, encompassing software, AI, and automation, have become integral to nearly every facet of human life, revolutionising communication, industry, and research (Brynjolfsson and McAfee, 2014).

The foregoing historical sketch highlights how tools extended the capabilities of early human beings. This is similar in many ways to how AI extends human cognitive abilities.

Artificial intelligence: the next epoch in tool-making

AI stands at the forefront of the next epoch in tool-making, representing a paradigm shift in how human beings interact with and harness technology. Unlike previous tools, which rely on human guidance, AI can learn, adapt, and make autonomous decisions based on vast datasets. This transformative capability has the potential to revolutionise numerous industries and redefine the boundaries of human achievement (Russell and Norvig, 2016). AI systems excel in tasks that traditionally require human cognitive abilities, such as pattern recognition, natural language processing, and complex decision-making. Machine learning algorithms, a subset of AI, allow systems to improve their performance over time through exposure to data, a feature akin to learning in humans (Goodfellow *et al*., 2016). This adaptability makes AI a powerful tool in contexts ranging from healthcare diagnostics to autonomous vehicles. Some of the AI tools are discussed in the ensuing subsections.

Robotics

The term ‘robotics’ refers to the interdisciplinary field of science and engineering that focuses on designing, developing, and applying autonomous machines capable of performing tasks with varying degrees of complexity (Khatib and Siciliano, 2016). These machines, known as robots, can range from simple manipulators to highly sophisticated humanoid or specialised robots, each tailored for specific functions and environments. One of the critical advancements in robotics is the development of autonomous navigation systems. These systems enable robots to perceive their surroundings and make decisions based on sensory inputs. This capability is fundamental for robots to operate effectively in dynamic and unstructured environments. Simultaneous Localization and Mapping (SLAM) techniques, for example, allow robots to create maps of their surroundings while simultaneously determining their position (Thrun *et al*., 2005).

Furthermore, advancements in robotic hardware have significantly expanded the capabilities of robots. Innovations in materials, actuators, and sensors have created more dexterous and adaptable robots. Soft robotics, for instance, employs flexible materials and pneumatic actuators to create robots capable of tasks in delicate or constrained environments (Trivedi *et al*., 2008). Robots have found applications across various industries. In manufacturing, they are employed for dangerous, monotonous tasks or require precision beyond human capabilities. For instance, in automotive assembly lines, robots perform tasks such as welding, painting, and assembling components (Nof, 2015). In healthcare, robots assist in surgery rehabilitation and even provide companionship for the elderly (Barrett *et al*., 2019).

Furthermore, robots play a crucial role in exploration and disaster response. Unmanned Aerial Vehicles (UAVs) and Unmanned Ground Vehicles (UGVs) are deployed in hazardous or inaccessible environments. For instance, drones are used for aerial surveillance, search and rescue operations, and environmental monitoring (Zhou *et al*., 2019). In summary, robotics represents a significant evolution in tool-making, transitioning from passive instruments to active agents with autonomy, adaptability, and learning capabilities. While offering numerous benefits in efficiency, productivity, and safety, robotics also presents challenges such as job displacement, ethical considerations, and economic disparities.

Chatbot

Chatbots, or conversational agents or virtual assistants, are computer programmes designed to engage in natural conversations with users through text or speech interfaces (Jurafsky and Martin, 2020). They employ a range of technologies, including Natural Language Processing (NLP), Machine Learning (ML), and sometimes AI, to understand and respond to user queries or commands. One of the critical features of chatbots is their ability to process and understand human language. NLP techniques enable chatbots to analyse and interpret the meaning of user inputs, allowing them to provide relevant and contextually appropriate responses. This is achieved through sentiment analysis, entity recognition, and language modelling (Zhou and Zhang, 2003).

In addition, chatbots can be categorised based on their underlying architecture. Rule-based chatbots operate on predefined rules and patterns, making them suitable for specific, structured tasks. On the other hand, AI-powered chatbots leverage machine learning algorithms and natural language understanding to handle a broader range of inquiries (Di Prospero *et al*., 2017). The applications of chatbots are diverse and widespread. In customer service, they provide instant responses to frequently asked questions, handle routine queries, and assist with basic troubleshooting (Rapp *et al*., 2021). E-commerce platforms often integrate chatbots to guide users through purchasing, recommend products, and provide personalised shopping experiences (Hsu and Lin, 2023).

Therefore, chatbots represent a significant evolution in tool-making, transitioning from physical, manually operated tools to virtual, autonomous agents. While previous tools often required creating new tools for different functions or tasks, chatbots can be updated and reprogrammed to handle new tasks or respond to changing user needs. Hence, integrating robotics and AI in chatbots has enabled sophisticated interactions, learning capabilities, and scalability. However, this evolution also brings challenges like those mentioned in the robotics section.

Machine Learning

Machine learning (ML) is an AI tool that develops algorithms and models that enable computers to learn and make predictions or decisions based on data (Mitchell, 1997). It involves using statistical techniques and computational methods to improve the performance of machines on a specific task over time. One of the fundamental concepts in machine learning is supervised learning, where models are trained on labelled data, meaning each data point is paired with the correct outcome. This allows the model to learn the mapping between input features and target values, making accurate predictions on new, unseen data (Hastie *et al*., 2009). Furthermore, unsupervised learning involves training models on unlabelled data, where the goal is to discover patterns or structures within the data without explicit guidance. Clustering, dimensionality reduction, and generative modelling are standard techniques in unsupervised learning (Bishop, 2006).

Furthermore, reinforcement learning is a branch of machine learning where agents learn to make sequential decisions in an environment. The agent receives feedback through rewards or penalties

based on its actions, allowing it to learn optimal strategies over time (Sutton and Barto, 1998). Deep learning, a subset of machine learning, involves using neural networks with multiple layers to extract high-level features from data. Deep learning has proven highly effective in tasks such as image and speech recognition, natural language processing, and autonomous driving (LeCun *et al*., 2015).
Therefore, machine learning represents a profound evolution in tool-making, as it enables the creation of autonomous, adaptive, and intelligent systems that go beyond traditional tools. Robots powered by machine learning algorithms offer enhanced efficiency, precision, versatility, and learning capabilities, revolutionising industries and opening new possibilities. However, this evolution also raises critical questions about human-robot collaboration, economic impacts, ethical considerations, and the need for robust regulatory frameworks.

The impact of AI on various sectors

Globally, AI is one of the newest developments. AI influences many industries, including business, industry, education, automobile, aviation, surgery, and medicine (Ali and Haider, 2016; Arlitsch and Newell, 2017). Below is an overview of how AI has impacted the healthcare, finance, manufacturing, logistics and agricultural sectors.

Healthcare

AI algorithms excel in interpreting medical images, such as X-rays, CT scans, and magnetic resonance imaging (MRI) images. They aid in detecting abnormalities, including tumours, fractures, and anomalies (Obermeyer and Emanuel, 2016). Deep learning models, in particular, have shown remarkable accuracy in image analysis. It also assists in tailoring treatment plans to individual patient profiles. It analyses patient data, genetic information, and treatment outcomes to recommend the most effective therapies (Esteva *et al*., 2019). It accelerates drug discovery by predicting the potential of chemical compounds for drug development. Also, it identifies novel drug candidates and optimises existing formulations, reducing time and costs (Kraus, 2020).

Furthermore, AI-enabled wearables and remote monitoring devices allow continuous tracking of patient health metrics. These technologies facilitate proactive intervention and reduce hospital readmissions (Haghi *et al*., 2017). Other impacts include predictive analytics, early intervention (Churpek *et al*., 2016), and medical education and training.

Finance

AI-powered algorithms are adept at detecting unusual patterns in financial transactions, helping to identify and prevent fraudulent activities in real time (Geetha and Thilagam, 2021), analysing vast amounts of financial data to make high-frequency trading decisions, and optimising trading strategies and execution (Biais *et al*., 2015). It can also evaluate credit and investment risk and market volatility by providing accurate risk assessments, enabling better-informed investment decisions (Ahmed *et al*., 2022). Furthermore, it automates compliance processes, ensuring adherence to regulatory standards and streamlining reporting requirements (Butler & O’Brien, 2019), assists in anti-money laundering efforts by analysing vast datasets to identify suspicious transactions and behaviours, ensuring compliance with regulatory frameworks (Raffel *et al*., 2018). These advancements can potentially increase efficiency, accuracy, and accessibility in financial operations.

Manufacturing

AI-driven predictive maintenance systems analyse real-time sensor data to anticipate equipment failures, allowing for timely repairs and minimising downtime (Zhang *et al*., 2019), inspect and assess the quality of products on the assembly line, detecting defects with high accuracy (Li *et al*., 2018), optimising production schedules, resource allocation, and inventory management, improving overall efficiency and reducing waste (Zhang *et al*., 2019). It also analyses data across the supply chain to optimise procurement, logistics, and inventory levels, reducing costs and improving delivery times (Wamba *et al*., 2020) while analysing historical data to identify patterns, enabling manufacturers to make data-driven decisions for process improvement and innovation (Wang *et al*., 2016). These advancements are driving increased competitiveness and innovation in the manufacturing industry.

Logistics

AI-powered algorithms analyse real-time data on traffic, weather, and road conditions to optimise delivery routes, reducing transit times and fuel consumption (Giuffrida *et al*., 2022), use historical data and predictive analytics to forecast demand, enabling companies to optimise inventory levels and reduce carrying costs (Wruck *et al*., 2017), enhance warehouse operations, including picking, packing, and sorting tasks, which leads to increased throughput and accuracy (Lun and Zhao, 2015). Furthermore, it optimises the allocation of orders to drivers, considering factors like location, traffic, and delivery time windows, leading to improved customer satisfaction (Chu *et al*., 2021), monitoring vehicle health and performance, predicting maintenance needs and reducing unplanned downtime, which increases the lifespan of assets and reduces maintenance costs (An *et al*., 2011). These advancements have driven significant improvements in supply chain operations.

Agriculture

AI-powered drones and autonomous machinery equipped with sensors and cameras gather data on soil health, moisture levels, and crop health, allowing for precise interventions (Kamilaris *et al*., 2017), analysing satellite imagery and drone-captured data to monitor crop growth, detect diseases, and optimise irrigation and fertilisation schedules, leading to increased yields (Gebbers and Adamchuk, 2010), identify and differentiate between crops, weeds, and pests. This enables targeted treatment, reducing the need for chemical inputs (Pardede *et al*., 2020). Furthermore, AI models utilise historical data, weather forecasts, and soil conditions to predict crop yields, aiding in decision-making for planting, harvesting, and marketing (Pradhan *et al*., 2015), process market data, commodity prices, and trade trends to provide farmers with insights for strategic decision-making on planting and marketing (Wang *et al*., 2018), while AI-driven technology facilitate digital inclusion in rural areas, providing access to market information, e-commerce platforms, and financial services for farmers (Liu, 2021). These advancements have the potential to address global food security challenges.

AI's potential of being the 'last tool'

AI holds the potential to be a transformative tool, often called the 'last tool', due to its capacity to autonomously adapt, innovate, and tackle a wide array of complex tasks. This stems from AI's ability to learn and improve from vast datasets, often in ways that surpass human capabilities. Unlike traditional tools, AI can evolve and adapt its capabilities without physical modification. It learns from new data and experiences, allowing it to improve and refine its performance over time (Russell, 2021). AI's versatility enables it to address various challenges across various sectors, from healthcare and finance to agriculture and transportation. This adaptability positions AI as a tool with broad-spectrum applicability (Russell, 2021). Its autonomous learning and problem-solving capacity can lead to novel solutions and innovations that may not have been conceivable through conventional means. This can potentially drive significant advancements in numerous fields (Bengio *et al*., 2021).

In addition, AI's ability to analyse vast datasets and make informed decisions can lead to more efficient resource allocation. This can result in cost savings, improved productivity, and reduced industry waste (Baldwin *et al*., 2020). Rather than replacing human beings, AI can potentially augment human skills and capabilities. By handling routine or data-intensive tasks, AI frees human professionals to focus on higher-level decision-making and creativity (Brynjolfsson and McAfee, 2014). Furthermore, AI's capacity to process and analyse massive datasets allows it to tackle complex, multifaceted problems that often exceed human cognitive capacity. This is particularly evident in climate modelling and drug discovery (Hasselgren and Oprea, 2023). Thus, AI holds immense potential to be the 'last tool' in various domains, as it is most effective when working in tandem with human expertise and values.

In summary, AI holds immense potential to be the 'last tool', as it can operate independently, making decisions without constant human intervention. Unlike early tools with fixed functions, AI tools can adapt to new situations and data, adjusting their actions and decisions in real time. It also could learn from data and experiences, enabling continuous improvement and refinement of its capabilities over time. Furthermore, rather than relying on separate tools for different tasks, AI can combine functionalities, simplifying workflows and processes. AI tools can be repurposed and retrained for new tasks, eliminating the need to create new physical tools for each function. Hence, AI represents a departure from static early tools towards systems that evolve and improve iteratively through data-driven learning. It represents a monumental leap in tool evolution, potentially culminating in the final step towards autonomous, adaptive, and intelligent tools. The transformative capabilities of AI, from autonomous decision-making to continuous learning and environmental adaptation, redefine how tools are conceived, created, and utilised.

Limitations and ethical considerations

AI holds immense potential, but it is important to acknowledge its limitations and consider the ethical implications of its widespread deployment. An example of the limitations is that AI systems often need more intuitive understanding and common sense reasoning that human beings possess, making them susceptible to misinterpretation of context (Marcus, 2018). According to Mateen (2018), AI models heavily rely on the data they are trained on. Biased or unrepresentative data can lead to biased outcomes, perpetuating existing inequalities. Many AI algorithms, particularly deep learning models, are often viewed as ‘black boxes’ owing to their complex inner workings, making it challenging to understand their decision-making process (Rudin, 2019). Furthermore, Szegedy *et al*. (2014) posited that AI models can be tricked or manipulated by introducing carefully crafted inputs to deceive the system.

For the ethical considerations of AI, Barocas *et al*. (2019) opined that ensuring that AI systems are fair and unbiased is critical. To promote equitable outcomes, steps must be taken to identify and rectify biases in training data and algorithms. AI systems should be designed to explain their decisions, especially in high-stakes applications like healthcare or criminal justice (Lipton, 2018). Also, establishing clear lines of accountability for AI systems is essential. Developers, operators, and organisations should take responsibility for the outcomes of AI applications (Morley et al., 2020). According to Mittelstadt et al. (2016), properly handling personal and sensitive data is crucial. AI systems must adhere to strict privacy regulations to protect individuals' rights and prevent misuse. Finally, maintaining a level of human oversight and control over AI systems is imperative, especially in critical domains where human judgment is crucial. Thus, AI applications should be designed and deployed to benefit humanity while minimising potential harm or negative consequences (Boddington, 2017; Russell, 2021).

Lessons Learnt

Part of the lessons learnt include the ceaseless innovation and refinement of tools that mirrored each era's aspirations and challenges, from the agricultural implements of ancient civilisations to the Industrial Revolution's machinery. They underscore humanity’s relentless drive to enhance productivity and conquer new frontiers. The narrative takes a decisive turn as humankind stands at the precipice of the AI era. Will AI be the last tool, the apotheosis of humankind’s tool-making journey?

Firstly, resilience and adaptability define progress. *Homo habilis* did not cling to the comfort of stone tools; instead, each epoch witnessed a willingness to discard outdated tools for more sophisticated ones. Similarly, in the age of AI, the ability to adapt to rapid technological advancements becomes paramount. For example, the AlphaGo computer programme developed by the London-based DeepMind Technologies showcases AI's potential to excel in tasks that require intuition, strategy, and adaptability, traits typically associated with human intelligence. Even the Tesla Autopilot illustrates AI's autonomy and ability to operate in dynamic environments. Its continuous learning from real-world driving data improves safety and efficiency. Integrating AI into transportation systems represents a shift towards intelligent, self-driving tools that can navigate complex scenarios without human intervention. Also, chatbots and virtual assistants demonstrate AI's role in enhancing user experience, providing personalised interactions, and adapting to user preferences. These examples and case studies of AI applications demonstrate its potential as the final step in tool evolution.

Secondly, the narrative underscores the dual nature of tools, creations and creators in symbiosis. Tools shape human existence, but human intent, creativity, and ethical considerations guide the trajectory of technological evolution.

The lessons call for a mindful approach to AI, recognising it as a tool shaped by human values. Furthermore, the narrative illuminates the need for continuous learning. *Homo habilis'* toolkit expanded with each discovery, and similarly, mankind's capacity to harness AI's potential hinges on ongoing education and exploration. In the rapidly advancing landscape of technology, a profound shift towards human-technology symbiosis is poised to redefine how humankind interacts with its creations. This vision entails a harmonious integration of human intelligence and creativity with the capabilities of advanced technologies, ushering in an era of unprecedented potential and innovation. At its core, human-technology symbiosis envisions a seamless collaboration between human beings and machines, where each complements the strengths and compensates for the limitations of the other. Rather than replacing human roles, technology is a powerful amplifier of human capabilities, unlocking new realms of productivity, creativity, and problem-solving. In the future, intelligent systems will anticipate human needs, enhance decision-making processes, and automate routine tasks. Virtual and augmented reality technologies offer immersive experiences revolutionising education, training, and entertainment. Healthcare is personalised and empowered by AI-driven diagnostics and treatment plans, extending and improving lives. Smart cities harness data and connectivity to optimise infrastructure, resource allocation, and sustainability efforts.

However, ethical considerations and human values remain paramount in this symbiotic future. Transparent algorithms and accountable systems ensure fairness, equity, and justice. Privacy protections are fortified, empowering individuals with control over their data. The symbiotic relationship between humans and technology is grounded in a shared commitment to improving society, the environment, and the human condition. Education and lifelong learning take centre stage, empowering individuals to adapt and thrive in this dynamic landscape. Interdisciplinary collaboration flourishes as experts from diverse fields converge to tackle complex global challenges. While embracing the potential of this vision, society also acknowledges the need for responsible development and deployment of technology. Regulations and governance structures evolved to ensure that ethical considerations guide innovation trajectory. Continuous dialogue and engagement with stakeholders from all walks of life foster a collective sense of ownership and shared responsibility. The lines between the physical and digital worlds will blur in the future, giving rise to boundless possibilities. It is a future where humans are expected to soar, empowered by the transformative potential of technological creations. Finally, "Homo Habilis Latest Tool: Artificial Intelligence, will it be the last tool?" beckons us to grasp the lessons embedded in the tool-making journey. It urges a harmonious coexistence with creations, a commitment to ethical innovation, an openness to adaptation, and a recognition that the last tool, AI, is a mirror reflecting the essence of Homo habilis' earliest endeavours, a testament to the enduring spirit of human ingenuity.

Conclusion

The journey from stone tools to the sophisticated realm of AI encapsulates a story of adaptation, resilience, and continuous learning. As *Homo habilis* discarded obsolete tools for more refined instruments, the narrative underscores the imperative of embracing change and harnessing innovation to navigate the complexities of the ever-evolving world. In contemplating the evolution from *Homo habilis*, the earliest tool-wielding hominid, to the present epoch dominated by AI, the question echoes: Will AI be the last tool? It prompts a reflection on the responsible stewardship of AI, recognising it not just as a tool but as a reflection of humankind's values, intentions, and ethical considerations. The paper cautions against complacency, urging a stance of perpetual curiosity and education. *Homo habilis* expanded their toolkit with each discovery. Likewise, the capacity to navigate the AI era hinges on a commitment to continuous learning, adaptation, and a keen awareness of the ethical dimensions of technological advancement. Finally, the "Will AI be the last tool?" question invites everyone to shape a future where the symbiosis between humans and technology transcends the confines of mere tools, ushering in an era where innovation, ethical considerations, and the heart of *Homo habilis* converge to define the next epoch of human evolution. In this future, AI stands as the apex of tool evolution, representing a new era where tools transcend their traditional limitations. As AI continues to evolve, its implications for society, industries, and the very concept of tools will shape the course of human progress. Thus, responsibly embracing AI development's potential will pave the way for a future where intelligent, adaptive tools empower mankind to tackle the greatest challenges, explore new frontiers, and unlock the full potential of human ingenuity and collaboration. The answer to whether AI will be the last tool lies not just in the capabilities of humankind's creations but in the wisdom, foresight, and compassion with which the tools are wielded.

References

Ahmed, S., Alshater, M. M., El Ammari, A. and Hammami, H. 2022. Artificial intelligence and machine learning in finance: A bibliometric review. *Research in International Business and Finance* 61, 101646.

Alaloul, W. S., Liew, M. S., Zawawi, N. A. W. A. and Kennedy, I. B. 2020. Industrial Revolution 4.0 in the construction industry: Challenges and opportunities for stakeholders. *Ain Shams Engineering Journal* 11(1), 225-230.

Ali, M. Y. and Haider, K. 2016. An analysis of digital reference services tools and usage in university libraries of Karachi. *Pakistan Library and Information Science Journal* 47(1), 20-29.

An, S. H., Lee, B. H. and Shin, D. R. 2011 July). A survey of intelligent transportation systems. In 2011, *Third International Conference on Computational Intelligence, Communication Systems and Networks* (pp. 332-337). IEEE.

Arlitsch, K. and Newell, B. 2017. Thriving in the age of accelerations: A brief look at the societal effects of artificial intelligence and the opportunities for libraries. *Journal of Library Administration* 57(7), 789-798.

Baldwin, D. R., Gustafson, J., Pickup, L., Arteta, C., Novotny, P., Declerck, J., ... and Gleeson, F. V. 2020. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules. *Thorax* 75(4), 306-312.

Barocas, S., Hardt, M. and Narayanan, A. 2019. *Fairness and Machine Learning*. fairmlbook. org.

Barrett, M., Boyne, J., Brandts, J., Brunner-La Rocca, H. P., De Maesschalck, L., De Wit, K., ... and Zippel-Schultz, B. 2019. Artificial intelligence supported patient self-care in chronic heart failure: a paradigm shift from reactive to predictive, preventive and personalised care. *EPMA Journal* 10, 445-464.

Bar-Yosef, O. 2002. The upper Paleolithic revolution. *Annual Review of Anthropology* 31(1), 363-393.

Bengio, Y., Lodi, A. and Prouvost, A. 2021. Machine learning for combinatorial optimisation: a methodological tour d'horizon. *European Journal of Operational Research* 290(2), 405-421.

Biais, B., Foucault, T. and Moinas, S. 2015. Equilibrium fast trading. *Journal of Financial Economics* 116(2), 292-313.

Bishop, C. M. and Nasrabadi, N. M. 2006. *Pattern Recognition and Machine Learning* (Vol. 4, No. 4, p. 738). New York: Springer.

Boddington, P. 2017. *Towards a Code of Ethics for Artificial Intelligence* (pp. 27-37). Cham: Springer.

Brynjolfsson, E. and McAfee, A. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. WW Norton and Company.

Butler, T. and O'Brien, L. 2019. Artificial intelligence for regulatory compliance: Are we there yet? *Journal of Financial Compliance* 3(1), 44-59.

Chu, H., Zhang, W., Bai, P. and Chen, Y. 2021. Data-driven optimisation for last-mile delivery. *Complex and Intelligent Systems* 1-14.

Churpek, M. M., Yuen, T. C., Winslow, C., Meltzer, D. O., Kattan, M. W. and Edelson, D. P. 2016. Multicentre comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Critical Care Medicine* 44(2), 368.

Coolidge, F. L. and Wynn, T. 2016. An introduction to cognitive archaeology. *Current Directions in Psychological Science* 25(6), 386-392.

Da Xu, L., Lu, Y. and Li, L. (2021). Embedding blockchain technology into IoT for security: A survey. *IEEE Internet of Things Journal* 8(13), 10452-10473.

Dibble, H. L. 1987. The interpretation of Middle Palaeolithic scraper morphology. *American Antiquity* 52(1), 109-117.

Di Prospero, A., Norouzi, N., Fokaefs, M. and Litoiu, M. 2017. Chatbots as assistants: an architectural framework. In *Proceedings of the 27th Annual International Conference on Computer Science and Software Engineering* (pp. 76-86).

Duan, L. and Da Xu, L. 2012. Business intelligence for enterprise systems: a survey. *IEEE Transactions on Industrial Informatics* 8(3), 679-687.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M. and Thrun, S. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542(7639), 115-118.

Gebbers, R. and Adamchuk, V. I. 2010. Precision agriculture and food security. *Science* 327(5967), 828-831.

Geetha, R. and Thilagam, T. 2021. A review on the effectiveness of machine learning and deep learning algorithms for cyber security. *Archives of Computational Methods in Engineering* 28, 2861-2879.

Goodfellow, I., Bengio, Y. and Courville, A. 2016. *Deep Learning*. MIT Press.

Giuffrida, N., Fajardo-Calderin, J., Masegosa, A. D., Werner, F., Steudter, M. and Pilla, F. 2022. Optimisation and machine learning applied to last-mile logistics: A review. *Sustainability* 14(9), 5329.

Haghi, M., Thurow, K. and Stoll, R. 2017. Wearable devices in medical internet of things: scientific research and commercially available devices. *Healthcare Informatics Research* 23(1), 4-15.

Hand, D. J. and Henley, W. E. 1997. Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 160(3), 523-541.

Harding, A. F. 2000. *European Societies in the Bronze Age*. Cambridge University Press.

Hasselgren, C. and Oprea, T. I. 2023. Artificial Intelligence for Drug Discovery: Are We There Yet? *Annual Review of Pharmacology and Toxicology* 64.

Hastie, T., Tibshirani, R., Friedman, J. H. and Friedman, J. H. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Vol. 2, pp. 1-758). New York: Springer.

Hsu, C. L. and Lin, J. C. C. 2023. Understanding the user satisfaction and loyalty of customer service chatbots. *Journal of Retailing and Consumer Services* 71, 103211.

Joel, M. 2016. *A Culture of Growth: The Origins of the Modern Economy*. Princeton University Press.

Johanson, D. C. 1989. A partial *Homo habilis* skeleton from Olduvai Gorge, Tanzania: a summary of preliminary results. See Giacobini 1989, pp. 155–66

Johanson, D. C., Masao, F.T., Eck G. G., White. T. D, Walter, R. C., Kimbel, W.H., Asfaw B., Manega, P., Ndessokia, P. and Suwa, G. 1987 New partial skeleton of *Homo habilis* from Olduvai Gorge, Tanzania. *Nature* 327:205–209

Jurafsky, D. and Martin, J. H. 2020. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*.

Kamilaris, A., Gao, F., Prenafeta-Boldu, F. X., and Ali, M. I. 2016. Agri-IoT: A semantic framework for Internet of Things-enabled smart farming applications. In *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)* (pp. 442-447). IEEE.

Khatib, O. and Siciliano, B. (eds) (2016). *Springer Handbook of Robotics*. Springer International Publishing.

Klein, R. G. 2009. *The Human Career: Human Biological and Cultural Origins*. University of Chicago Press.

Koren, Y. 2009. The bellkor solution to the Netflix grand prize. *Netflix prize Documentation* 81(2009), 1-10.

Kraus, J. L. 2020. Can artificial intelligence revolutionize drug discovery? *AI and Society* 35, 501-504.

Leakey, M. D., Hay, R. L., Thurber, D. L., Protsch, R. and Berger, R. 1972. Stratigraphy, archaeology, and age of the Ndutu and Naisiusiu beds, Olduvai Gorge, Tanzania. *World Archaeology* 3(3), 328-341.

Leakey, L. B., Tobias, P. V. and Napier, J. R. 1964. A new species of the genus Homo from the Olduvai Gorge. *Nature* 202, 7–9.

LeCun, Y., Bengio, Y., and Hinton, G. 2015. Deep learning. *Nature* 521(7553), 436-444.

Li, F., Wu, J., Dong, F., Lin, J., Sun, G., Chen, H., and Shen, J. 2018. Ensemble machine learning systems for the estimation of steel quality control. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 2245-2252). IEEE.

Lipton, Z. C. 2018. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* 16(3), 31-57.

Liu, J. 2021. Bridging the digital divide amidst educational change for socially inclusive learning during the COVID-19 pandemic. *Sage Open* 11(4), 21582440211060810.

Lun, R. and Zhao, W. 2015. A survey of applications and human motion recognition with Microsoft Kinect. *International Journal of Pattern Recognition and Artificial Intelligence* 29(05), 1555008.

Marcus, G. 2018. Deep learning: A critical appraisal. arXiv preprint arXiv:1801.00631.

Mateen, H. 2018. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. *Berkeley Journal of Employment and Labor Law* 39(1), 285-292.

McHenry, H. M. and Coffing, K. 2000. Australopithecus to Homo: transformations in body and mind. *Annual Review of Anthropology* 29(1), 125-146.

Mitchell, T. M. 1997. Does machine learning work?. *AI Magazine* 18(3), 11-11.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. 2016. The ethics of algorithms: Mapping the debate. *Big Data and Society* 3(2), 2053951716679679.

Morgan, T. J., Uomini, N. T., Rendell, L. E., Chouinard-Thuly, L., Street, S. E., Lewis, H. M., ... and Laland, K. N. 2015. Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications* 6(1), 6029.

Morley, J., Machado, C. C., Burr, C., Cowls, J., Joshi, I., Taddeo, M. and Floridi, L. 2020. The ethics of AI in health care: a mapping review. *Social Science and Medicine*, 260, 113172.

Nof, S. Y. 2009. *Springer Handbook of Automation* (pp. 1379-1396). Berlin, Heidelberg: Springer Berlin Heidelberg.

Obermeyer, Z. and Emanuel, E. J. 2016. Predicting the future—big data, machine learning, and clinical medicine. *The New England Journal of Medicine* 375(13), 1216.

Olawumi, T.O., Chan, D.W.M. and Wong, J.K.W. 2017 Evolution in the intellectual structure of BIM research: a bibliometric analysis. *Journal of Civil Engineering and Management* 23(8), pp. 1060-1081

Pardede, H. F., Suryawati, E., Krisnandi, D., Yuwana, R. S. and Zilvan, V. 2020. Machine learning based plant diseases detection: A review. In *2020 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications* (ICRAMET) (pp. 212-217). IEEE.

Pradhan, P., Fischer, G., Van Velthuizen, H., Reusser, D. E. and Kropp, J. P. 2015. Closing yield gaps: how sustainable can we be? *PloS One* 10(6), e0129487.

Raffel, M., Willert, C. E., Scarano, F., Kähler, C. J., Wereley, S. T. and Kompenhans, J. 2018. *Particle Image Velocimetry: A Practical Guide*. Springer.

Rapp, A., Curti, L. and Boldi, A. 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. International *Journal of Human-Computer Studies* 151, 102630.

Rudin, C. 2019. Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1(5), 206-215.

Russell, S. J. and Norvig, P. 2010. *Artificial Intelligence is a Modern Approach*. London.

Russell, S. (2021). Human-compatible artificial intelligence. *Human-like Machine Intelligence* 3-23.

Shea, J. J. 2013. *Stone Tools in the Palaeolithic and Neolithic Near East: A Guide*. Cambridge University Press.

Sutton, R. S. and Barto, A. G. 2018. *Reinforcement Learning: An Introduction*. MIT Press.

Szegedy, C., Reed, S., Erhan, D., Anguelov, D. and Ioffe, S. (2014). Scalable, high-quality object detection. arXiv preprint arXiv:1412.1441.

Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., ... and Mahoney, P. (2006). Stanley: The robot that won the DARPA Grand Challenge. J*ournal of Field Robotics* 23(9), 661-692.

Toth, N. and Schick, K. 2009. The Oldowan: the tool making of early hominins and chimpanzees compared. *Annual Review of Anthropology* 38, 289-305.

Trivedi, D., Rahn, C. D., Kier, W. M. and Walker, I. D. 2008. Soft robotics: Biological inspiration, state of the art, and future research. *Applied Bionics and Biomechanics* 5(3), 99-117.

Wamba, S. F., Dubey, R., Gunasekaran, A. and Akter, S. 2020. The performance effects of big data analytics and supply chain ambidexterity: The moderating effect of environmental dynamism. *International Journal of Production Economics* 222, 107498.

Wang, A. X., Tran, C., Desai, N., Lobell, D. and Ermon, S. 2018. Deep transfer learning for crop yield prediction with remote sensing data. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies* (pp. 1-5).

Wang, S., Wan, J., Li, D. and Zhang, C. 2016. Implementing smart factory of industrie 4.0: an outlook. *International Journal of Distributed Sensor Networks* 12(1), 3159805.

Wruck, S., Vis, I. F. and Boter, J. 2017. Risk control for staff planning in e-commerce warehouses. *International Journal of Production Research* 55(21), 6453-6469.

Wynn, T. 2009. Hafted spears and the archaeology of mind. *Proceedings of the National Academy of Sciences* 106(24), 9544-9545.

Zhang, Y., Cheng, Y., Wang, X. V., Zhong, R. Y., Zhang, Y. and Tao, F. 2019. Data-driven smart production line and its common factors. *The International Journal of Advanced Manufacturing Technology* 103, 1211-1223.

Zhang, W., Yang, D. and Wang, H. 2019. Data-driven methods for predictive maintenance of industrial equipment: A survey. *IEEE Systems Journal* 13(3), 2213-2227.

Zhou, B., Gao, F., Wang, L., Liu, C. and Shen, S. 2019. Robust and efficient quadrotor trajectory generation for fast autonomous flight. IEEE Robotics and Automation Letters, 4(4), 3529-3536.

Zhou, L. and Zhang, D. 2003. NLPIR: A theoretical framework for applying natural language processing to information retrieval. *Journal of the American Society for Information Science and Technology* 54(2), 115-123.

Leveraging the Affordances of Artificial Intelligence for the Benefit of Higher Education in Developing Countries: A Conceptual Framework

O. Olaitan
Walter Sisulu University
oolaitan@wsu.ac.za

Abstract

This paper presents a framework that seeks to address recurring issues pertaining to the adoption and use of artificial intelligence in higher education. Addressing these issues is essential to broader sociological transformation, equitable and inclusive access in higher education.. This paper examines the landscape of adoption and use of artificial intelligence in education, with particular focus on South Africa. It delves into the challenges and opportunities inherent in use of artificial intelligence in higher education and explores how South African institutions can strategically harness artificial intelligence to navigate the changing global higher education terrain. Recognising the ethical implications of artificial intelligence, the paper also highlights the importance of having regulatory measures to safeguard student data and ensure responsible AI use.

Keywords: Artificial intelligence, artificial intelligence ethics, equity in education, higher education, sociological transformation

Introduction and background

In the digital age, the transformative potential of artificial intelligence (AI) has permeated various facets of human life, from personalised recommendations on streaming platforms to advanced medical diagnostics. In the realm of higher education, AI emerges as a powerful tool that promises to revolutionise the activities of teaching, learning, and administrative processes. At its core, AI is a multidisciplinary field of computer science that aims to imbue machines with the capacity to simulate human intelligence. This simulation encompasses various facets of cognition, including problem-solving, learning, reasoning, perception, and language understanding. AI systems are constructed upon intricate algorithms and datasets, allowing them to discern patterns, make informed decisions, and adapt their behaviour based on added datasets. Machine learning, a subset of AI, empowers systems to iteratively improve their performance through exposure to data, thereby driving the development of intelligent applications. With the aid of data, AI systems are trained to independently make predictions, make recommendations and decisions akin to those usually made by human intelligence (European Union, 2019).

In the realm of higher education, AI has revolutionised both teaching and learning experiences. For instance, AI-powered adaptive learning platforms personalise educational content to cater to individual student needs, thereby improving learning outcomes. Moreover, AI-driven tools like chatbots and virtual assistants offer immediate support to students, addressing their queries and providing guidance outside of traditional classroom hours. Similarly, AI technologies facilitate data analysis, enabling institutions to gather insights into student performance patterns and optimise educational strategies accordingly. AI is assisting educators to make data-informed decisions about course design, identify areas where students may need additional support, thus adapting their teaching strategies accordingly. AI applications are widely employed to complete administrative tasks in higher education. Tasks such as admissions, enrolment management, and student support are being allocated to AI applications to promote operational efficiency. Furthermore, AI-powered predictive analytics are employed in identifying at-risk students and implementing timely interventions to improve student success rates and graduation throughput.

In light of these considerations, this paper presents the findings of a comprehensive inquiry into the impact of AI within the realm of higher education, including the challenges and prospects thereof. It provides critical insights into how higher education institutions in South Africa can strategically leverage AI to navigate the ever-evolving global higher education educational ecosystem. It also discusses the role of AI in promoting equitable access to education without exacerbating the existing socio-economic disparities prevalent in South African society as it contends that AI can play a transformative role in addressing and ameliorating persistent and pertinent issues related to transformation and equity, which are pivotal concerns within the higher education landscape of the country.

Objectives and guiding question

The primary objective of this paper is to explore the potential challenges posed by the adoption and use of AI in higher education within developing economies, with a specific focus to South Africa. The question that the paper seeks to answer is: “How can the affordances of AI be leveraged to achieve transformation, inclusivity and equity in South African higher education context?” The paper attempts to answer this question by (a) examining the current state of education in South Africa and identify the existing challenges related to transformation, inclusivity, and equity in the system; (b) exploring the current and potential application of AI in education and its ability to address specific issues related to equity and inclusive access, such as language barriers, infrastructural challenges, and remote learning opportunities; (c) analysing the ethical considerations and potential biases associated with the implementation of AI in education, particularly regarding fairness, privacy, and transparency in decision-making algorithms; and (d) proposing a conceptual framework for policymakers, educational institutions, and other stakeholders to effectively harness AI technologies in promoting equitable, inclusive, and transformative education in South Africa.

Methodology

The material presented and discussed in the paper was obtained through systematic literature search and review. This methodology enabled the use of findings of numerous studies that have been published in scholarly literature. The material used was sourced from databases including EBSCO, Web of Science, Google Scholar, and Science Direct. Searches were confined to the domains of Education, Information Systems, Computer Science, and Information Science. The rationale for this restriction was to maintain relevance, as AI permeates multiple sectors, but the critical factors under investigation in this study were specific in nature. The aforementioned databases were also chosen for their high citation rates and the quality of systematic literature reviews available in their stables. Furthermore, the inclusion criteria were chosen to ensure that the literature selected for review aligns with the aims of the study, focusing on recent, peer-reviewed articles from reputable sources to guarantee relevance, currency, and rigorous evaluation. Exclusion criteria were designed to maintain the integrity and focus of the research by excluding materials that might not directly contribute to addressing the research question or meet the desired standards of research quality. For instance, the exclusion of papers published before 2015 serves to prioritise contemporary insights and developments in the field of AI in higher education. Similarly, the exclusion of non-English language papers mitigates potential language barriers and ensures accessibility to a broader audience. Moreover, the decision to exclude editorial reviews, opinion pieces, and preprints aims to maintain a focus on empirical research and scholarly discourse, thereby enhancing the robustness and credibility of the study's findings. By adhering to these carefully crafted exclusion criteria, the study maintains a clear and targeted approach, facilitating a more meaningful exploration of the transformative potential of AI in higher education within developing economies.

The combined use of these inclusion and exclusion criteria was intended to ensure that the paper focuses exclusively on the most pertinent and contemporary research within the sphere of artificial intelligence in higher education, specifically concerning its impact on transformation, inclusivity, and equity.

The initial database search using the relevant search string yielded a total of 3, 214 papers. Appendix 1 shows the search strings that were used to frame and interpret each objective, and to refine the search and reduce the more relevant sources to just 176 articles. A sentiment analysis was conducted to establish a scientific basis for reviewing these articles and to ascertain how scholarly discussions regarding the application of AI for educational transformation and equity within the global South. Sentiment analysis serves as a valuable tool for discerning nuances, tones, and opinions within a collection of articles *(Wankhade et al.*, 2022). Transformation and equity are emotive subjects, particularly among faculty and students in previously disadvantaged communities, such as those in South Africa, given its historical context. This methodological approach is particularly pertinent for providing insights into how the discourse is evolving, especially in comparison to the developed world where the context differs significantly. The search strings along with graphical representations from the sentiment analysis for each research objective are attached as Appendix 1 at the end of the paper.

Presentation and discussion of findings

The findings from the systematic literature review and analysis are presented and discussed under the following main headings: state of higher education provisioning in sub-Sharan Africa, state of higher education provisioning in South Africa, adoption and use of AI in higher education, ethical concerns regarding development and use of AI, and conceptual framework for ethical and transformative utilisation of AI in developing countries.

The state of higher education provisioning in sub-Saharan Africa

In an era where the intrinsic value of higher education has been overshadowed by the pursuit of survivalist affluence, the landscape of higher education in sub-Saharan Africa presents a stark contrast to global trends (Boughey and McKenna, 2021). According to UNESCO's data spanning from 2000 to 2022, there has been a noteworthy 40% increase in higher education enrolment worldwide. However, sub-Saharan Africa's contribution to this increase in 2022 stood at a mere 9%, a striking disparity when compared to the remarkable 79% increase witnessed in Europe and North America during the same period (UNESCO, 2022). This glaring inequality in enrolment growth reflects the entrenched challenges faced by the Sub-Saharan African higher education system.

The Covid-19 pandemic laid bare the chasm in access, provision, and quality of education across different global regions, serving as a harsh reminder of the inequities embedded in higher education. UNESCO's research, as illustrated on Table 1, succinctly captures the inequality that defines the state of higher education in sub-Saharan Africa. This disparity, which extends across various dimensions, underscores the urgent need for comprehensive reforms and concerted efforts to bridge the educational divide, ensuring that higher education in the region can fulfil its potential as a catalyst for socio-economic development and equitable progress (Formunyam, 2020; Wamba *et al*., 2023).

Table 1:The impact of Covid-19 on Higher Education around the World. I4IJ Global Survey (Marinoni *et al*, 2020)

Global region	**Not affected**	**Classroom learning replaced by distance teaching and learning**	**Teaching suspended but the institution is developing solutions**	**Teaching cancelled**
Africa	3%	29%	43%	24%
Americas	3%	72%	22%	3%
Asia and Pacific	1%	60%	36%	3%
Europe	Almost zero	85%	12%	3%

Information in Table 1 also reveals that African higher education institutions (HEs) faced significant challenges. They experienced a teaching cancellation rate of 24%, a stark contrast to the global average of just 3%. This report underscores a crucial point: the enthusiastic promotion of digital transformation, including Artificial Intelligence Digital Transformation (AIDT), does not adequately account for the unique challenges faced by Africa. SDG 4 aims to ensure equal access to education for individuals of all racial backgrounds. However, the reality in most of the global South is far removed from this ideal. Over the years, discrimination, racism, and negative stereotypes have insidiously taken root within institutional structures, perpetuating disparities in educational opportunities (Kiemde and Kora, 2022; Mhlanga and Moloi, 2020).

The state of higher education provisioning in South Africa

The South Africa's higher education system has undergone significant changes since the end of apartheid. However, transformation and equity remain a slow and contentious process (Mhlanga and Moloi, 2020). The legacy of apartheid, characterised by institutionalised racism and discrimination, continues to cast a long shadow over South African higher education (Boughey and McKenna, 2021). While policies and frameworks exist to promote transformation, their implementation is inconsistent and lacks the necessary funding and oversight (Formunyam, 2020). Government officials often prioritise maintaining political stability over confronting systemic inequalities head-on, contributing to the persistence of an inequitable system (Williamson *et al*. 2020). Furthermore, South African universities continue to utilise curricula that are rooted in colonial knowledge systems, perpetuating the marginalisation of indigenous knowledge and epistemologies. These curricula prioritise Western ways of knowing and learning, leaving many students disconnected from their cultural heritage and epistemological traditions (Boughey and McKenna, 2021). This disconnect further intensifies feelings of alienation and disempowerment among Black students. The epistemological perspectives of the Black majority in South Africa often remain on the periphery of the higher education system (Boughey and McKenna, 2021). The lack of recognition and incorporation of indigenous epistemologies into the curriculum perpetuates a hierarchical and exclusionary system that places non-Western ways of knowing at a disadvantage. This marginalisation reinforces the sense of 'othering' experienced by Black students and hinders their full participation and empowerment within the academic community. The slow pace of transformation and equity in South African higher education is a deeply rooted issue with multifaceted challenges (Adams, 2021; Boughey and McKenna, 2021; Chaka, 2022; Davids, 2016; Mhlanga and Moloi, 2020). Achieving transformation requires a commitment to dismantling these barriers through comprehensive policy reform, development academic and support staff, and a shift in institutional culture.

Adoption and use of AI in higher education

AI has been adopted in higher education institutions in the midst of all the challenges discussed above, and there is an expectation that its affordances should contribute towards addressing the challenges. In line with Millenium Sustainable Development Goal (SDG) number 4 UNESCO (2018) states that the desired role of AI is ensuring inclusive and equal opportunities for all people in spite of their station in life. The use of AI includes learning analytics to facilitate informed decision-making, fostering collaborative learning among individuals situated in diverse geographical locations, delivering tailored and personalised educational experiences, and offering dual teacher/bot-assisted learning approaches (Okewu *et al*., 2021; Wamba e*t al*., 2023; Williamson *et al*., 2020). AI is also being leveraged to map learning trajectories, customise learning pathways, and provide specialised assistance to learners. Notably, AI-driven grading systems are being deployed for essays in countries like Uruguay, China, and the United States (Anshari and Hamdan, 2022). In a study to assess the impact of Covid-19 on higher education, Zhou (2023) found that, in South Africa, during the lockdown, a variety of AI tools were utilised across all spheres of education and training from primary education to higher and tertiary education where educational activities switched to remote (online) learning. The pandemic presented an opportunity to assess successes and failures of deployed technologies, costs associated with them, and scaling these technologies to improve access (Slimi and Carballido, 2023).These AI-enabled technologies present valuable opportunities to automate routine and administrative tasks, such as grading and recordkeeping, which are presently conducted by educators. The automation of these tasks has the potential to liberate educators' time, enabling them to redirect their efforts towards the more creative, empathetic, and inspirational aspects of their profession.

A study by Chaka (2020) revealed that one of the primary AI technologies shaping the landscape of learning are chatbots. Furthermore, AI exhibits the potential to revolutionise education by offering the promise of tailored, scalable, and cost-effective learning solutions. Additionally, the study further stated that there is focus on the exploratory nature of robotics applications within the educational sphere, with an emphasis on meta-teaching and meta-learning paradigms. Additionally, the integration of blockchain technology is manifesting in several areas, including digital grading, digital credentialing, and digital certification processes. Moreover, blockchain finds relevance in real-time contracting and time stamping of learning activities, further underscoring its multifaceted impact on education (Chaka, 2022). AI technologies have primarily found traction in the amelioration of administrative functions within the academic sphere. Notably, these AI applications have been directed towards enhancing administrative tasks traditionally undertaken by faculty, such as the implementation of blockchain technology for credentialing, fostering collaborative learning environments, and monitoring the progression of student learning. In contrast, the deployment of AI in a transformative or pioneering manner, one that fundamentally augments educational opportunities for marginalized or decontextualized student populations, appears to be scarce within the current academic landscape. These observations are consistent with the findings of Boughey and McKenna (2021), who have illuminated these trends in their scholarly contributions.

Okonkwo and Ade-Ibijola (2021) document the use of chatbots to teach students programming language. Benefits of using chatbots highlighted in the paper include the integration of content, immediacy of assistance, instantaneous support for students and in some cases, personalised learning support. One of the most important benefits of chatbots is the repeatability of its tasks. The repetitive tasks of disseminating basic information are now carried out by machines, freeing administrative staff as well as faculty to engage in more meaningful, human centred tasks(Williamson *et al*., 2020). It is interesting to note that these uses of AI do not focus on sociological transformation or equity in the South African context as it presupposes that all students have equal access to the platforms that enable the AI bots.

The research conducted by Boughey and McKenna (2021) illuminates the challenges faced by numerous students who do not receive full financial support from the national government. Consequently, these students were compelled to undertake extensive daily commutes from their residences to the campus and back. Their access to unlimited Wi-Fi connectivity is confined to the campus environment, where they are obliged to attend classes and fulfil various academic obligations. Consequently, there exists limited leeway for them to explore the integration of chatbot technologies as a means of enhancing their academic pursuits. This circumstance underscores the dependence of such students on on-campus resources, constraining their capacity to leverage AI-driven chatbots for academic enrichment. The implications of this constraint warrant further examination.

Mhlanga and Moloi (2020) highlight the pivotal role played by Industry 4.0-related technology during the Covid-19 pandemic, particularly in facilitating educational continuity while adhering to imperative social distancing measures. Mhlanga and Moloi (2020) further posit that technology, including AI, possesses the inherent potential to broaden access to education, particularly in the realm of online learning. They assert that it effectively transcends spatial limitations, rendering education more accessible to a broader spectrum of learners. This attribute has minimal importance in regions where physical infrastructure may be lacking, further emphasizing the potential of AI technology to widen the digital divide.

AI has a substantial impact on assisting students in selecting courses (Maphosa and Maphosa 2022). In the context of online education, teacher-bots are progressively taking the place of teaching assistants during tutorial sessions and managing specific administrative tasks in the teaching process (Popenici and Kerr 2017). This shift relieves educators from the repetitive and routine aspects of their work, enabling them to concentrate their energies on more advanced and intellectually challenging pursuits.

Mhlanga and Moloi (2020) report that AI is used to promote inclusivity and equity in South African higher education, making higher education more accessible but also create opportunities for lifelong learning, thereby addressing disparities and ensuring that education is available to all. Another popular use of AI is found in the field of data analytics, machine learning and artificial neural networks. Due to the widespread adoption of technology-driven educational approaches such as virtual learning, e-learning, and blended learning, concerns among researchers and stakeholders about the ability of higher education to produce graduates with the essential analytical skills and adaptability needed in the 21st century knowledge economy, both for personal and national success, are being alleviated.

From the perspective of higher education management, the remedy lies in the realm of data analytics (Okewu *et al*., 2021). Researchers have intensified their efforts to leverage data for the enhancement of teaching and learning. This involves the utilisation of insights derived from academic records, employing techniques such as artificial neural networks, predictive analytics, and machine learning. These methods are harnessed to make sense of student data in ways that foster student success and boost overall academic progress (Okewu *et al*., 2021). AI possesses the capability to forecast learners' academic performance, thereby exerting a direct influence on their scholastic achievements (Teng *et al*., 2023). Such predictive abilities enable the implementation of timely remedial measures to mitigate the risk of course failure.

Ethical concerns regarding development and use of AI

It is important to note that a significant challenge known as 'overfitting' may present in the outcome of datasets for students. Overfitting occurs when a machine learning model becomes too fixated on the specific details of the training data, fitting it excessively closely *(Okewu et al*., 2021). Consequently, it becomes less adept at predicting new, unseen data in the future. This is akin to a student who memorises answers for a specific test but struggles to apply their knowledge to answer new, unfamiliar questions. In underdeveloped countries, where resources and expertise in machine learning may be scarce, overfitting can be particularly problematic. Models used for decision-making in critical higher education domains could be susceptible to overfitting, yielding unreliable results and failing to address the unique challenges facing these countries. This could have adverse repercussions for decision-making based on the strength of predictive analytics with regards to higher education and student profiling in developing countries, as it can lead to unreliable predictions and impede the effective resolution of pressing issues. Additionally, biased algorithms tend to favour data creators, empathisers, pattern-recognition experts, and meaning makers (Slimi and Carballido, 2023). This the basis of concerns about to the impartiality of AI outcomes and the subsequent decisions made, particularly in the context of promoting equity and driving transformation in higher education within South Africa and similar regions in the global South. The use of AI and other advanced technologies in this instance can be detrimental rather than progressive to transformational efforts.

Additional concerns revolve around the utilisation of predictive analysis in decision-making and student support, often without taking into account the unique circumstances of each student, and the collection of student data without obtaining their consent (Huang *et al*., 2021). The pervasive practice of datafication has permeated every facet of tertiary education, ranging from recruitment to assessment practices. While this system may appear highly efficient, it is disconcerting that there is currently no evidence to suggest adequate monitoring of the data input into AI systems to detect potential biases and racial profiling (Gardner, 2022; Hagerty and Rubinov, 2019; Slimi and Carballido, 2023). Huang *et al*. (2021) argue that the biases ingrained in algorithm construction and the involuntary, yet inescapable, surveillance of students demand an immediate discussion regarding the fairness of quantifying students' experiences based on data collected without their consent or awareness.

Related to this loss of agency over one's data are common practices in the contentious field of AI data analytics for human-impacting decision making. Academic staff, management, and funding bodies all employ predictive analytics in decision-making that affects the careers and educational experiences of students. Some scholars consider predictive analytics as a novel and relevant tool in assessing student progress and targeting definite 'pain points' (Otto *et al*., 2023). Others believe that learning analytics do not take the context and lived experiences of the student into consideration, thereby obliterating the individualism of persons and reducing them to a set of data (Williamson *et al.*, 2020). With data analytics, sense making or in-depth analysis seems to have been taken from students, thus shifting agency to the AI which acts and perceives students as 'data transmitters' rather than human agents possessing willpower and opinions who can be engaged in dialogue about the role of machines in their lives (Williamson *et al* 2020; Bearman 2022).Those who are mostly affected by the sweeping and disproportionate opportunities are the marginalised minority, in which case the technology, rather than advance the cause of equality and transformation, has become another tool of oppression and othering of certain members of society (Formunyam, 2020; UNESCO, n.d.).

Advocating for the essential role of human engagement within technology-driven higher education, Popenici and Kerr (2017) emphasises that, despite the remarkable advancements in AI, over-reliance on technology poses a perilous path. In the South African education sector, there is a concerning rise in the level of specialisation required for technology-centric occupations like machine operators, surpassing what was previously observed, leading to an exacerbation of the digital divide, and increased economic disparities (Formunyam, 2020). Significantly, attention should be directed toward the pivotal contribution of human beings in processes such as problem identification, critical evaluation, risk assessment, and the formulation of essential inquiries. These inquiries may encompass a wide range of concerns, including issues related to privacy, authority structures, human intuition, and governance frameworks. Furthermore, fostering an environment conducive to creativity, unexplored domains, and serendipitous discoveries is deemed integral to the teaching and learning experience in higher education (Okewu *et al*., 2021; UNESCO, n.d.). One of the challenges associated with excessive dependence on AI is the potential loss of opportunities for critical thinking, independent interpretation of pedagogical principles, and exploratory discourse within the intricate landscape of machine learning and automated tools (Huang *et al*., 2021).

Kiemde and Kora (2022) found that most research on AI ethics and guidelines on ethical use of AI is from Europe and America. The authors reported a paucity of research into ethical use of AI in higher education in global publications, and an acute shortage of ethical guidelines authored by Africans to guide the use of AI on the continent. This presents a paradox because the continent is the hardest hit when it comes to the unethical application of AI, including racially biased algorithmic data, datafication taken out of context, and the extensive commercialization of AI by private entities marketing their products as the ultimate solution of our modern era. The mandate of the Independent High-Level Expert Group on Artificial Intelligence established by the European Commission was to formulate ethical guidelines promoting the development and deployment of trustworthy AI within the European Union (EU, 2019). These guidelines serve as a blueprint for fostering responsible AI practices and underscore key principles such as fairness, transparency, and accountability, with the overarching goal of ensuring that AI technologies are human centred and conform to European values and legal structures. The guidelines have been used referentially to further develop AI ethical policies for European nations (EU, 2020).

Arthur (2019) and Sallstrom *et al* (2019), among others, advocate measures to combat the dangers of unstructured use of AI Africa. Among the proposed principles are the definition and integration of African values in the contextual application of AI in education, and a community and co-creation approach to socially acceptable AI. African researchers, with an understanding of the socio-cultural contexts of the continent, must be at the forefront of ensuring African focused datasets are employed on the continent. The underlying logic of an AI technology must align with the value and context of the African people otherwise Africa is entering into what Kieme and Karo (2022) termed as “cyber-colonialism” Although the ideas and principles advocated by these authors are bound to reduce the ethical conundrum, it is vague with regards to how the proposed changes will be effected. Kiemde and Kora, (2022), therefore proposed a comprehensive and diversified team of developers in the development of AI technology in, and for Africans. The dictatorial and prescriptive manner in which the West has traditionally imposed its values and systems on Africans will continue and expand through AI if Western technologists are able to write the rules by themselves but for Africa (Boughey and McKenna, 2021; Hagerty and Rubinov, 2019). Kiemde and Kora, (2022) believe that datasets generated without the input of Africans stand the risk of accelerating already deep- seated inequality in society. Onaolapo and Onifade (2020) propose a comprehensive categorisation of the challenges facing HE in Africa concerning AI into three overarching domains: policy, techno-economic, and security. Within the existing global order, Onaolapo and Onifade (2020) argue that the benefits derived from AI capabilities are inherently skewed in favour of Western nations, raising a significant concern.

The optimisation of AI for educational purposes necessitates the presence of a conducive environment characterised by adequate infrastructure, cutting-edge data facilities, and a proficient AI workforce (Nawi *et al.*, 2021; Tundrea, 2020; UNESCO, n.d.). Unfortunately, these essential elements are notably deficient across the African continent. Furthermore, the insufficient technical infrastructure, coupled with a deficit in AI expertise, data deficits, and unfavourable regulatory frameworks, constrains the application of AI for educational purposes in Africa.

The ramifications of AI-induced biases extend far beyond the confines of algorithms and circuits, delving into the intricate fabric of society itself. Hagerty and Rubinov (2019)'s research underscores how these biases breed a pervasive culture of societal othering, where specific groups are systematically marginalized and excluded. This culture of exclusion reverberates throughout higher education, from the seemingly objective admissions processes to the nuanced terrain of student monitoring and assessment, as poignantly outlined by Huang et al. (2021). The consequence of these biases is a deeply rooted social division that may undermine the very foundations of equitable access to education.

Formunyam, (2020) underscores the paramount importance of decolonising AI implementation in Africa, emphasizing its potential to revolutionize educational curricula and establish AI as a catalyst for equity and transformation, rather than perpetuating a colonialist agenda embedded with biased algorithms that marginalize minority communities. AI systems, as part of the justification for concerns about bias in AI decision-making, tend to perpetuate and normalise social inequality. This issue is evident in notable cases in the United States (Hagerty and Rubinov, 2019).. For instance, job recruitment tools have exhibited bias against women, discriminatory credit algorithms have posed challenges for Latino and African American borrowers, and issues related to race, gender, and sexual orientation bias have been observed in sentiment analysis systems, natural language processing technologies, and the datasets utilized to train image recognition systems(Hagerty and Rubinov, 2019).

In light of the findings regarding the non-neutrality of AI, and negative perceptions about its ethical trustworthiness, it is important to develop guidelines and recommendations for policymakers, educational institutions, and other stakeholders to effectively harness AI technologies in promoting equitable and transformative education in South Africa.

A thorough analysis of the use of AI, its ethical conundrum and placement within the context of transformation and equity, as defined by this study has been conducted in this paper. The discussion has highlighted some of the challenges around the application of AI in higher education, especially in developing economies. Several authors have proposed different ways of realigning the use of AI in ways that do no harm but encourages the application of the affordances to promote equity and transformation. According to the expert group, the most important task of guidelines and policy for AI use is that it must be human-centred, legislature compliant and promote fairness. To this end, the EU's expert group Ethics Guidelines For Trustworthy AI Report provides critical insights into the key factors for the equitable use of AI in education, with implications for sociological transformation and equity, particularly in the global south. The report emphasises transparency and accountability as fundamental principles, insisting that educational stakeholders must be fully informed about AI-driven educational tools and that mechanisms for accountability are in place to rectify any adverse AI-generated outcomes. Furthermore, it underscores the paramount importance of data privacy and protection, ensuring that sensitive information, especially of vulnerable individuals in the global south, is safeguarded in the implementation of AI for education.

Non-discrimination and inclusivity are central tenets of the guidelines, condemning any AI-driven discrimination and calling for AI systems designed to cater to a diverse array of learners. The report advocates for preserving human oversight and control in educational AI, recognizing the irreplaceable role of human educators in maintaining educational quality and equity. Finally, it calls for continuous evaluation and accountability to ensure AI systems align with ethical standards, emphasizing the need for collaboration between AI developers and educators in designing education-specific AI solutions. By adhering to these principles, the responsible use of AI in education can significantly contribute to fostering equitable, inclusive, and transformative educational practices, particularly in regions facing sociological challenges in the global South.

The guideline recommends three critical areas where AI developers/creators must align any AI system with throughout its lifespan, namely (a) compliance with overarching legal requirements of the land, (b) an ethical framework to guide the AI in its work and decision making and (c) flexibility in both socio-technical and academic instances, in ways that promote the greater good and equity in society. The guidelines are aligned with recommendations from (UNESCO, n.d., 2019). Policies governing AI utilisation in developing economies should address questions about the empowerment of disadvantaged groups and populations. These questions should explore the potential of AI in enhancing education for underprivileged communities, examining how digital education and AI technologies can be harnessed to accelerate educational progress in developing nations and bridge the stark divide between affluent and marginalized students globally. Additionally, such policy considerations should delve into the realm of gender equality, probing the best practices in leveraging AI to narrow gender gaps, particularly focusing on women and girls.

Conceptual framework for ethical and transformative utilisation of AI in developing countries

The conceptual framework for promoting the ethical and transformative utilisation of AI in developing countries is underpinned by the observed impacts of AI in higher education in developing countries, with a focus on South Africa. The framework seeks to build on the positive impacts of AI, while addressing the negative impacts of AI. The impacts as, discussed, are summarised in Table 2 below.

Table 2: Summary of Findings on the Impact of AI in Higher Education (Olaitan, 2024)

Finding	Key Points	Authors
Inequality persists in South African higher education	Historical disparity on access, success, and support systems.	Boughey and McKenna, 2021; Wamba *et al*., 2023; Mhlanga and Moloi, 2020.
Slow pace of transformation	Many socio-economic and political factors hinder efforts to transform.	Boughey and McKenna, 2021; Huang *et al*., 2021, Chaka, 2022; Davids, 2016; Mhlanga and Moloi, 2020.
Persistent disparities hindering equitable performance outcomes	Structural and technical disparities persist between different types of institutions.	Davids, 2016; Kiemde and Kora, 2022; Williamson *et al*., 2020; Boughey and McKenna, 2021.

Impact of AI in higher education is slow but on the uptake.	Learning analytics, data analysis and bot assisted learning are on the uptake at many institutions.	Bearman *et al* (2022), Okewu *et al*., 2021; Wamba *et al*., 2023; Williamson *et al*., 2020.
Impact of Covid-19 and opportunities for AI	This crisis highlighted the potential of AI to increase access to education through remote learning and automate administrative tasks.	Formunyam, 2020; Wamba *et al*., 2023; Mhlanga and Moloi, 2020.
Government initiatives and AI integration	Some countries, such as Brazil, have introduced state-endorsed educational platforms incorporating AI.	UNESCO, n.d; Durso and Arruda, 2022.
Limited African representation in literature	There is a notable scarcity of African representation in literature on AI in higher education.	Kayembe and Nel, 2019; Badat, 2010; Boughey and McKenna, 2021.
Focus on administrative enhancement	AI applications in South African higher education predominantly target administrative tasks traditionally performed by faculty.	Otto et al., 2023; Okewu *et al*., 2021; Wamba *et al*., 2023; Williamson *et al*., 2020.
Challenges for students from formerly disadvantaged background	Certain marginalized students are precluded from leveraging AI-driven technologies for academic enrichment. Students reliant on on-campus resources have limited opportunities to benefit from AI applications due to spatial constraints.	Boughey and McKenna, 2021; Chaka, 2022; Moosa, 2018; Mhlanga and Moloi, 2020.
Potential of AI in promoting inclusivity	Despite challenges, AI, including teacher-bots and data analytics, has the potential to promote inclusivity and equity in South African higher education.	Okonkwo and Ade-Ibijola, 2021; Mhlanga and Moloi, 2020.
Concerns about bias and overfitting	Biased algorithms and overfitting in machine learning models underscore the importance of ensuring fairness and reliability in AI applications.	Formunyam, 2020; Popenici and Kerr, 2017.
Limited customisation for local contexts	Despite the potential for AI to address unique challenges in South African higher education, there is limited customisation for local context.	Boughey and McKenna, 2021; Davids, 2016; Kiemde and Kora, 2022; Williamson *et al*., 2020.
Dominance of AI as digital literacy and administrative assistance	AI applications in South African higher education predominantly focus on enhancing digital literacy, providing administrative support through chatbots, and improving assessment processes.	Okewu *et al*., 2021; Teng *et al*., 2023.
Ethical concerns and potential biases	The literature highlights instances where AI algorithms perpetuate racial biases and societal othering, exacerbating existing inequalities within South African higher education.	Huang *et al*., 2021; Hagerty and Rubinov, 2019; Gardner, 2022; Slimi and Carballido, 2023.

Need for localised ethical guidelines	The absence of ethical guidelines risks exacerbating inequalities and perpetuating neo-colonialism through AI technologies imported from Western countries.	Arthur, 2019; Sallstrom *et al*., 2019; Kiemde and Kora, 2022. Slimi and Carballido, 2023; Arthur, 2019; Sallstrom *et al*., 2019.
Infrastructure and expertise constraints	The effective implementation of AI in South African higher education is hindered by inadequate technical infrastructure, a shortage of AI expertise, and limited access to relevant data.	Nawi *et a*l., 2021; Tundrea, 2020; UNESCO, n.d.; UNICEF,2020.
Biases in automated proctoring	Studies reveal biases in automated proctoring software used in higher education settings, with darker-skinned and black students disproportionately targeted for scrutiny.	Himes *et al*., 2023; Intahchomphoo and Gundersen, 2020; Williamson *et al*., 2020; UNESCO, n.d.
Impact on equity and social division	AI-induced biases contribute to social division and undermine equitable access to education within South African higher education.	Hagerty and Rubinov, 2019; Huang *et al*., 2021; Williamson *et al*., 2020; Formunyam, 2020; (Davids, 2016; Gardner, 2022; Mhlanga and Moloi, 2020; Moosa, 2018.
A dearth of policy guidelines for AI use in higher education	A multifaceted approach is required for policy on AI driven decision-making, which could negatively impact the students if their context is not taken into consideration in the application of AI analysis and decisions.	Tundrea, 2020; UNICEF, 2020.

The framework is cyclical as a way of depicting the fact that it must constantly adapt to changing circumstances, be open for improvement and allow for regular improvement, updates and revisions. At the centre, the framework shows the critical factor (AI) and the unit of analysis (HE), surrounded by all the components that are important to the activation of an inclusive and equitable use of AI in higher education.

The framework comprises six integral components aimed at fostering the ethical and equitable integration of artificial intelligence (AI) in the higher education sector. Grounded in African values and guided by ethical considerations, the framework emphasises the intersection of AI with socio-cultural context and ethical principles. These components are envisioned as a cyclical framework, where each component interacts with and influences the others, yielding an interconnected approach to the ethical and inclusive application of AI in higher education. The components are discussed below.

Inclusive Data Development

The first component addresses the issue of inclusive data development, particularly focusing on empowering disadvantaged groups and promoting inclusion and equality. As the literature has revealed, the majority of creators and developers of AI originate in the Global North where technology affordances are more widespread. This is not the case in South Africa where connectivity is still difficult for people in informal settlements and wetlands (Mhlanga and Moloi, 2020; Toda *et al*., 2022),

Sociological transformation, ethical equity and African values

This component of the framework underscores the paramount importance of aligning AI applications for South Africa with African values and ethical guidelines (Boughey and McKenna, 2021; Williamson *et al*., 2020). This component advocates for engagement and co-creation of relevant data, in AI development and stresses the necessity of utilizing African-focused datasets with socio-cultural context. If this is done, AI datasets will be truly representative of the values, which are unique and meaningful to citizens and residents of the global South. Decision-making based on such datasets are likely to hold contextual nuances which represent the beliefs and lifestyles of the African people. Socio-technical approaches to algorithmic audits acknowledge that algorithms are not impartial technical instruments; instead, they are intertwined with intricate social and cultural environments. Upholding principles of equality, non-discrimination, and solidarity are integral in this component. Furthermore, it emphasizes inclusive data representation and respect for human dignity, calling for the decolonization of AI and reorienting higher education curricula(Bearman *et al.*, 2023; Formunyam, 2020). In this component, it is expected that AI becomes an enabler of local languages to provide an edge to students whose first language is not English. AI should also develop rather than truncate epistemic ideologies out of the global South and design datasets in a manner that is respectful of the lived experiences of the students *(Bearman et al*., 2023; Gardner, 2022; Slimi and Carballido, 2023). It should promote a nuanced understanding of cultural context and suggests investigating social implications of AI through ethnographic research. Prioritising funding for comprehensive research on AI's social impacts is an essential part of this component (Durso and Arruda, 2022).

Policy, techno-economic, and security domains

This component focuses on the overarching policy, techno-economic, and security challenges surrounding AI integration in education. It underscores the need to prioritise network infrastructure, computing resources, and connectivity to enhance AI applications. Furthermore, the framework advocates for the development of regulatory frameworks for AI in education and underscores the importance of comprehensive data protection and privacy(European Commission, HLEG on Artificial Intelligence., 2019; Kiemde and Kora, 2022). Recognising the internet as a fundamental human right is another tenet, alongside advocating for international collaborations for infrastructure development. Widespread internet connectivity is a vital key to successful implementation of AI in higher education.

Human-centred and trustworthy AI

Human-centred AI is the core principle of the framework, emphasising the ethical and fair design and use of AI systems. Transparency, accountability, and maintaining human oversight are pivotal components within this section (Hagerty and Rubinov, 2019; Huang *et al*., 2021). Continuous evaluation and accountability mechanisms are stressed to ensure adherence to ethical standards.

Human-centeredness in AI appropriation in HE will ensure that decisions which could impact the lives of students are not left solely in the hands of AI (Huang *et al*., 2021; Zhou, 2023). This component of the framework advocates recommendations for policymakers and educational institutions to promote responsible AI development and deployment(Slimi and Carballido, 2023; Zhou, 2023). It highlights the need for ethical AI practices and mechanisms for accountability and transparency (European Commission, 2019; Kiemde and Kora, 2022).

Global cooperation and stakeholder collaboration

Global collaboration is key to harmonising AI guidelines and policies. This component emphasises the sharing of best practices for AI application in diverse cultural settings (Kiemde and Kora, 2022). The component calls for international cooperation to ensure AI's equitable and transformative use globally (European Commission, 2019). It also emphasises stakeholder collaboration as central to navigating the uncertain future of AI in education (Davids, 2016; UNESCO, n.d.). It calls for fostering collaboration between AI developers, educators, and policymakers globally, and encouraging multidisciplinary teams to ensure AI systems align with ethical principles and societal needs for diverse people groups (Davids, 2016, 2020; Williamson *et al*., 2020).

Equity in educational opportunities

This component focuses on ensuring that AI technologies facilitate equitable access to education, bridging socio-economic divides, and catering to marginalised populations. It addresses issues related to cultural relevance and accessibility in AI-driven education (Bearman *et al*., 2023; Formunyam, 2020; UNESCO, n.d.). The literature has shown that many students from disadvantaged backgrounds have limited access to the technologies that are available for their assistance outside the campus, yet they are tested at par with those with unlimited access (Boughey and McKenna, 2021; David *et al*., 2022; Gardner, 2022). The critical concern of bias in AI decision-making must be addressed in this component.

To effectively implement an AI framework for education within the context of South Africa, particularly in higher education, it is crucial to consider various factors ensuring inclusivity, equity, and ethical use. The next section proposes some practical guidelines for implementing the proposed Ethical AI Framework in South African Higher Education.

The framework is heavily influenced by the AI expert group’s report on the Ethics of AI, and the components are derived from the recommendations from several authors on the subject. It is important to state that the framework focuses mainly on how AI can bring about transformation, equity and inclusivity in HE. Although digital transformation is a subset of sociological transformation, the latter is the main focus of this paper. The framework is thus presented in Figure 1.

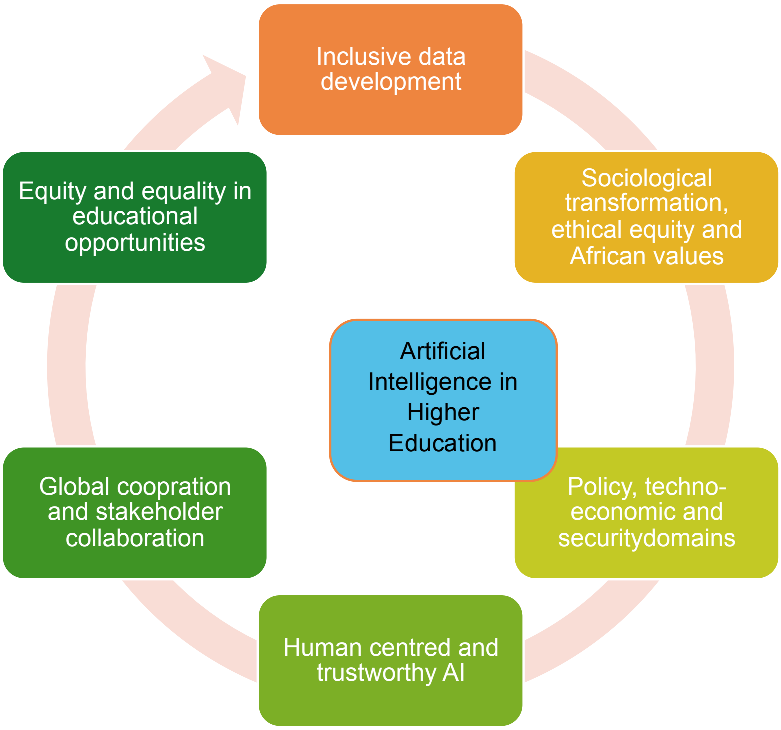

Figure 1: Framework for Transformational and equitable use of AI in Higher Education (Olaitan, 2024)

The effective implementation of the framework should be undertaken with the following factors in mind.

Contextual sensitivity: Understand the unique challenges and contextual intricacies of South Africa. Conduct a feasibility study to assess factors such as infrastructure limitations, access to technology, language barriers, and cultural nuances relevant to educational outcomes.

Stakeholder engagement in policy development: Identify key stakeholders and involve them in the policy development process. Develop policies addressing identified challenges collaboratively with stakeholders to ensure inclusivity and relevance.

Multifaceted approach: Adopt a systems approach considering each layer of AI application and design it around the user for optimized process management, equity, and user support.

Cultural and social context considerations: Acknowledge regional differences and cultural contexts in the impact of AI on society. Invite experts and scholars in cultural and social contexts to ensure AI systems are developed and deployed with consideration for cultural nuances.

Research and ethical considerations: Undertake rigorous, independent ethnographic research to understand the ethical and social implications of AI across different cultures. Adhere to UNESCO's AI ethical principles emphasizing human autonomy, prevention of harm, fairness, and explicability transparently.

Human-centred guidelines: Craft guidelines prioritising inclusion, sociological transformation, and ethical equity. Uphold fundamental principles such as equality, non-discrimination, and solidarity to guide AI system development and deployment.

Inclusive and respectful data collection: Treat individuals with respect as moral agents, not merely data points. Consider potentially vulnerable individuals and communities to avoid biased outcomes. Safeguard the well-being, personal and cultural identities of all individuals during data collection.

Drawing insights from existing recommendations: Implementers should draw insights from existing recommendations and expert reports on AI ethics to address recurrent issues and assure AI trustworthiness.

Fundamental principles such as equality, non-discrimination, and solidarity, encompassing the rights of individuals susceptible to exclusion, should serve as the guiding beacons in crafting human-centred guidelines for AI use in HE. Crucially, equal reverence for the moral value and dignity of all human beings must be upheld, transcending the realm of mere non-discrimination, which, at times, permits distinctions founded on justifiable criteria in disparate circumstances.

Within the sphere of AI, equality assumes a pivotal role, mandating that the operations of AI systems refrain from generating unfairly biased outcomes. To actualise this, it becomes imperative that the data employed to train AI systems be as inclusive as possible, accurately representing diverse demographic and societal groups. Moreover, such inclusivity demands due consideration for potentially vulnerable individuals and communities, encompassing but not limited to labourers, women, persons with disabilities, ethnic minorities, children, consumers, and other marginalised segments of society.

The concept of respecting human dignity takes on profound significance. It underscores the necessity of treating all individuals with the respect due to them as moral agents, recognising their intrinsic worth and agency. AI systems, therefore, should be meticulously developed to uphold and safeguard the physical and psychological well-being, personal and cultural identities, and the fulfilment of essential needs of all human beings, rather than treating them as things requiring sifting, sorting and labelling for data processing only (Dalalah and Dalalah, 2023; Davids, 2016; Hagerty and Rubinov, 2019; Kiemde and Kora, 2022).

Adams (2021) stated that to embark on the mission of harnessing AI for the betterment of the global South in a way that promotes transformation and equity, it is essential to recognise that decolonising AI necessitates a critical examination of its underpinnings, which are intrinsically linked to colonial power structures and the segregating mechanisms of racialisation. In the same vein, Formunyam, (2020) argues for the reorienting the higher education curriculum to reflect the post- colonial realities of the global South. However, the incursion of unregulated AI in Higher Education would further widen the divide hence the important step of enacting policies and programmes to guide its use for all aspects of the education cycle in tertiary institutions.

Conclusion

In conclusion, the proposed framework for the transformational and equitable use of AI in higher education in the global South is aimed at addressing the complexities and challenges of integrating AI into higher education systems. This framework, grounded in African values and ethics, emphasises the need to align AI applications with local socio-cultural contexts, prioritise inclusive data development, and promote equity and inclusivity. It also addresses key policy, techno-economic, and security considerations, ensuring that the infrastructure is in place for successful AI implementation.

Human-centred and trustworthy AI is at the core of the framework, emphasising transparency, accountability, and the importance of maintaining human oversight in AI systems. Moreover, it highlights the need for collaboration among various stakeholders, both at a national and international level, to ensure that AI in education aligns with ethical principles and serves the diverse needs of the Global South. The framework recognises that equity in educational opportunities is paramount, and AI should be harnessed to bridge socio-economic divides and cater to marginalised populations. It addresses the issue of bias in AI decision-making, which is crucial in ensuring that AI technologies do not exacerbate existing inequalities. Practical guidelines for implementing the framework are presented. The use of the framework will facilitate inclusiveness and fairness to counter marginalisation or discrimination against any cohort of students within the South African higher education.

Recommendations

It is recommended that higher education institutions in the global South should actively participate in the development of AI policies that consider their unique socio-cultural contexts. These policies should prioritise equity, inclusivity, and ethical use of AI. Furthermore, governments and institutions should invest in improving data accessibility and connectivity in underserved areas to ensure that inclusive data development can take place. This can involve public-private partnerships and infrastructure development initiatives. Higher education curricula should incorporate ethics and socio-cultural considerations in AI development. This will prepare the next generation of AI professionals to develop systems that respect local values and needs. Governments and funding agencies should prioritise and allocate funding for research on AI's social impact in education. This can help drive evidence-based decision-making and policy development. Educational institutions should establish mechanisms for continuous evaluation and accountability in AI systems, ensuring that ethical and equity standards are met. Equity impact assessments should be part of the implementation process. If there is commitment to these recommendations and continuous refinement of the framework through research and practical application, AI will be well positioned to play a positive role in higher education while promoting social transformation, equity, and inclusivity in the global South.

Future research

Future research should focus on the development of AI algorithms that are not only ethical but also culturally sensitive. Investigating the potential impact of culturally-aware AI on educational outcomes and experiences is an essential area of exploration for achieving the desired outcome for the study. Additionally research into an understanding of how the policy, techno-economic, and security components are implemented and the associated challenges encountered in real life settings, when these steps are taken, is crucial. This can involve case studies of different countries and regions within the global South.

Research should delve into how the human-centred and trustworthy AI component affects student experiences, practices of academic staff, and administrative decision-making in higher education institutions. It is also important to explore the effectiveness of international partnerships in promoting equity and inclusivity. Finally, all of the components involve a commitment and dedication to social transformation, inclusivity and equality in higher education. Studies reflecting continuous monitoring of AI systems for bias and discrimination must remain an on-going area of research. Developing tools and methodologies for identifying and rectifying bias in AI decision-making is crucial and should form part of empirical research into the topic in the future.

References

Adams, R. 2021. Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews 46* (1–2), 176–197. https://doi.org/10.1080/03080188.2020.1840225.

Afrad, M. S. I. and Barau, A. A. 2019. Employability of Bachelor of Science. *Journal of Technical Education and Training 11*(2), 68–78. Scopus. https://doi.org/10.30880/jtet.2019.11.02.007.

Anshari, M. and Hamdan, M. 2022. Understanding knowledge management and upskilling in Fourth Industrial Revolution: Transformational shift and SECI model. *VINE Journal of Information and Knowledge Management Systems*. https://doi.org/10.1108/VJIKMS-09-2021-0203.

Badat, S. 2010. The challenges of transformation in higher education and training institutions in South Africa. *Development Bank of Southern Africa 8*(1),1-37.

Bearman, M., Ryan, J. and Ajjawi, R. 2023. Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education*, *86*(2), 369–385. https://doi.org/10.1007/s10734-022-00937-2

Boughey, C. and McKenna, S. 2021. *Understanding Higher Education: Alternative Perspectives*. African Minds: Cape Town.

Chaka, C. 2022. Fourth industrial revolution—A review of applications, prospects, and challenges for artificial intelligence, robotics and blockchain in higher education. *Research and Practice in Technology Enhanced Learning 18*, 002. https://doi.org/10.58459/rptel.2023.18002

Dalalah, D. and Dalalah, O. 2023. The false positives and false negatives of generative AI detection tools in education and academic research: The case of ChatGPT. *International Journal of Management Education 21*(2). https://doi.org/10.1016/j.ijme.2023.100822

David, L. O., Nwulu, N. I., Aigbavboa, C. O. and Adepoju, O. O. 2022. Integrating fourth industrial revolution (4IR) technologies into the water, energy and food nexus for sustainable security: A bibliometric analysis. *Journal of Cleaner Production*, *363*, 132522. https://doi.org/10.1016/j.jclepro.2022.132522

Davids, N. 2016. On extending the truncated parameters of transformation in higher education in South Africa into a language of democratic engagement and justice. *Transformation in Higher Education*, *1*(1). https://doi.org/10.4102/the.v1i1.7

Durso, S. and Arruda, E. 2022. Artificial intelligence in distance education: a systematic literature review of Brazilian studies. *Problems of Education in the 21st Century 80*(5), 679–692. https://doi.org/10.33225/pec/22.80.679.

European Commission. Directorate General for Communications Networks, Content and Technology. and High-Level Expert Group on Artificial Intelligence. 2019. *Policy and Investment Recommendations for Trustworthy AI.* Publications Office. https://data.europa.eu/doi/10.2759/465913.

Formunyam, K. G. 2020. De-territorialising to re-territorialising the curriculum discourse in African higher education in the era of the Fourth Industrial Revolution. *International Journal of Higher Education 9*(4), 27. https://doi.org/10.5430/ijhe.v9n4p27.

Gardner, A. 2022. Responsibility, recourse, and redress: a focus on the three Rs of AI ethics. *IEEE Technology and Society Magazine*, *41*(2), 84–89. https://doi.org/10.1109/MTS.2022.3173342.

Hagerty, A., and Rubinov, I. 2019. *Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence*. https://doi.org/10.48550/ARXIV.1907.07892.

Huang, C., Samek, T. and Shiri, A. 2021. AI and Ethics: Ethical and Educational Perspectives for LIS. *Journal of Education for Library and Information Science 62*(4), 351–365. https://doi.org/10.3138/jelis-62-4-2020-0106.

Hussein, D. M. E.-D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University - Engineering Sciences*, *30*(4), 330–338. https://doi.org/10.1016/j.jksues.2016.04.002.

Intahchomphoo, C. and Gundersen, O. E. 2020. Artificial Intelligence and Race: A Systematic Review. *Legal Information Management 20*(2), 74–84. https://doi.org/10.1017/S1472669620000183.

Kayembe, C. and Nel, D. 2019. Challenges and opportunities for education in the Fourth Industrial Revolution. *African Journal of Public Affairs* 11(3): 79-94.

Kiemde, S. M. A. and Kora, A. D. 2022. Towards an ethics of AI in Africa: Rule of education. *AI and Ethics 2*(1), 35–40. https://doi.org/10.1007/s43681-021-00106-8.

Maphosa, V. and Maphosa, M. 2020. E-waste management in Sub-Saharan Africa: A systematic literature review. *Cogent Business and Management*, *7*(1), 1814503.

Marinoni, G., Van't Land, H. and Jensen, T. 2020) The impact of Covid-19 on higher education around the world. *IAU Global Survey Report*, 23(1), 1-17.

Mhlanga, D. and Moloi, T. 2020. Covid-19 and the Digital Transformation of Education: What Are We Learning on 4IR in South Africa? *Education Sciences 10*(7), 180. https://doi.org/10.3390/educsci10070180

Moosa, R. 2018. World university rankings: reflections on teaching and learning as the cinderella function in the South African higher education system. *African Journal of Business Ethics 12*(1). https://doi.org/10.15249/12-1-165.

Nawi, A., Yaakob, M., Ren, C., Khamis, N. and Tamuri, A. 2021. A preliminary survey of Muslim experts' views on artificial intelligence. *Islamiyyat-The International Journal of Islamic Studies 43*(2), 3–16. https://doi.org/10.17576/islamiyyat-2021-4302-01.

Okewu, E., Adewole, P., Misra, S., Maskeliunas, R. and Damasevicius, R. 2021. Artificial Neural Networks for Educational Data Mining in Higher Education: A Systematic Literature Review. *Applied Artificial Intelligence*, *35*(13), 983–1021. https://doi.org/10.1080/08839514.2021.1922847.

Okonkwo, C. W. and Ade-Ibijola, A. 2021. Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence 2*, 100033. https://doi.org/10.1016/j.caeai.2021.100033.

Onaolapo, O. and Onifade. T. 2020. *Teaching and Learning in the Cloud: Prospects and Challenges of Artificial Intelligence for Education in Africa*. 1–9.

Otto, F., Kling, N., Schumann, C. A. and Tittmann, C. 2023. Conceptual approach to an AI-based adaptive study support system for individualised higher education. *International Journal of Advanced Corporate Learning (iJAC)*, *16*(2), 69–80. https://doi.org/10.3991/ijac.v16i2.35699.

Popenici, S. A. and Kerr, S. 2017. Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1), 22.

Sallstrom, L., Morris, O. and Mehta, H. 2019. Artificial intelligence in Africa's healthcare: ethical considerations. *ORF Issue Brief*, *312*, 1-11.

Slimi, Z. and Carballido, B. 2023. Navigating the Ethical Challenges of Artificial Intelligence in Higher Education: An Analysis of Seven Global AI Ethics Policies. *TEM Journal-Technology Education Management Informatics 12*(2), 590–602. https://doi.org/10.18421/TEM122-02.

Teng, Y., Zhang, J. and Sun, T. 2023. Data-driven decision-making model based on artificial intelligence in higher education system of colleges and universities. *Expert Systems 40*(4), e12820. https://doi.org/10.1111/exsy.12820.

Toda, A., Palomino, P., Rodrigues, L., Klock, A., Pereira, F., Borges, S., Gasparini, I., Teixeira, E., Isotani, S. and Cristea, A. 2022. Gamification Through the Looking Glass—Perceived Biases and Ethical Concerns of Brazilian Teachers. In M. Rodrigo, N. Matsuda, A. Cristea, and V. Dimitrova (eds). *Universidade de Sao Paulo* (WOS:000877459600047; Vol. 13356, pp. 259–262). https://doi.org/10.1007/978-3-031-11647-6_47.

Tundrea, E. 2020) *Artificial Intelligence in Higher Education: Challenges and Opportunities*. 2041–2049. https://doi.org/10.21125/inted.2020.0644.

UNESCO. (n.d.). Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development. *Working Papers Series*, *7*, 1–48.

Wamba, S. F., Gumbo, S., Twinomurinzi, H., Bwalya, K. and Mpinganjira, M. 2023. Digital transformation under Covid-19: A Bibliometric Study and future research agenda. *Centeris – International Conference on Enterprise Information Systems / ProjMAN – International Conference on Project Management / HCist – International Conference on Health and Social Care Information Systems and Technologies 2022*, *219*, 271–278. https://doi.org/10.1016/j.procs.2023.01.290.

Wankhade, M., Rao, A. C. S. and Kulkarni, C. 2022. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review 55*(7), 5731–5780. https://doi.org/10.1007/s10462-022-10144-1.

Williamson, B., Bayne, S. and Shay, S. (2020). The datafication of teaching in Higher Education: Critical issues and perspectives. *Teaching in Higher Education 25*(4), 351–365. https://doi.org/10.1080/13562517.2020.1748811.

Zhou, C. 2023. Integration of modern technologies in higher education on the example of artificial intelligence use. *Education and Information Technologies 28*(4), 3893–3910. https://doi.org/10.1007/s10639-022-11309-9.

Appendix 1: Search strings for the literature review

Search strings generated on the 12th of August 2023

Explore the current and potential applications of AI in education and its ability to address specific issues related to transformation, inclusivity and equity in higher education..

"South Africa education system" AND "educational challenges"

"Educational disparities" AND "transformation" AND "equity" AND "South Africa"

"Access to education" AND "educational inequalities" AND "South Africa"

Analyse the ethical considerations and potential biases associated with the implementation of AI in education.

"AI in education" AND "language barriers" AND "teacher shortages" AND "remote learning"

"Digital divide" AND "AI solutions" AND "marginalized communities" AND "South Africa"

"AI in special education" AND "equity" AND "South Africa"

Develop guidelines and recommendations for policymakers, educational institutions, and other stakeholders

"AI ethics" AND "educational technology ethics"

"Algorithmic fairness" AND "privacy in education"

"Transparency in AI algorithms" AND "ethical AI implementation" AND "education"

"AI policy guidelines" AND "equitable AI adoption" AND "South Africa"

"Strategies for AI implementation" AND "education policy recommendations"

"Future of AI in education" AND "AI trends" AND "equity-focused AI"

Navigating the Generative Artificial Intelligence Revolution in Higher Education: A Kotter Management Model Approach

F. Waghid
Cape Peninsula University of Technology
WaghidF@cput.ac.za

Abstract

Opportunities and challenges exist in adopting generative artificial intelligence (AI) in higher education. This article examines how adopting generative artificial intelligence (AI) in higher education can be navigated using John Kotter's Eight-Step Change Management approach. Generative AI holds much potential to personalise education, augment operational efficiencies, and facilitate teaching and learning, which is the impetus behind the emerging necessity to adopt generative AI. However, the erosion of human interaction, academic integrity, and ethics should be considered. The article explores how the Kotter approach can be used in the context of higher education to promote the adoption of generative AI. Kotter's steps, which include introducing new practices into the institutional culture, draw on a sense of urgency. Ensuring that technological advancements are congruent with educational outcomes and ethical norms, the article illustrates that Kotter's approach serves as a guide to circumnavigate the complexities of generative AI. In this way, higher education institutions can enhance the quality and accessibility of education provision in the age of AI by employing change management, drawing on collaborative engagement while simultaneously capitalising on the advantages of AI and confronting its challenges.

Keywords: Academic integrity, artificial intelligence, change management, higher education, university of technology

Introduction

In recent years, higher education has had to deal with several crises. These crises have implications for students and educators in higher education institutions around South Africa. For instance, student protests over increases in tuition fees brought attention to the escalating problems with students' access to and affordability of higher education. Consequently, higher education institutions have to re-evaluate their revenue generation models and operational efficiencies because the South African government has, over the year decreased spending on education, which has further exacerbated these issues. Additionally, the Covid-19 pandemic disrupted long-standing traditional pedagogical approaches at some higher education institutions, forcing a rapid transition to online learning, irrespective of whether it was effective or not. Even before these crises, there was a need to address the decolonisation of university curricula in Africa. During this time, students engaged in protests calling for more inclusive pedagogical practices that acknowledged historical biases and the students' voices. Despite these challenges, higher education has navigated them, drawing on the urgency of these crises. Each crisis played a role in raising urgency levels, culminating in the development of thoughtful responses and structural changes that have confirmed the resilience and relevance of higher education.

The emergence of generative artificial intelligence (AI) presents a response to prospective crises for some long-standing practices in higher education. Similarly, urgency levels are being raised to bring about change so that the potential for generative AI to transform the higher education sector exists. Some benefits of adopting generative AI include the potential to enhance the quality of teaching and learning, improve student attainment of outcomes, foster innovation, and automate administration. Generative AI can assist in updating curricula, offering opportunities for developing innovative assignments, and may assist in creating more authentic learning experiences. It may also support the development of more frequent formative assessments, including low-stakes quizzes, which not only reduce cheating, but also offer more constructive opportunities for student feedback. However, incorporating these assessments and feedback will necessitate reconsidering class sizes, teaching loads, and grading support availability (Yin, 2024). Some educationists see AI-generated content in higher education as inevitable, given its potential to significantly improve productivity and augment learning.

Student perceptions of generative AI have, for the most part, been positive. Students consider it a helpful tool for augmenting learning experiences. For instance, in language learning, generative AI tools have helped provide grammar assistance, idea generation, and communication support in the target language. Chan and Hu (2023) also highlight that using generative AI-based chatbots for learning support has improved students' learning achievement, self-efficacy, and enthusiasm. Furthermore, students recognise the impact of generative AI on their disciplines and future careers. This has led to calls for integrating generative AI into university curricula as an essential component (Chan and Hu, 2023).

Despite these perceived benefits of adopting generative AI, there are notable challenges and concerns. These relate to academic integrity, the need to critically interrogate the ethical use of the technology, and the challenge of ensuring that the adoption of generative AI tools does not replace human interaction (Chan and Hu, 2023). The potential of AI to restrict human interaction, pose data leakage risks, lack emotional connection, breach ethics, and affect job opportunities has been highlighted as areas of concern by students (Chan and Hu, 2023).

Navigating the generative AI landscape in higher education through thoughtful consideration of its implications on teaching and learning practices is essential. A change management model, John Kotter's Eight-Step Change Management Model, is used to assist in the navigation, as it is widely recognised for its effectiveness in implementing complex organisational change initiatives. The model specifically references the role of urgency in an organisation to bring about change. In the context of generative AI for higher education, the urgency exists to ensure higher education can capitalise on how it can augment teaching and learning practices and address challenges raised earlier. This model was developed by observing leaders and organisations attempting to transform their practices. Kotter identified eight steps to guide organisations through meaningful and effective change (Kotter 2014).

Kotter's model can offer a strategic framework for navigating the challenges and opportunities of integrating generative AI into higher education (Kotter 2014). Integrating generative AI into higher education entails more than simply adopting a novel technology. It also requires cultural and procedural transformations to improve the quality of teaching and learning. Consequently, the Kotter model may be attuned to the needs of higher education attempting to navigate the generative AI landscape successfully. Given the rapid advancements and disruptions caused by generative AI, it is argued that higher education institutions should adopt a systematic approach outlined by the Kotter model. This article examines the potential of John Kotter's Eight-Step Change Management Model as a strategic framework for effectively managing the benefits and challenges associated with adopting generative AI. Furthermore, to support the ongoing discussion, the article provides a case study that exemplifies how Kotter's Eight-Step Change Management Model is applied to adopt generative AI within a higher education institution.

The generative AI landscape in higher education

Generative AI can create or produce text, images, videos, music, and software code. This functionality not only holds the potential to transform the processes of educational content creation, but also gives rise to significant concerns regarding access to the AI technology, and academic integrity. Today, students can use generative AI to generate content and answers to assessments, for instance, which has sparked specifically vigorous debates regarding academic integrity. This has resulted in higher education institutions developing new policies and procedures to support the ethical use of generative AI.

Further, there are concerns regarding copyright infringements and whether the data generated is trustworthy. The balanced integration of generative AI technologies into education is crucial to optimise its benefits for personalised learning and creative expression while addressing challenges such as the digital divide and content homogenisation, as highlighted by Miao and Holmes (2023).

The introduction of generative AI into the higher educational sector presents many potential benefits that could significantly enhance teaching, learning, and research. According to Miao and Holmes (2023), generative AI offers innovative approaches to education by automating basic levels of information processing. This automation can free educators and students to focus on higher-order thinking skills and personalised learning experiences.

Higher education institutions have the potential to enhance the comprehension of complicated course material and foster greater engagement among students by utilising generative AI to design interactive curricula that are personalised to meet the unique needs of each student. In addition, the ability of generative AI to generate new content can encourage students to investigate novel concepts and points of view, thereby stimulating innovation and creativity in the classroom. Miao and Holmes (2023) emphasise the importance of implementing generative AI with an emphasis on human needs. They suggest supporting policies that promote cultural and linguistic diversity, human agency, inclusivity, equity, and gender parity. By ensuring adherence to such supporting policies when integrating generative AI into the realm of education, proponents assert that educators can optimise the advantages of the technology while mitigating potential challenges (Miao and Holmes, 2023). Embracing such an approach not only enhances instruction and knowledge dissemination but also ensures ongoing relevance to a heterogeneous student cohort (Miao and Holmes, 2023). Generative AI can significantly benefit researchers, scholars, and educators, provided that appropriate policies are established, and human-centric values are given priority (Miao & Holmes, 2023).

As previously mentioned, although generative AI does provide advantages for the field of education, it also introduces a series of challenges that require thoughtful deliberation and management. As highlighted by Miao and Holmes (2023), the integration of generative AI into education presents several significant challenges that must be addressed to fully capitalise on its potential while also safeguarding against these concerns. For example, higher education institutions must implement adequate safeguards to protect data privacy (Miao and Holmes, 2023). The widening of the digital divide represents a further challenge. The adoption of generative AI technologies into higher education may give rise to new disparities that result from the uneven accessibility of generative AI platforms (Miao and Holmes, 2023). To confront this challenge, it is critical to guarantee that all students have access to generative AI platforms. The issues highlight the critical nature of implementing safeguards to avoid the unauthorised utilisation of copyrighted materials and the generation of manipulative or harmful content via generative AI (Miao and Holmes, 2023). Another potential challenge of generative AI is that it may diminish the diversity of viewpoints and opinions, thus restricting the inclusivity of the pedagogical process (Miao and Holmes, 2023). Ensuring that the design and utilisation of generative AI platforms foster inclusivity should be paramount.

Adopting generative AI in higher education presents various challenges, including regulatory, ethical, accessibility, and diversity concerns. According to Miao and Holmes (2023), effectively dealing with these challenges necessitates a careful and all-encompassing strategy that weighs the possible benefits of generative AI against the imperative of minimising its challenges. This way, generative AI adoption can adhere to ethical principles and promote fairness and inclusivity.

A brief overview of Kotter's model

In today's era, organisations, including higher education institutions, encounter economic volatility and disruption (Kotter 2014). This necessitates these organisations to adapt to these uncertainties by pursuing opportunities for expansion and developing cost-effective ways to enhance the quality of their offerings (Kotter 2014). As organisations such as higher education institutions navigate technological progress, such as the advent of generative AI, change is an unavoidable element (Self, Armenakis, and Schraeder 2007). There is a common contention that change is a continuous process since it never begins and never ends (Weick, Sutcliffe, and Obstfeld 2005). Much effort has been devoted to enabling organisations, such as higher education institutions, to navigate change effectively. Change management encompasses the procedures and undertakings that direct an organisation through change (Schneidt 2022). It details the approach by which stakeholders will be informed, motivated, and assisted in implementing the change initiative. Higher education institutions are intricate entities requiring ongoing transformation to align with vital societal and technological advances. According to Kotter and Schlesinger (1989), change often results from the introduction of perceived threats through changes and disruptions to the *status quo*.

In the context of this article, the advent of generative AI is considered a disruption to the existing traditional teaching and learning practices currently employed at higher education institutions. Moreover, they argue that change frequently generates organisational resistance, as stakeholders remain apprehensive of failure or uncertainty. As a result, stakeholders, including lecturers, might be apprehensive about the potential benefits of using generative AI. Kotter and Schlesinger (1989) suggest that an organisation's efficiency might be negatively affected by fears of failure and uncertainty. There exists widespread evidence discussing how the application of the Kotter model can facilitate change during economic instability and disruption. A revised iteration of Kotter's eight-step change model was introduced in 2014 (Kotter 2014). The process steps are executed continuously and concurrently in the enhanced mode, as opposed to sequentially as in the original model.

The eight-step Kotter model comprises the following steps:

- Creating and maintaining a sense of urgency through the identification and deliberation of prevailing crises, potential crises, and the prospects to improve an organisation;
- Forming a guiding coalition with identified managers and leaders within the organisation where the coalition will have the impetus to drive change within the organisation;
- Creating a shared vision and developing strategies to ensure the realisation of such a vision;
- Empowering action by identifying impediments that may constrain the change process. These may include the systems and structures existing within the organisation;
- Planning changes that have the potential to generate discernible improvements representing short-term wins;

- Ensuring credibility to change structures, policies, and systems further; and
- Integrating new approaches into the newly established organisational culture.

Kotter and Schlesinger (2014) define the phase in which the *status quo* is disrupted as steps one through four. The changes are introduced and anchored in steps five through eight. The decision to utilise Kotter's model for change management in integrating AI into higher education is supported by many crucial steps in Kotter's model that are well-suited to the unique contexts of higher education settings. Kotter's model starts with creating a sense of urgency, an essential step for higher education institutions confronted with rapid technological advancements such as the advent of generative AI. This initial stage, which is critical for overcoming complacency and resistance to change, might not be given priority in other change management models (Kotter, 2012). Al-Haddad and Kotnour (2015) suggest that comparing various change management models and emphasising the significance of urgency supports Kotter's emphasis on urgency.

In contrast to models that emphasise hierarchical approaches to support change, Kotter's model proposes the formation of a guiding coalition comprising all stakeholders (Kotter, 2012). In higher education, students, faculty, administrators, and IT personnel are included, which ensures a multitude of viewpoints, and cultivates an environment conducive to collaboration. Weiner (2009) emphasises the significance of stakeholder engagement in healthcare reform initiatives, a sector that shares similar complexities with higher education. Kotter (2012) highlights the significance of formulating a strategic vision and initiatives and implementing efficient communication strategies. Effective communication is critical in higher education environments, wherein heterogeneous groups must comprehend the advantages and goals associated with proposed changes (Gilley, McMillan, and Gilley, 2009). Similarly, for long-term initiatives such as AI integration, Kotter's steps, which include removing obstacles to action and generating short-term victories, are essential for sustaining momentum and interest (Kotter, 2012). Adopting a proactive and optimistic approach is especially beneficial in higher education contexts, where concrete evidence of accomplishments can effectively maintain backing and financial investment.

Kotter (2012) asserts that the final stage, institutionalising new approaches, entails integrating changes into the organisational culture. Maintaining a long-term perspective is crucial in higher education contexts, for instance, where sustainably integrating generative AI necessitates modifications to established policies, practices, and norms. This contrasts with models that prioritise the operational aspects of change over its cultural assimilation. The significance of cultural change in facilitating successful organisational transformation is supported by Schein (1992). Therefore, Kotter's eight-stage model may enable institutions to have a coherent and sequential structure that can serve as a navigational aid during the intricate undertaking of AI integration. This sequential approach facilitates the management of the complex elements of change, as opposed to the emergent or flexible change models emphasised by Bridges (2009). The structured approach can provide organisations with explicit direction and milestones, which is especially advantageous in higher education, where coordinating numerous stakeholders and processes is essential.

Applying Kotter's Model to the integration of AI in higher education

Creating a sense of urgency

The first step of the Kotter Model involves diminishing satisfaction with he *status quo* within an organisation and, consequently, raising urgency levels. In the context of generative AI and higher education, this can relate to higher education yearning to become competitive and abreast with the developments in generative AI. For instance, the World Economic Forum (2020) underscores the significance of adapting pedagogical approaches and higher education curricula to equip students for changing labour markets shaped by rapid generative AI within the education sector. Students increasingly expect their educational experiences to be enriched with the latest technologies, such as artificial intelligence, according to the Educause Centre for Analysis and Research (2020). According to McKinsey and Company (2020), the schooling sector has also experienced positive effects of adopting generative AI, specifically regarding administrative efficiencies and personalised learning.

Responsible adoption of generative AI in higher education is imperative to effectively deal with challenges such as the digital divide, ethical concerns, and data privacy issues. Floridi *et al*. (2018) emphasise that incorporating AI technologies requires a delicate equilibrium between safeguarding individual rights and promoting innovation. They advocate for establishing ethical frameworks and policies that govern the utilisation of AI in an educational capacity, guaranteeing that it does not compromise privacy or worsen inequalities while simultaneously augmenting educational opportunities. By utilising such an approach, one can effectively navigate the complexities of AI adoption, capitalise on its benefits, and minimise its risks. Highlighting to higher education stakeholders the need to adopt generative AI to stay competitive along with observing the ethical imperatives may raise urgency levels that can serve as a driving force to bring about change.

Building a guiding coalition

According to Kotter (2012), a strong leadership team drives change. This team must have sufficient influence to inspire participation and buy-in from the organisation. The effective implementation of generative AI in higher education requires establishing a leadership team of all relevant stakeholders. Including a diversity of perspectives while attending to the interests and apprehensions can ensure the legitimacy of the change process. For instance, Bryson, Crosby, and Stone (2015) emphasise the significance of inclusivity in higher education environments. He suggests that including representatives from across the institution can aid in addressing the ethical, technical, and pedagogical implications of adopting generative AI. This is further supported by Kezar and Gehrke (2015), who argue that successful change efforts in higher education require creating structures and cultures that support shared goals. This can be realised through the provision of opportunities for collaborative decision-making that can empower all stakeholders and foster a sense of shared ownership of the change process. This can occur in meetings, workshops, and forums where individuals can discuss progress, challenges, and insights related to generative AI adoption.

Forming a strategic vision

A vision, according to Kotter (2014), must be compelling, communicable, attainable, desirable, and focused. A future state substantially distinct from the current state of affairs ought to be clarified. Concerning the adoption of generative AI, this vision may encompass the aim of enhancing academic outcomes via technological means, equipping students for a future characterised by the ubiquity of AI. Ertmer and Newby (2016) argue that technological innovation should be harmonised with pedagogical practices. Consequently, generative AI could strengthen student-centredness by promoting personalised learning experiences that generative AI enables. This can be extended to include generative AI into curriculum development to create adaptive learning environments, leveraging generative AI to augment research capabilities, or integrating AI tools to personalise student learning. Luckin, Holmes, Griffiths, and Forcier (2016) argue how generative AI can be integrated into educational contexts to enable more personalised learning. They suggest that utilising analytics data to drive AI use can create opportunities to track student progress and adapt curricula to individual learning. By creating a shared vision that is aspiring and aligned with the institution's goals, higher education institutions can effectively navigate the complexities of AI adoption and harness its potential to transform pedagogical practices.

Enlisting a volunteer army

This step in the Kotter Model is essential for attaining the broad-based support necessary for making substantial organisational changes. This step goes beyond forming a guiding coalition. Kotter (2014) suggests that for change to be successful and sustainable, it should be encompassed by many individuals. In the context of higher education, this may equate to engaging not only educators but also students and staff developers. Henderson, Beach, and Finkelstein (2011) suggest that change initiatives in the higher education context are more successful when there is input from various stakeholders. This approach can encourage the community to unite around the change initiative and, in the context of generative AI, sustain its momentum for widespread adoption. This could stem from smaller initiatives to develop pilot projects or research into how AI can be leveraged in their unique contexts.

Removing barriers

Kotter (2014) highlights numerous barriers that can impede change initiatives. This includes an unsupportive organisational structure, lack of training, and restrictive policies. In higher education, these barriers can manifest as a lack of generative AI access and resistance from faculty or staff who are sceptical of AI's impact on teaching and learning or research. Furthermore, there may also be existing policies that need to be attuned to the developments in generative AI. George Westerman, Didier Bonnet, and Andrew McAfee (2014) suggest that where organisations have overcome resistance to digital transformation, there was the fostering of digital culture, reskilling the workforce, and altering policies to support change (Westerman *et al*., 2014). In the context of generative AI, this might involve training staff on using new tools, updating data governance policies to ensure ethical AI usage, and reforming curriculum design to integrate AI into teaching and learning practices.

By removing these barriers to the change initiative, higher education institutions may potentially pave the way for generative AI tools to augment teaching and learning practices, streamline administrative tasks, and foster innovative research. This includes not only technological adaptation but also necessitates cultural shifts and policy revisions to support the change initiative.

Generating short-term wins

Generating short-term wins is an essential phase in Kotter's change model. It builds momentum by providing evidence of the benefits of a change initiative. This is especially important in higher education institutions adopting generative AI for teaching and learning, where the immediate benefits may not be instantly observable. Kotter (2014) explains that short-term wins create positive feedback loops that reinforce the change effort, sustain the commitment of involved parties, and can help to quell cynics and resistors. In higher education, such wins could be highlighted by the successful adoption of AI in a particular course or administrative function, demonstrating the practical benefits and positive outcomes associated with its adoption. Bates (2015) supports this approach, emphasising the importance of using technology to enhance teaching and learning in ways that can be quickly demonstrated and communicated to stakeholders. For instance, an AI-powered programme might improve student engagement or reduce the time taken to grade assessments, which could serve as early indicators of success. Change initiatives need to plan for these wins, communicate them effectively, and use them as stepping stones toward the larger goals of the change initiative (Kotter, 2014).

Sustaining acceleration

The penultimate step of the Kotter model entails embedding small-scale changes into the organisational culture. The credibility that short-term successes lend to the change process enables an organisation to scale up the implemented changes. This may include broadening the implementation of generative AI beyond individual use. Therefore, this consolidation phase ensures the sustained benefits of change. During this phase, leaders should focus on communicating and reinforcing the organisation's vision for change, as Kotter (2012) notes. The organisation will then understand how each member contributes to this vision for change. In the context of generative AI adoption, this involves using the gains from early AI initiatives to maintain momentum for ongoing change. Thus, generative AI has the potential to become an integral part of a higher education institution’s structure and culture.

Instituting change

The final step of the Kotter model is to embed proposed changes as part of the organisational culture. In the context of generative AI and higher education this could relate to six related activities. Firstly, revising institutional policies is necessary to support ongoing generative AI initiatives. These policies may relate to data governance, privacy, and intellectual property. Secondly, developing resources for best practices on the use of generative AI. This might involve creating guidelines for generative AI applications in course development, student interaction, and data analysis. Thirdly, continuous advocacy from leadership is essential to promote the use of generative AI should it be aligned with the strategic goals of the institution. Fourthly, providing professional development opportunities for staff is key to equipping them with the knowledge required to adopt generative AI effectively. Fifthly, regular assessments of the impact of generative AI initiatives should also be carried out. And finally, fostering an institutional culture that values innovation, risk-taking, and adaptability to change is fundamental. This could include celebrating achievements in generative AI and recognising individuals or departments that successfully integrate generative AI into their operations.

Implementing the Kotter Model to Support Generative AI Adoption at a University of Technology: A Case Study

Integrating generative AI into higher education presents a complex challenge that requires a nuanced and strategic approach. This case study explores how a university of technology is navigating the AI landscape utilising the Kotter's 8-Step Change Management Model as a framework. As the generative AI landscape faces rapid advancements and changing educational needs, the university recognises the potential of generative AI for enhancing teaching, learning, and research. However, it also acknowledges the associated risks, including concerns over academic integrity, ethical use, and the potential reduction impact on its pedagogical activities.

Creating a sense of urgency

Many higher education institutions in South Africa acknowledge the pressure from technological advancements and the expectations of digital-native students. In response, the higher education institutions have begun compiling guidelines and policies to navigate the generative AI landscape for teaching, learning, and research in late 2022. In much of 2023, the university of technology in question was under much pressure from staff and students to formulate a response to the disruption generative AI was causing in the institution. This is related primarily to academic integrity concerns and the understudy of how technology could augment the institution's practices. Consequently, the urgency to formulate a response for the institution was present.

Building a guiding coalition

A diverse task team was established within the institution. It comprised various stakeholders within the institution, from the deputy vice-chancellor, directors, heads of department, teaching and learning coordinators, lecturers and students. This task team was tasked with steering the generative AI adoption process, ensuring a comprehensive approach considering technical, pedagogical, and ethical dimensions.

Forming a strategic vision

The guiding coalition developed a clear and compelling vision for generative AI adoption. There was much deliberation pertaining to the nature of the institutional response. There were discussions as to whether the institution needed to formulate a new policy or a guide. Ultimately, the task team agreed that the existing policies and frameworks within the institution could adequately address the disruption of generative AI, especially regarding academic integrity and ethical considerations. For instance, the institution's plagiarism policy, although not specifically mentioning generative AI, described plagiarism as using electronic sources without referencing as plagiarism. It was, therefore, agreed that the existing policies within the institution were robust enough to deal with issues about academic integrity, to mention but one example.

The strategic vision was, therefore, to formulate guidelines that speak to the institutional vision, advocating for the need to change traditional practices to embrace digital technology to augment teaching, learning and research.

Enlisting a volunteer army

To garner widespread support, the task team actively engaged the broader university community. This was done through teaching and learning coordinators and academic staff developers, who were part of the task team offering workshops for faculties highlighting the functionality of generative AI to support teaching and learning. During these workshops, lecturing staff were encouraged to contribute ideas and express concerns, creating a collaborative atmosphere that empowered individuals to participate in the change process, manifest through the formulation of generative AI guidelines.

Removing barriers

Kotter (2012) proposes that four organisational impediments may hinder employees from partaking in a change process, namely the structures, skills, systems, and supervisors. In the context of this change initiative, access to generative AI platforms, resistance from faculty, and concerns over academic integrity were identified as impediments. Furthermore, the task team identified that more than merely formulating generative AI guidelines to support the change initiative would be required to bring about change. Coupled with the generative AI guide was the need to develop a staff development programme to assist lecturing staff in enacting various aspects covered in the generative AI guidelin*es.*

Generating short-term wins

The short-term win in this change initiative was formulating generative AI guidelines. These were designed to assist staff in ethically and practically incorporating generative AI tools into their work. They were developed using prompts specifically designed to reflect the unique institutional culture and contributions from all staff. The guidelines highlight generative AI's potential effects on teaching methods, curriculum development, and academic honesty, promoting a balanced approach to education in the AI era. Furthermore, the guidelines encourage ongoing dialogue about generative

AI's benefits and challenges, aiming to equip staff with knowledge and strategies for integrating generative AI. Additionally, the guidelines reference a framework of relevant policies to be considered alongside the guidelines, emphasising the importance of integrating these guidelines with existing policy structures.

The guidelines were then workshopped at numerous institutional forums, including workshops with individual departments within faculties. Throughout these sessions, staff were encouraged to provide recommendations for improvement. The workshop feedback forms indicated were overwhelmingly positive. In this way, the guidelines became an institutional initiative rather than a task team initiative.

Sustaining acceleration

Leveraging these initial successes, the university expanded its AI initiatives, with academic staff development units requested by numerous departments to support lecturers in effectively navigating generative AI integration through workshops covering the generative AI guidelines.

Instituting change

To solidify the changes, the university is revising its policies and practices to reference the generative AI guidelines and support workshops. Professional development and ethical guidelines ensured that the use of AI aligned with the university's values and mission. Thus, adhering to Kotter's approach in facilitating a generative AI strategy for a university of technology, it seems as if the leaning is towards the cultivation of democratic action. This implies that people work together, engage one another's diverse thoughts, and encourage one another to urgently respond to emerging complexities and challenges *vis-à-vis* generative AI implementation.

What democratic engagement also accentuates is a need for ways of engaging in higher education discourse in the context of continuous change. Based on democratic engagement the implementation of generative AI would be subjected to engaged action whereby people consider emerging possibilities and dissonant perspectives in the manifestation of AI. In short, when critique and dissonance hold sway in the cultivation of generative AI, higher education institutions invariably become more adept at implementing widespread change.

Conclusion

In conclusion, the adoption of generative AI in higher education, drawing on Kotter's Eight-Step Change Management Model, offers an opportunity for augmenting educational practices and mitigating many challenges. This article has highlighted the potential of generative AI to enhance the efficacy and accessibility of pedagogical practices. It has also highlighted the need to consider ethical and academic integrity concerns. The Kotter Model emphasises the significance of fostering an inclusive environment, encouraging the participation of all stakeholders, establishing coalitions, and generating urgency as important variables in a change management effort.

The case study of the university of technology has exemplified how Kotter's model can be effectively implemented to navigate the generative AI landscape. This illustrates the model's efficacy in cultivating an environment that encourages ongoing discourse and changes to emerging technologies. Through the establishment of clear guidelines, facilitating widespread engagement, and promoting ethical practices, the university sets a good example for how institutions can potentially manage the transition towards a more technologically integrated educational context which incorporates generative AI. Given the rapid advancements in digital technology and ever-evolving educational demands, higher education institutions may find the strategic implementation of Kotter's model valuable. Its use may ensure higher education remains relevant and malleable to the demands of society by furnishing students with the expertise and competencies required to prosper in an ever more digital environment.

References

AI Index Steering Committee. 2021. *The AI Index 2021 Annual Report*. Stanford University, Human-Centred AI Institute. https://aiindex.stanford.edu/ai-index-report-2021/.

Al-Haddad, S. and Kotnour, T. 2015. integrating the organisational change literature: a model for successful change. *Journal of Organizational Change Management* 28(2), pp.234-262.

Bates, A. W. 2015. *Teaching in a Digital Age: Guidelines for Designing Teaching and Learning*. Vancouver, BC: Tony Bates Associates Ltd.

Bridges, W. 2017. *Managing Transitions: Making the Most of Change*. 25th anniversary edition. Boston, MA: Da Capo Lifelong Books.

Bryson, J. M., Crosby, B. C. and Stone, M. M. 2015. Designing and implementing cross-sector collaborations: needed and challenging. *Public Administration Review*, 75(5), pp. 647-663.

Chan, C. K. Y. and Hu, W. 2023. Students voices on generative AI: perceptions, benefits, and challenges in higher education', arXiv preprint arXiv:2305.00290. https://doi.org/10.48550/arXiv.2305.00290

Ertmer, P. A. and Newby, T. J. 1993. Behaviourism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, 6(4), pp. 50-72.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... and Shaheed, A. 2018. AI4People—An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), pp. 689-707.

Gilley, A., Gilley, J. W. and McMillan, H. S. 2009. Organisational change: Motivation, communication, and leadership effectiveness. *Performance Improvement Quarterly*, 21(4), pp. 75-94. https://doi.org/10.1002/piq.20039

Henderson, C., Beach, A. and Finkelstein, N. 2011. Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching*, 48(8), pp. 952-984.

Kezar, A. and Gehrke, S. 2015. Communities of Transformation and Their Work Scaling STEM Reform. Los Angeles, CA: Pullias Centre for Higher Education, Rossier School of Education, University of Southern California. http://files.eric.ed.gov/fulltext/ED574632.pdf.

Kotter, J. P. 2012. *Leading Change*. Boston, MA: Harvard Business Review Press.

Kotter, J. 2014. *Accelerate (XLR8): Building Strategic Agility for a Faster Moving World*. Harvard Business Review Press.

Kotter, J. P. and Schlesinger, L. A. 1989. Choosing Strategies for Change, in Kotter, J. P. and Schlesinger, L. A. (eds) *Readings in Strategic Management*. London: Macmillan Education UK. pp. 294-306. https://doi.org/10.1007/978-1-349-20317-8_21

Luckin, R., Holmes, W., Griffiths, M. and Forcier, L. B. 2016. *Intelligence Unleashed: An argument for AI in Education*. London: Pearson Education.

Miao, F. and Holmes, W. 2023. Guidance for generative AI in education and research. UNESCO. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research.

McKinsey and Company. 2020. Global AI Survey 2020. [online] Available at: https://ceros.mckinsey.com/global-ai-survey-2020-a-desktop

Schein, E. H. 1992. *Organisational Culture and Leadership*. 2nd ed. San Francisco, CA: Jossey-Bass.

Self, D. R., Armenakis, A. A. and Schraeder, M. 2007. Organisational content, process, and context: A simultaneous analysis of employee reactions. *Journal of Change Management*, 7(2), pp. 211-229.

Weick, K. E., Sutcliffe, K. M. and Obstfeld, D. 2005. Organising and the process of sensemaking. *Organisation Science*, 16(4), pp. 409-421. https://doi.org/10.1287/orsc.1050.0133.

Weiner, B. J. 2009. A theory of organisational readiness for change. *Implementation Science*, 4, Article number: 67.

Westerman, G., Bonnet, D. and McAfee, A. 2014. *Leading Digital: Turning Technology into Business Transformation*. Boston, MA: Harvard Business Review Press.

World Economic Forum. 2020. *The Future of Jobs Report 2020*. Geneva, Switzerland: World Economic Forum. The Future of Jobs Report 2020.

Yin, W. J. 2024 Will our educational system keep pace with ai? a student's perspective on ai and learning. EDUCAUSE Review, 24 January. https://er.educause.edu/articles/2024/1/will-our-educational-system-keep-pace-with-ai-a-students-perspective-on-ai-and-learning.

Opportunities Presented by Artificial Intelligence to Teaching and Learning in Higher Education: A Review of Literature

L. V. Mulaudzi
University of the Free State
munyailv@ufs.ac.za

Abstract

As higher education continues to evolve in response to technological advancements, the use of artificial intelligence appears to be gaining traction in teaching and learning environments. This paper critically examines the growing significance of artificial intelligence in higher education. The paper is based on the review of relevant literature, guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses principles. It employs a thematic synthesis method to analyse and synthesise data, providing a comprehensive understanding of the opportunities presented by artificial intelligence in higher education. It highlights the potential for innovative teaching methods, personalised learning experiences, and improved student engagement; as well as potential to improve administrative efficiency. It is further argued that the use of artificial intelligence in teaching and learning promotes equity and inclusivity by addressing educational disparities. However, adoption of artificial intelligence in teaching and learning varies across institutions and disciplines.

Keywords: Artificial intelligence, higher education, pedagogical innovation, student engagement, administrative efficiency

Introduction and background

Artificial Intelligence (AI) is rapidly transforming various aspects of society. In higher education, AI offers significant opportunities to revolutionise teaching and learning practices, enhance student engagement, and improve institutional efficiency. Before the era of AI, teaching and learning in higher education primarily relied on traditional methods within a structured framework with in-person lectures, physical textbooks, and limited interaction beyond the classroom (Diwaker *et al*., 2021; Popenici and Kerr, 2017; Yang *et al*., 2020). Lecturers delivered lectures to large groups of students, often following a standardised curriculum with little room for customisation (Diwaker *et al*., 2021; Popenici and Kerr, 2017; Yang *et al*., 2020). Assessments typically revolved around examinations, essays, and assignments, which evaluated students' ability to memorise and regurgitate information rather than fostering critical thinking or practical skills.

Pedagogically, the emphasis was on transmitting knowledge from instructor to student, with limited opportunities for active engagement or hands-on learning. Students were expected to absorb information passively, often without much opportunity for questioning or discussion (Kordrostami and Seitz, 2022). Classrooms were hierarchical, with lecturers holding authority over the flow of information and learning activities (Diwaker *et al*., 2021; Popenici and Kerr, 2017; Yang et al., 2020). Furthermore, the traditional higher education model lacked flexibility because it could not accommodate diverse learning styles and preferences. Students with different backgrounds, interests, or learning paces may have struggled to engage fully with the material or find relevance in their studies.

Administratively, processes such as registration, grading, and course scheduling were often manual and paper-based, time-consuming, leading to resource allocation and decision-making inefficiencies and delays (Zouhaier, 2023). Decision-making within institutions relies heavily on bureaucratic structures and institutional hierarchies, with limited input from students or faculty, leading to inefficiencies. The traditional higher education model prioritised knowledge transmission over active learning, standardised assessments over skill development, and bureaucratic processes over efficiency and innovation (Zouhaier, 2023). This model worked well for many years but became increasingly outdated as the educational landscape evolved, leading to the exploration of new approaches like AI to address its limitations and challenges (Zouhaier, 2023). Furthermore, the educational landscape lacked the personalised learning experiences, data-driven insights, and automation capabilities that AI now offers.

The integration of Artificial Intelligence (AI) technologies into higher education has emerged as a transformative force in recent years, producing new teaching and learning solutions globally (United Nations Educational, Scientific and Cultural Organization (UNESCO), 2023; Chu *et al*., 2022). AI encompasses a wide range of tools and applications, including machine learning algorithms, natural language processing, data analytics, and automation, which have the potential to revolutionise various aspects of teaching and learning within HE institutions (Roumate, 2023; Sadiku et al., 2022; Peng *et al*., 2021). As the demand for accessible and innovative education continues to grow, educators and institutions are increasingly exploring the possibilities presented by AI to enhance the quality and effectiveness of educational experiences (Sadiku *et al*., 2022; Arakpogun *et al*., 2021).

AI technologies promise personalised learning experiences, efficient administrative processes, data-driven decision-making, and enhanced student engagement (Khadimally, 2022; Cağataylı and Celebi, 2022). These advancements hold the potential to address longstanding challenges in higher education, such as improving retention rates, accommodating diverse learning styles, and adapting to rapidly changing educational landscapes (Crompton and Burke, 2023; Roumate, 2023; Khadimally, 2022; Sadiku *et al*., 2022). These technologies offer new vistas of opportunity and pose novel challenges for educators, administrators, and students alike (Akinwalere and Ivanov, 2022). Additionally, institutions grapple with administrative complexities and the need to harness data-driven insights to enhance decision-making and resource allocation (Elgendy *et al*., 2022). These multifaceted challenges necessitate a reimagining of the traditional higher education model.

This paper reviews relevant literature to explore the transformative potential of AI in higher education, focusing on its implications for pedagogy, student outcomes, administrative processes, and equity.

Problem statement, guiding questions, aim and objectives

Higher education is mostly characterised by long standing traditions including in teaching methodologies and in curricula that do not keep abreast with the technological developments (Eager and Brunton, 2023; O'Dea and O'Dea, 2023; Zouhaier, 2023). The result is that higher education has been producing graduates who are underprepared for the demands of the modern workforce, where AI proficiency is increasingly valued. Additionally, current teaching pedagogies often prioritise knowledge transmission over skill development, hindering students' ability to adapt to an AI-driven future (Eager and Brunton, 2023; O'Dea and O'Dea, 2023; Zouhaier, 2023). Higher education has been slower than the corporate sector in adopting and using AI (Jain and Jain, 2019). While AI has gained substantial attention and adoption across various sectors, its integration into higher education remains complex and multifaceted.

The paper was prompted by the question: *What opportunities are presented by AI for teaching and learning in higher education?* It is this question that that guided the review of relevant literature resulting the the compilation of this review paper whose aim is to explore opportunities presented by AI to teaching and learning in higher education. The specific objectives are to (a) systematically identify and analyse the various opportunities AI offers within the context of teaching and learning in higher education; and (b) evaluate the impact of AI on pedagogical practices, including its influence on student outcomes, engagement, and satisfaction.

Methodology

The review of literature was undertaken according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) principles of transparency, rigour, and a controlled review process. According to Page *et al*. (2021), a comprehensive review benefits greatly from the well-established methodology that the PRISMA guidelines offer. The PRISMA guidelines help with a rigorous, repeatable review process (Page *et al*., 2021). These actions involve establishing inclusion and exclusion criteria, creating a thorough search strategy, methodically picking research, and transparently reporting the results (Page *et al*., 2021). The qualitative thematic analysis was used for a comprehensive literature assessment on AI's prospects for teaching and learning in higher education to identify and synthesise salient themes, concepts, or patterns among the chosen studies.

Specific inclusion criteria were defined following the thematic analysis to ensure the rigour and focus of the review. Primary research articles were extracted. This data includes relevant quotes, passages, or findings that pertain to AI opportunities in higher education. The review encompasses studies published within a specific timeframe from 2019 to 2023 (5 years). This time frame enabled capturing recent developments while considering foundational research. Studies published in English or could be translated into English, given the common use of the languages in academic publications within the field of education, were the focus. The review prioritised primary research studies, including empirical research, case studies, reports, conference proceedings, and qualitative or quantitative investigations. Secondary literature was excluded.

The literature search strategy was designed to comprehensively identify relevant articles and research papers from various academic sources. A combination of database searching and reference tracking was employed, including PubMed, Education Resources Information Centre (ERIC), PsycINFO, IEEE Xplore, Scopus, Semantic Scholar, Kovsie Cat, Kovsie Scholar and Google Scholar.
A list of search terms and Boolean operators was used to refine the search. The Boolean search contained phrases relating to AI, HE, and learning appropriate to the research topic and questions.
The approach to screening and selecting studies follows a structured and transparent process guided by the PRISMA guidelines. The initial search yielded a substantial number of results from each database, reflecting the depth and breadth of research on AI in HE. The total number of initial search results across all databases was 35804. The screening and selection of studies followed a structured and transparent process. A two-stage screening approach was employed: stage 1: title and abstract screening and stage 2: full-text review.

Following the initial search, a two-stage screening was conducted to identify studies that met the inclusion criteria. The first stage involved screening titles and abstracts, resulting in the exclusion of irrelevant or non-conforming studies. The second stage consisted of a thorough full-text review of selected articles to ensure alignment with the research question and objectives. After screening and applying the inclusion criteria, this systematic literature review included 47 studies.

These studies represent a diverse body of research exploring the opportunities presented by AI in HE. The entire selection process was documented using a PRISMA flow diagram (Figure 1), which was used to illustrate the number of studies identified, screened, included, and excluded at each stage, providing a clear visual representation of the review process.

The relevant data points were extracted from each study, including study design, study type, AI key findings, and recommendations. The extracted data were synthesised thematically to identify patterns, trends, and key insights related to the opportunities presented by AI in higher education. A categorisation approach was adopted based on themes and key topics related to AI in higher education to systematically analyse and synthesise the selected studies.

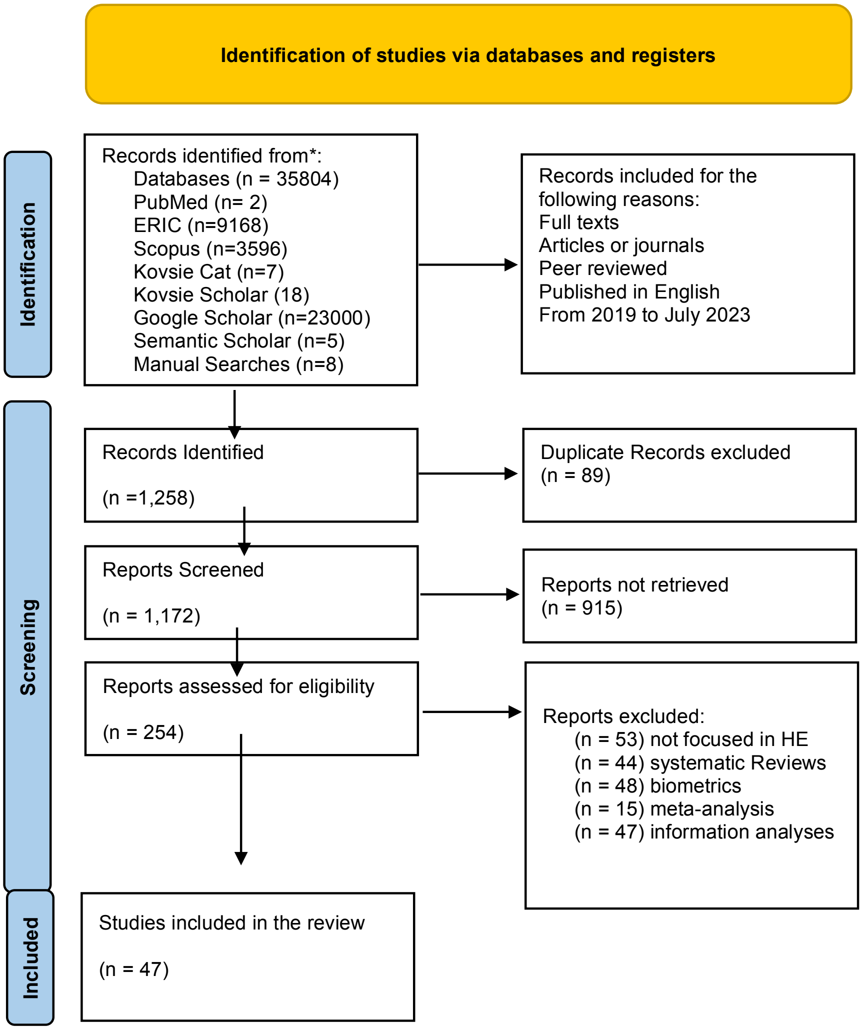

Figure 1: PRISMA Flow Diagram

Presentation of results

The results of the literature review are presented under the following headings pedagogical transformation, student engagement and learning Outcomes, administrative efficiency, resource allocation optimisation, data-driven decision-making, equity and inclusivity, adoption trends, faculty and student perspectives.

Pedagogical transformation

According to Abgaryan, Asatryan and Matevosyan (2023), introducing AI has caused important paradigm shifts in higher education, altering how educational pedagogies and academic endeavours are carried out. These advanced technologies have reshaped educators' methods and approaches and empowered students to participate in enriched, personalised, and adaptive learning experiences (Abgaryan *et al*., 2023; Mishra and Shinde, 2023; Vanichvasin, 2022). Numerous studies have highlighted that AI uses better learning styles and teaching methods in higher education than human lecturers, and that AI-driven technologies have enabled educators to explore and implement innovative teaching methods (Zouhaier, 2023; Turugare and Rudhumbu, 2020). These methods encompass a spectrum of pedagogical approaches that are discussed below.

Personalised learning, adaptive learning and interactive learning

AI-powered platforms can tailor learning experiences to individual student's needs, preferences, and learning styles (Abgaryan *et al*., 2023). Findings reveal that personalised learning pathways have significantly increased student engagement and achievement (Tapalova and Zhiyenbayeva, 2022; Vanichvasin, 2022). Personalised learning platforms use AI to adapt content, resources, and assessments to individual student needs, preferences, and progress (Mishra and Shinde, 2023). They boost engagement, improve learning outcomes, and cater to diverse learning styles, leading students to have ownership of their learning (Yu, 2023; Zouhaier, 2023).

Educators can leverage personalised learning platforms to differentiate instruction and provide targeted support to students with varying abilities (Mishra and Shinde, 2023). Yu (2023) revealed that AI algorithms adapt the difficulty and pace of instruction to match each student's proficiency level (Yu, 2023). Rodríguez, Guerrero-Roldán, Baneres and Karadeniz (2022) attest that using algorithms leads to more effective knowledge acquisition and retention of students. Furthermore, AI-driven interactive tools and simulations have revolutionised classroom dynamics, fostering active student participation and collaborative problem-solving (Mbodila and Ndebele, 2022).

Abgaryan *et al*. (2023) also reveal that AI facilitates personalised and adaptive student learning outcomes. Modern technology has made it possible to analyse massive amounts of data to identify individual learning needs, provide individualised feedback, and customise educational paths for pupils, improving academic achievements (Abgaryan *et al*., 2023; Mishra and Shinde, 2023). The adaptability and personalisation offered by AI have been linked to higher knowledge retention rates, with students demonstrating a deeper understanding of course material (Leoste *et al*., 2022; (Mishra and Shinde, 2023). AI interventions have effectively reduced dropout rates, particularly in challenging courses, by providing the necessary support and resources (Rodríguez *et al*., 2022; Shalini and Tewari, 2020).

Curriculum development

Studies suggest that AI technologies have facilitated the creation of dynamic, up-to-date curricula that align with rapidly evolving industry demands (Kuleto *et al*., 2021). Furthermore, AI-driven data analysis has enabled institutions to identify knowledge gaps and refine course content accordingly (Yu, 2023). Another study indicates that AI has simplified the curriculum by streamlining content creation, which generates lecture materials and automates the development of assessment items (Jiayu, 2023; Rawas, 2023). Almufarreh, Arshad, and Mohammed (2021) agree that AI tools such as Blackboard Ally assist in increasing the usability of their curriculum. Their finding further indicates that AI assists in reviewing course material to look for accessibility difficulties, including a ranking of the course's overall accessibility, a breakdown of the course material by kind of curriculum, and a list of all accessibility issues (Almufarreh *et al*., 2023).

It is also important to note that AI in curriculum development faces several challenges, including the lack of human elements, biases, and the dynamic nature of industries. AI can analyse data but may overlook nuanced aspects of learning, perpetuating biases and limiting inclusivity (Hutson and Ceballos, 2023; Rufrano and Yeung, 2023). To ensure AI-driven curricula remain relevant, strategies include continuous monitoring and evaluation, collaboration with industry partners, agile curriculum design, incorporating real-world projects and experiences, and data-driven decision-making (Hutson and Ceballos, 2023; Rufrano and Yeung, 2023). These measures help align curricula with industry needs, provide practical experience, and ensure students are prepared for the modern workforce. Addressing these limitations and implementing proactive strategies, AI-driven curricula can better prepare students for the modern workforce and contribute to their long-term success and employability.

Delivery of educational content

Accessibility issues are addressed by Jiayu (2023), Shalini and Tewari (2020), and Holmes *et al*. (2021) by highlighting the transformation of the delivery of educational content with AI-driven systems supporting various modalities such as Intelligent Tutoring Systems (ITS), Chatbots and Virtual Assistants (VA). These systems offer real-time feedback and guidance to students, improving their understanding and retention of course material (Abgaryan et al., 2023). Intelligent tutoring systems can customise their educational tactics to address certain learning gaps and provide additional practice or explanations as needed by learning the student's strengths and limitations (Abgaryan *et al*., 2023; Hui *et al*., 2021; Mishra and Shinde, 2023).

A study by Vanichvasin (2022) indicated that chatbots were an interesting, innovative, and fun teaching way. While a working policy paper by Pedro et al. (2019) revealed that VAs freed up teachers' time by doing typical duties, allowing them to concentrate on student mentoring and one-on-one interaction. Similarly, others such as Jiayu (2023), Gawande, Al Badi and Al Makharoumi (2020), and Holmes et al. (2021) attest that AI-powered chatbots and virtual assistants have enhanced communication between students and institutions, offering instant responses to queries and support.

Furthermore, Abgaryan *et al*. (2023) acknowledged that VAs play a role in virtual advising and tutoring, fortifying student engagement and retention. VAs, as revealed by Tapalova and Zhiyenbayeva (2022), provide students with 24/7 access.

The integration of AI into educational content delivery faces several challenges, including equitable access to AI technologies, teacher training and professional development, curriculum alignment and customisation, data privacy and security, ethical and bias considerations, and cost and sustainability (Owan *et al*., 2023; Penicini, 2023). Access to technology and infrastructure is crucial for ensuring equal opportunities for all students. Teachers need comprehensive training programs to use AI tools effectively, and customising AI tools to meet specific needs is also essential (Owan et al., 2023; Penicini, 2023). Data privacy and security are crucial for protecting sensitive information and complying with regulations (Owan *et al*., 2023; Penicini, 2023). Ethical practices and bias mitigation are essential for promoting fairness and inclusivity. Cost and sustainability are crucial for long-term success. Collaboration between educators, policymakers, technology providers, and other stakeholders is needed to overcome these challenges and harness the transformative potential of AI in education.

Student engagement and learning outcomes

Integrating AI technologies in higher education has resulted in a noteworthy transformation in student engagement, motivation, and learning outcomes (Mbodila and Ndebele, 2022; Jiayu, 2023). The literature consistently demonstrates that AI-driven educational interventions have significantly heightened levels of student engagement (Mbodila and Ndebele, 2022; Jiayu, 2023; Pedro *et al*., 2019). Abgaryan *et al*. (2023) attest that AI fortifies student engagement because when teachers' paperwork tasks are automated, they have sufficient time to engage with students (Mishara and Shinde, 2023).

Findings by Lee *et al*. (2022) revealed that increased cognitive presence and deeper cognitive engagement with course topics are related to superior learning outcomes for students. Additionally, a randomised controlled trial by Nazari *et al*. (2021) revealed that students who used AI significantly improved their scores statistically significantly. Lee *et al*. (2022) attest that there were percentage increases in student achievement when AI-driven interventions were compared to traditional educational approaches in various disciplines (Rodríguez et al., 2022). Hence, AI-driven platforms such as gamification and real-time feedback promote active participation and sustained interest, motivating students to continuously improve (Qureshi, 2023; Tapalova and Zhiyenbayeva, 2022; Vanichvasin, 2022; Rutner and Scott, 2022; Yu, 2023; Dever *et al*., 2020).

Furthermore, AI systems allow students to track their progress and set achievable goals, fostering a sense of accomplishment and motivation to excel (Shalini and Tewari, 2020). On the other hand, the adaptive nature of AI platforms ensures that students are appropriately challenged, preventing boredom and sustaining motivation (Rutner and Scott, 2022). Kuleto *et al*. (2021) add that students prefer new technologies in education due to their high interactivity, motivation, enthusiasm, and opportunities for experimentation and simulation.

Various metrics are essential to effectively measure student engagement and learning outcomes in AI-driven educational interventions. Quantitative metrics include attendance, participation, completion, and assessment scores (Henri, 2015; Lee, 2019). Qualitative metrics include self-reported feedback, observational data, learning outcomes assessment, learning analytics, predictive modelling, retention and persistence metrics, and social and collaborative metrics (Henri, 2015; Lee, 2019). Quantitative metrics like attendance, participation, completion rates, and assessment scores provide insights into student engagement and learning experiences (Henri, 2015; Lee, 2019). Qualitative metrics like self-reported feedback, observational data, and performance-based assessments provide insights into student perceptions and satisfaction (Henri, 2015; Lee, 2019). Predictive modelling uses machine-learning algorithms to predict outcomes based on engagement patterns and historical trends. Social and collaborative metrics include peer interaction, communication, knowledge sharing, and peer evaluation.

Administrative efficiency

Various scholars highlight the automation of various administrative tasks in higher education. The review consistently reveals that AI technologies have effectively automated administrative tasks (Zouhaier, 2023). In a conference proceeding, Turcu and Turcu (2020) indicated that HE systems face various problems in a huge bureaucratic environment. Time restraints, tight budgets, and a lack of human resources hamper the management of numerous tasks. There is a need for workable and simple-to-implement solutions to lessen the stress experienced by educators, students, and administrative employees in many departments (Pedro et al., 2019). Robotic Process Automation technology, which is still relatively new, may make it possible to lessen the burden of repetitive, simple tasks on employees (Mishara and Shinde, 2023; Mosteanu, 2022).

Abgaryan *et al*. (2023) suggest that AI can assist in automating different tasks to free up teacher's time. Additionally, AI-based administrative tools can optimise administrative tasks like scheduling, grading, and enrolment management, thereby reducing administrative workload and giving faculty members more time to focus on their pedagogical and scholarly endeavours (Abgaryan *et al*., 2023).
Mosteanu (2022) concurs with Turcu and Turcu (2020) by revealing three potential partnerships between artificial intelligence and human intelligence in higher education institutions: (a) Physical Robots for guidance, (b) Robotic Process Automation for admission and registration, and (c) Robotic Process Automation for student evaluation. AI-powered chatbots and virtual assistants assist prospective students throughout admissions, providing timely information and support (Mosteanu, 2022). This could result in expedited admissions and enrolment procedures.

AI-driven systems offer immediate responses to student queries, reducing response times and facilitating efficient support services (Mosteanu, 2022; Yu, 2023; Almufarreh *et al*., 2021; Zouhaier, 2023). AlAfnan, Dishari, Jovic and Lomidze (2023) attest that because ChatGPT gives students accurate and trustworthy information, it can replace search engines. The study discovered that ChatGPT offers a venue for students to seek solutions to theory-based queries and develop concepts for application-based queries (AlAfnan *et al*., 2023). Additionally, AI-enabled data processing and management systems have automated data entry, validation, and reporting, reducing manual workload and minimising errors (Mosteanu, 2022).

The use of AI technology in education has raised ethical and employment concerns. Privacy and data security are crucial, as AI systems rely on vast amounts of student and institutional data (Klimova et al., 2023; Reiss, 2021; Tarisayi, 2023). Algorithmic bias and fairness are also important, as AI algorithms may perpetuate biases in historical data. Transparency and accountability are also crucial, as stakeholders may have limited visibility into decision-making processes (Klimova et al., 2023; Reiss, 2021; Tarisayi, 2023). Job displacement and redistribution are also concerns, as automation may disrupt traditional roles. However, automation also creates employment opportunities in AI development, implementation, and maintenance (Klimova et al., 2023; Reiss, 2021; Tarisayi, 2023). AI technology can augment human capabilities and productivity through collaboration, fostering a symbiotic relationship between humans and machines. Educational institutions must provide ongoing training, professional development, and lifelong learning opportunities to empower staff to thrive in an AI-driven workplace.

Resource allocation optimisation

Findings highlight the significant impact of AI on resource allocation. AI-driven predictive analytics assist institutions in optimising financial resource allocation, ensuring that funds are allocated efficiently to programs and initiatives that yield the highest returns (Mosteanu, 2022). AI algorithms have been utilised to optimise course scheduling, considering faculty availability, student preferences, and room availability (Mosteanu, 2022). This could result in more efficient use of institutional resources. The research by Khiat (2022) revealed that using a time management enabling system helped students study the material consistently, assisted them in doing so, and helped them practice more effective time management, which affected performance.

Furthermore, this indicates that AI-driven systems can enhance library resource management by optimising book and resource placement, automating cataloguing, and providing personalised recommendations to students and faculty (Khiat, 2022; Mosteanu, 2022). Aldulaimi, Abdeldayem, Abo Keir and Al-Sanjary, (2021) further indicate in their findings that utilising electronic technologies and methods to distribute material, give tasks and assignments to students, and assess their performance to lessen the administrative costs associated with academic programs. Additionally, Mishra and Shinde (2023) reveal that AI-powered platforms can analyse students' knowledge, learning styles, and preferences to provide customised content. By analysing data from previous interactions, AI algorithms can recommend appropriate learning resources, such as articles, videos, interactive modules, or simulations.

AI-driven resource allocation in educational institutions faces several challenges, including data quality, algorithmic bias, complexity, resource constraints, change management, ethical considerations, and adaptability (Klimova *et al*., 2023; Reiss, 2021; Tarisayi, 2023). High-quality, relevant data is crucial for accurate predictions, but fragmented or outdated sources can hinder data integration and preprocessing. AI algorithms may perpetuate biases, leading to unfair treatment or discrimination (Klimova *et al*., 2023; Reiss, 2021; Tarisayi, 2023). Complexity and interpretability can erode trust and confidence among users. Limited resources, such as funding, faculty, facilities, and support services, require careful prioritisation. Change management and stakeholder engagement are essential for acceptance and collaboration. Ethical considerations are crucial, as AI-driven decisions can impact student opportunities, academic programs, research initiatives, and institutional priorities (Klimova *et al*., 2023; Reiss, 2021; Tarisayi, 2023). Adaptability and robustness are essential for accommodating changes and uncertainties. A multidisciplinary approach involving education, data science, ethics, policy, and stakeholder engagement can help overcome these challenges.

Data-driven decision-making

The integration of AI analytics has revolutionised the decision-making landscape within HE institutions (AlAfnan *et al*., 2023). AI-powered predictive analytics help forecast enrolment trends, optimise resource allocation, and align course offerings with student demand (Mbodila and Ndebele, 2022; Jiayu, 2023). AI-generated insights also enhance student support services, providing personalised recommendations for academic advising, course selection, and career guidance (Trust, Whalen and Mouza, 2023). AI chatbots and sentiment analysis tools help identify students needing mental health support (Rodríguez *et al*., 2022; Shalini and Tewari, 2020). Operational efficiency gains have been facilitated by AI, including financial optimisation and maintenance and resource optimisation. AI-driven insights have become indispensable tools for better HE decision-making by aligning course offerings with anticipated student demand and reducing under-enrolled courses (Mosteanu, 2022).

The integration of AI analytics in higher education institutions presents numerous benefits but also raises ethical concerns and potential risks. Data privacy, bias, and transparency are key concerns. AI systems rely on vast amounts of student data, which must be handled responsibly to protect students' rights and maintain trust. Algorithmic bias can perpetuate or amplify existing inequalities or discrimination in the data used to train them, leading to unfair treatment of certain student groups (Ntoutsi *et al*., 2020). Transparency and accountability are also crucial. AI algorithms often operate as "black boxes," making it difficult to understand decisions and who should be responsible for errors or adverse outcomes (Lo Piano, 2020; Webb, 2019). Overreliance on AI-generated insights can diminish human judgment and critical thinking skills (Lo Piano, 2020; Webb, 2019). To address these ethical considerations, robust governance frameworks, clear policies for data governance, algorithmic transparency, and accountability mechanisms are necessary. Additionally, ethical AI training for staff involved in decision-making is essential to ensure awareness of potential risks and ethical best practices. By proactively addressing these ethical considerations, higher education institutions can harness the benefits of AI while mitigating potential risks and ensuring fair and equitable decision-making processes.

Equity and inclusivity

Integrating AI technologies in HE presents the opportunity to address issues of equity and inclusivity as well as existing disparities. AI-driven personalised learning pathways cater to individual student needs, potentially closing achievement gaps by providing targeted support to underserved or struggling students (Mosteanu, 2022; AlAfnan *et al*., 2023; Rodríguez *et al*., 2022; Shalini and Tewari, 2020; Qureshi, 2023). AI-based recommendations can direct students to resources, such as online courses and tutorials, addressing student resource disparities (Almufarreh *et al*., 2021; Aldulaimi *et al*., 2021). AI-powered language translation and accessibility tools can facilitate learning for non-native English speakers, promoting inclusivity (Zouhaier, 2023).

Findings highlight how AI-driven initiatives enhance accessibility. AI-driven tools, such as text-to-speech and speech-to-text technology, benefit students with disabilities by providing alternative means of engaging with course content (Mishra and Shinde, 2023; Almufarreh *et al*., 2021; Zouhaier, 2023). AI can adapt content for diverse learning needs, supporting students with varying abilities and preferences (Yu, 2023; Mishra and Shinde, 2023). AI can play a role in early intervention to prevent educational disparities by using AI predictive models to identify at-risk students falling behind or dropping out, enabling timely intervention and support (Rodríguez *et al*., 2022). AI-generated insights can guide institutions in allocating resources, such as tutoring or counselling services, to students who need them most (Yu, 2023).

A multifaceted approach is needed to ensure equitable access to AI tools in higher education. One strategy is prioritising accessibility in designing and implementing AI technologies, ensuring they cater to diverse learning needs and abilities (Klimova *et al*., 2023; Ntoutsi *et al*., 2020; Reiss, 2021; Tarisayi, 2023). Institutions must provide training and support to ensure all students have the skills and knowledge to effectively utilise AI tools, particularly those from marginalised or underserved communities. Collaboration with community organisations and stakeholders is another critical strategy, extending access to AI resources and support services to underserved populations Klimova *et al*., 2023; Ntoutsi *et al*., 2020; Reiss, 2021; Tarisayi, 2023). Transparent and inclusive procurement processes are also essential to prevent the exacerbation of inequalities in AI adoption. Institutions should prioritise vendors and technologies that adhere to principles of accessibility, fairness, and non-discrimination Klimova *et al*., 2023; Ntoutsi *et al*., 2020; Reiss, 2021; Tarisayi, 2023). Thorough evaluations of AI products and services can assess their potential impact on marginalised groups and ensure alignment with equity goals. However, there are inherent risks that AI technologies may inadvertently perpetuate existing inequalities if not carefully implemented and monitored. One risk is algorithmic bias, where AI systems may reflect and amplify biases in the data used to train them, leading to unfair treatment or discrimination against certain groups. Another risk is the digital divide, where disparities in access to technology and internet connectivity create barriers to equitable participation in AI-enhanced learning environments.

Adoption Trends

The adoption of AI technologies within HE exhibits notable trends, reflecting varying adoption rates and patterns across institutions and academic disciplines. Integrating AI technologies in higher education faces varying perspectives among faculty and students. Faculty members may fear job displacement or loss of autonomy due to AI, while students may embrace it to enhance teaching effectiveness and improve student outcomes. Students may welcome AI-driven initiatives as opportunities for personalised learning, increased engagement, and improved academic performance. In contrast, others express concerns about data privacy, algorithmic bias, and the potential for AI to replace human interaction (Chan and Tsi, 2023).

Institutions can implement transparent communication, open dialogue, feedback mechanisms, and professional development workshops to address these concerns. Prioritising inclusivity and diversity in AI development and implementation can mitigate potential biases and ensure equitable access to AI-enhanced educational experiences (Chan and Tsi, 2023). Moreover, integrating AI into curriculum design and pedagogical practices should be guided by learner-centeredness, autonomy, and critical reflection principles. Academics can use AI technologies to create adaptive learning environments, foster collaboration and creativity, and empower students as active agents in their learning journey.
These trends are discussed below.

Patterns of adoption and emerging trends

Literature reveals varying patterns of AI adoption across higher education institutions. Research-intensive universities such as East China Normal University (ECNU) often leads in AI adoption owing to its endowment with resources as well as to its research focus (Gawande *et al*., 2020). Although the educational technology sector has not stopped producing innovations, there are no signs that AI-based applications for teaching and learning or system management have been used widely (Pedro *et al*., 2019).

Although the adoption of AI in higher education remains behind that of the business sector, those educational institutions that have already done so and are investing more money into AI applications will undoubtedly continue to be in the lead over their rivals (Netragaonkar, 2022). Smaller institutions, including community colleges and liberal arts colleges, are increasingly adopting AI, often through collaborations or partnerships with edtech companies.

Leoste *et al*. (2022) argue that Telepresence Robots (TPR) use in educational or professional settings has not yet attained a degree of widespread adoption comparable to other technologies like videoconferencing. In agreement, Yu (2023) indicates that to meet the demand for various delivery options, online learning environments have implemented multimodal interfaces and provided learners with the tools to interact with multimodal content (Yu, 2023).

The adoption of AI is higher in science, technology, engineering, and mathematics (STEM) disciplines. AI is integrated into these fields' research, coursework, and laboratory settings (Sadiku *et al*., 2022). The adoption in humanities and social science has been lagging, but is now showing signs of picking up, with AI tools being used in text analysis, language translation, and cultural studies (Sadiku *et al*., 2022). Disciplinary differences in AI adoption may be attributed to resource availability, expertise of academics, and perceived relevance to the curriculum (Turugare and Rudhumbu, 2020). Sadiku *et al*. (2022) attest that some academics are fearful of adopting AI as they believe AI will make them redundant.

Factors that influence adoption

The findings suggest that financial resources, leadership commitment, and institutional culture influence the adoption of AI (Turugare and Rudhumbu, 2020). Institutions with open culture tend to adopt AI technologies more readily. The challenge with such institutions is that sometimes the adopt new technological development without thorough evaluations of the advantages and disadvantages of the technological development (Pedro *et al*., 2019).

The commitment to technology demonstrated by those in leadership positions also influence the adoption of technological developments such as AI. For example, how professors and other academics feel about technology greatly impacts how it is adopted and integrated into the classroom.

Findings consistently highlight that institutions with substantial financial resources tend to adopt AI technologies more rapidly. Funding availability for technology investments plays a critical role (Turugare and Rudhumbu, 2020). Strong leadership support from institutional administrators and decision-makers drives AI adoption significantly. Key factors include leadership vision and commitment to technological innovation (Turugare and Rudhumbu, 2020). The buy-in of professors and other academics critical. Their support and positive attitudes are crucial for initiating and implementing educational technology in the classroom (Sadiku *et al*., 2022; Pedro *et al*., 2019). Institutions involving academics in selecting and implementing AI technologies tend to have higher adoption rates (Turugare and Rudhumbu, 2020). AI technologies that align with pedagogical goals and student learning outcomes are more likely to be adopted. Demonstrated benefits in teaching and learning are influential (Turugare and Rudhumbu, 2020). Collaborations with external partners, such as edtech companies and research organisations, facilitate AI adoption by providing access to expertise and resources (Turugare and Rudhumbu, 2020).

Netragaonkar (2022) and Jain and Jain (2019) found that the adoption of AI in higher education lags behinds the adoption in other sectors. The main barriers to adoption in higher education   are a lack of awareness, limited financial resources, data privacy concerns, and staff resistance (Turugare and Rudhumbu, 2020).

Perspectives of academic staff and students

Quite often there are differences in the perspectives of academic staff and students regarding the adoption of new technological developments such as AI. The literature reveals that some academic staff view AI as a valuable tool for enhancing teaching and reducing administrative burdens, while others express concerns about job security and the potential dehumanisation of education (Roumate, 2023; Sadiku *et al*., 2022; Lin and Streinz, 2021). Among the academic staff, some are innovators and early adopters who are enthusiastic about integrating AI into their teaching methods, while others exhibit reluctance or scepticism (Mbodila and Ndebele, 2022; Sadiku *et al*., 2022). Academic staff members who have experience with AI tools often report positive outcomes, including improved student engagement, personalised learning experiences, and more efficient administrative processes (Mbodila and Ndebele, 2022; Pedro *et al*., 2019). Academic staff members may require training and professional development opportunities to effectively integrate AI technologies into their teaching practices (Pedro *et al*., 2019). There is a significant number of academics who believe that excessive use of AI could hinder students' exploration and learning opportunities.

On the other hand, students generally have high expectations for technology in education and accept AI integration. They appreciate AI's convenience and personalised experiences (Tapalova and Zhiyenbayeva, 2022; Vanichvasin, 2022). Those who have encountered AI-driven tools report positive impacts on their learning, such as improved comprehension and more interactive educational experiences (Jiayu, 2023; Mbodila and Ndebele, 2022; Vanichvasin, 2022). However, they are concerned about the impersonal nature of the support provided by AI which may negatively impact learning ability. They believe AI can address privacy concerns and improve learner-instructor connections by providing social interaction cues without personal camera information. However, students find eye tracking and facial expression analysis uncomfortable, as they feel under surveillance. Tuomi (2019) attests that some students express concerns about data privacy and AI systems' potential misuse of personal information.

Abgaryan *et al*. (2023), and Pedro *et al*. (2019) highlight the digital divide, with disparities in access to technology and digital literacy affecting students' experiences with AI. Bridging this gap is essential for equitable AI integration.

Conclusion

The main conclusion is that the opportunities presented by AI have far-reaching implications for the future of higher education, particularly teaching and learning. The findings demonstrate that AI has the potential to revolutionise pedagogical practices, foster student engagement, streamline administrative processes, and promote data-driven decision-making. AI can drive the process if pedagogical transformation by equipping educators with tools to implement innovative teaching methods that cater to individual learning styles and paces. This enhances the learning experience and allows instructors to focus on higher-order pedagogical activities, such as critical thinking and problem-solving. The implications for teaching and learning are profound as students become active participants in their education, leading to deeper comprehension and improved knowledge retention.

Similarly AI plays a critical role in enhancing student engagement and improving learning outcomes. By providing personalised learning experiences, AI technologies address students' diverse needs and abilities. The implications are evident in higher academic achievement and increased motivation among students. The potential to narrow achievement gaps and enhance the overall quality of education is promising.

Furthermore, AI streamlines administrative tasks, leading to cost savings and operational efficiency in HE institutions. Using AI, institutions can allocate resources more effectively, direct funds toward strategic initiatives, and optimise administrative workflows. This, in turn, contributes to the sustainability and competitiveness of HE institutions.

AI-generated insights empower institutions to make data-driven decisions, improving institutional planning and student support services. The implications extend to enhanced institutional performance, efficient resource allocation, and a more responsive educational ecosystem. The ability to proactively identify and address challenges, such as student retention and course optimisation, aligns institutional priorities with student success.

AI has also demonstrated potential to address issues of equity and inclusivity. By providing tailored support and accommodations, AI can bridge educational disparities and ensure accessibility for all learners. The implications extend to promoting diversity, supporting underserved populations, and fostering a more inclusive HE landscape.

The significance of AI in higher education extends beyond immediate improvements in teaching and learning. The broader implications include enhancing global competitiveness, attracting students and faculty seeking cutting-edge educational experiences. AI-equipped graduates are better prepared for the modern workforce, where AI proficiency is increasingly valued. AI-driven personalised learning experiences extend beyond traditional degree programs, facilitating lifelong learning and skill development. AI accelerates research and innovation in higher education, facilitate data-driven discoveries and advance academic knowledge.

References

Abgaryan, H., Asatryan, S. and Matevosyan, A. 2023. Revolutionary changes in higher education with artificial intelligence. *Main Issues Of Pedagogy And Psychology 10*(1), 76-86.

Akinwalere, S. N. and Ivanov, V. 2022. Artificial intelligence in higher education: challenges and opportunities. *Border Crossing 12*(1), 1-15.

AlAfnan, M. A., Dishari, S., Jovic, M. and Lomidze, K. 2023. ChatGPT as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. *Journal of Artificial Intelligence and Technology 3*(2), 60-68.

AlDhaen, F. 2022. The use of artificial intelligence in higher education–systematic review. *COVID-19 Challenges to University Information Technology Governance*, 269-285.

Aldulaimi, S. H., Abdeldayem, M. M., Abo Keir, M. Y. and Al-Sanjary, O. I. 2021. E-Learning in higher education and Covid-19 outbreak: Challenges and opportunities. *Psychology and Education, 58*(2), 38-43. Retrieved from http://www.psychologyandeducation.net.

Alexander, B., Ashford-Rowe, K., Barajas-Murph, N., Dobbin, G., Knott, J., McCormack, M., ... and Weber, N. 2019. *Horizon Report 2019: Higher Education Edition* (pp. 3-41). EDU19, United States.

Almufarreh, A., Arshad, M. and Mohammed, S. H. 2021. An efficient utilisation of blackboard ally in higher education institution. *Intelligent Automation and Soft Computing*, *29*(1), 74-87.

Arakpogun, E.O., Elsahn, Z., Olan, F. and Elsahn, F. 2021. Artificial intelligence in africa: challenges and opportunities. In: Hamdan, A., Hassanien, A.E., Razzaque, A. and Alareeni, B. (eds) *The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success*. Studies in Computational Intelligence, vol 935. Springer, Cham. https://doi.org/10.1007/978-3-030-62796-6_22.

Barakina, E. Y., Popova, A. V., Gorokhova, S. S. and Voskovskaya, A. S. 2021. Digital technologies and artificial intelligence technologies in education. *European Journal of Contemporary Education 10*(2), 285-296.

Bethencourt-Aguilar, A., Castellanos-Nieves, D., Sosa-Alonso, J. J. and Area-Moreira, M. 2023. Use of generative adversarial networks (gans) in educational technology research. *Journal of New Approaches in Educational Research*, Spain.

Brown, M., McCormack, M., Reeves, J., Brook, D. C., Grajek, S., Alexander, B., ... and Weber, N. 2020. 2020 Educause Horizon. *Report Teaching and Learning Edition* (pp. 2-58). Educause, United States.

Butcher, N., Wilson-Strydom, M., and Baijnath, M. 2021. *Artificial intelligence Capacity in sub-Saharan Africa: Compendium Report*.

Cağataylı, M. and Celebi, E. 2022. Estimating academic success in higher education using big five personality traits,a machine learning approach. *Arab Journal Scientific Engineering, 47*, 1289–1298. https:// doi. org/ 10. 1007/s13369- 021- 05873-4.

Chan, C.K. and Tsi, L.H. 2023. The AI Revolution in Education: Will AI Replace or Assist Teachers in Higher Education? *ArXiv, abs/2305.01185*.

Chatterjee, S. and Bhattacharjee, K. K. 2020. Adoption of artificial intelligence in higher education: A quantitative analysis using structural equation modelling. *Education and Information Technologies 25*, 3443-3463.

Chu, H., Tu, Y. and Yang, K. 2022. Roles and research trends of artificial intelligence in higher education: A systematic review of the top 50 most-cited articles. *Australasian Journal of Educational Technolo*gy 38(3), 22–42. https:// doi.org/10.14742/ajet.7526

Crompton, H. and Burke, D. 2023. Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education 20*(1), 1-22.

Dever, D. A., Azevedo, R., Cloude, E. B. and Wiedbusch, M. 2020. The impact of autonomy and types of informational text presentations in game-based environments on learning: Converging multi-channel processes data and learning outcomes. *International Journal of Artificial Intelligence in Education 30*(4), 581–615. https:// doi. org/ 10. 1007/s40593- 020- 00215.

Devi, D. and Rroy, A. D. 2023. Role of artificial intelligence (ai) in sustainable education of higher education institutions in Guwahati City: Teacher's perception. *International Management Review 19, 111-116*.

Diwaker, C., Sharma, A. and Tomar, P. 2021. Artificial intelligence in higher education and learning. *Impact of AI Technologies on Teaching, Learning, and Research in Higher Education (pp. 62-72) IGI Global*.

Eager, B. and Brunton, R. 2023. Prompting higher education towards AI-augmented teaching and learning practice. *Journal of University Teaching and Learning Practice 20*(5), 02.

Elgendy, N., Elragal, A. and Päivärinta, T. 2022. DECAS: a modern data-driven decision theory for big data and analytics. *Journal of Decision Systems 31*(4), 337-373.

Gawande, V., Al Badi, H. and Al Makharoumi, K. 2020. An empirical study on emerging trends in artificial intelligence and its impact on higher education. *International Journal of Computer Applications 175*(12), 43-47.

Gocen, A. and Aydemir, F. (2021). Artificial intelligence in education and schools. *Research on Education and Media 12*(1), 13-21.

Henrie, C.R., Halverson, L.R. and Graham, C.R. 2015. Measuring student engagement in technology-mediated learning: A review. *Computers and Education. 90*, 36-53.

Holmes, W. and Porayska-Pomsta, K. (eds.). 2022. *The Ethics of Artificial Intelligence in education: Practices, challenges, and debates*. Taylor and Francis: New York.

Hutson, J. and Ceballos, J. 2023. Rethinking education in the age of AI: the importance of developing durable skills in the Industry 4.0. *Journal of Information Economics 1*(2), 26-36.

Issa, T. (ed). 2015. *Artificial Intelligence Technologies and the Evolution of Web 3.0*. IGI Global, Hershey PA, United States.

Jain, S. and Jain, R. 2019. Role of artificial intelligence in higher education—An empirical investigation. *IJRAR-International Journal of Research and Analytical Reviews* 6(2), 144z-150z.

Jiayu, Y. 2023. Challenges and opportunities of generative artificial intelligence in higher education student educational management. *Advances in Educational Technology and Psychology 7*(9), 92-96.

Khadimally, S. 2022. Mastery of learning via social networking technology: a comprehensive literature review with a systems approach. *Applications of Machine Learning and Artificial Intelligence in Education*, 142-161.

Khiat, H. 2022. Using automated time management enablers to improve self-regulated learning. *Active Learning in Higher Education 23*(1), 3-15.

Klimova, B.F., Pikhart, M. and Kacetl, J. 2023. Ethical issues of the use of AI-driven mobile apps for education. *Frontiers in Public Health, 10, 1118116*. https://doi.org/10.3389/fpubh.2022.1118116

Kordrostami, M. and Seitz, V. 2022. Faculty online competence and student affective engagement in online learning. *Marketing Education Review* 32, 240 - 254. https://doi.org/10.1080/10528008.2021.1965891.

Kuleto, V., Ilić, M., Dumangiu, M., Ranković, M., Martins, O. M., Păun, D. and Mihoreanu, L. 2021. Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions. *Sustainability 13*(18), 10424.

Lee, J., Soleimani, F., Hosmer IV, J., Soylu, M. Y., Finkelberg, R. and Chatterjee, S. 2022. Predicting Cognitive Presence in at-scale Online Learning: MOOC and For-Credit Online Course Environments. *Online Learning 26*(1), 58-79.

Lee, J., Song, H. and Hong, A.J. 2019. Exploring factors, and indicators for measuring students' sustainable engagement in e-learning. *Sustainability, 11(4), 985*.

Leoste, J., Virkus, S., Talisainen, A., Tammemäe, K., Kangur, K. and Petriashvili, I. 2022. Higher education personnel's perceptions about telepresence robots. *Frontiers in Robotics and AI 9*, 976836.

Lo Piano, S. 2020. Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanities and Social Sciences Communications, 7 (1), 1-7*.

Martin, S. M. 2019. *Artificial Intelligence, Mixed Reality, and the Redefinition of the Classroom*. Rowman and Littlefield.

Mbodila, M. and Ndebele, C. 2022. The impact of social networks on teaching and learning at a rural university in South Africa. *Journal of Educators Online*, *19*(3), n3.

Metcalfe, J.S., Perelman, B.S., Boothe, D.L. and McDowell, K. 2021. Systemic Oversimplification Limits the Potential for Human-AI Partnership. *IEEE Access, 9*, 70242-70260.

Mishara, A. and Shinde, V. B. 2023. The impact of artificial intelligence on students: shaping education for the future. *The International Journal for Innovative Engineering and Management Research*. 576-586. https://ijiemr.org/downloads/Volume-12/Issue-05.

Mosteanu, N. R. 2022. Machine learning and robotic process automation take higher education one step further. *Science and Technology 25*(1), 92-99.

Nazari, N., Shabbir, M. S. and Setiawan, R. 2021. Application of artificial intelligence powered digital writing assistant in higher education: randomised controlled trial. *Heliyon*, *7*(5), E07014. https://doi.org/10.1016/j.heliyon.2021.e07014

Netragaonkar, D. Y. 2022. Role of artificial intelligence (AI) in higher education. *IJFANS International Journal of Food and Nutritional Sciences, UGC CARE Listed (Group-I) Journal*, *11*(10), 1-7.

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K.E., Wagner, C., Karimi, F., Fernández, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., Broelemann, K., Kasneci, G., Tiropanis, T. and Staab, S. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1356*.

O'Dea, X. C. and O'Dea, M. 2023. Is Artificial intelligence really the next big thing in learning and teaching in higher education? a conceptual paper. *Journal of University Teaching and Learning Practice*, *20*(5), article 05. http://dx.doi.org/10.53761/1.20.5.05

Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., and Bassey, B. A. 2023. Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education 19*(8), em2307.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... and Moher, D. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International Journal of Surgery 88*, 105906.

Pedro, F., Subosa, M., Rivas, A. and Valverde, P. 2019. *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*. UNESCO, France.

Peng, S. Y., Lin, C. F. and Streinz, T. (eds.). 2021. *Artificial Intelligence and International Economic Law*. Cambridge University Press. United Kingdom.

Poole, D. L. and Mackworth, A. K. 2010. *Artificial Intelligence: Foundations of computational Agents*. Cambridge University Press. United Kingdom.

Popenici, S. 2023. The critique of AI as a foundation for judicious use in higher education. *Journal of Applied Learning and Teaching 6*(2), 1-7.

Popenici, S.A. and Kerr, S. 2017. Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning, 12 (1), 22*.

Qureshi, B. 2023. Exploring the use of ChatGPT as a tool for learning and assessment in undergraduate computer science curriculum: Opportunities and challenges. *arXiv preprint arXiv:2304.11214*.

Rawas, S. 2023. ChatGPT: Empowering lifelong learning in the digital age of higher education. *Education and Information Technologies* 1-14.

Reiss, M.J. 2021. The use of AI in education: Practicalities and ethical considerations. *London Review of Education, 19(1), p.n1*.

Rodríguez, M. E., Guerrero-Roldán, A. E., Baneres, D. and Karadeniz, A. 2022. An intelligent nudging system to guide online learners. *International Review of Research in Open and Distributed Learning*, *23* (1), 41-62.

Roumate, F. (ed). 2023. *Artificial Intelligence in Higher Education and Scientific Research : Future Development*. Springer.

Rufrano, M. and Yeung, J. E. 2023. Digital pedagogy for the present: an artificial intelligence methodology for curriculum development. 9th International Conference on Higher Education Advances (HEAd'23) Universitat Politecnica de Valencia, Valencia, 2023 (pp. 757-764). http://dx.doi.org/10.4995/HEAd23.2023.16175

Rutner, S. M. and Scott, R. A. 2022. Use of artificial intelligence to grade student discussion boards: An exploratory study. *Information Systems Education Journal 20*(4), 4-18.

Sadiku, M. N., Musa, S. M. and Chukwu, U. C. 2022. *Artificial intelligence in education*. iUniverse. Artificial Intelligence in Education - Matthew N.O. Sadiku, Sarhan M. Musa, Uwakwe C. Chukwu - Google Books.

Seo, K., Tang, J., Roll, I., Fels, S. and Yoon, D. 2021. The impact of artificial intelligence on learner–instructor interaction in online learning. *International journal of educational technology in higher education*, *18*(1), 1-23.

Shalini, and Tewari, A. 2020. Sustainable education in India through artificial intelligence: Challenges and opportunities. In *Companion Publication of the 12th ACM Conference on Web Science*, 41-47.

Tambe, M. and Rice, E. (eds). 2018. *Artificial Intelligence and Social Work*. Cambridge University Press. United Kingdom.

Tapalova, O. and Zhiyenbayeva, N. 2022. Artificial intelligence in education: AIEd for personalised learning pathways. *Electronic Journal of e-Learning 20*(5), 639-653.

Tarisayi, K. S. 2023. Strategic leadership for responsible artificial intelligence adoption in higher education. *CTE Workshop Proceedings*.

Trust, T., Whalen, J. and Mouza, C. 2023. Editorial: ChatGPT: Challenges, opportunities, and implications for teacher education. *Contemporary Issues in Technology and Teacher Education 23*(1), 1-23.

Tuomi, I. 2019. The Impact of Artificial Intelligence on Learning, Teaching, and Education: Policies for the Future. *JRC Science for Policy Report*. European Commission, Belgium.

Turcu, C. and Turcu, C. 2020. On robotic process automation and its integration in higher education. In *Conference Proceedings. The Future of Education 2020*. ICT4777. 6243-ICT4777-FP-FOE10.pdf (pixel-online.net)

Turugare, M. and Rudhumbu, N. 2020. Integrating technology in teaching and learning in universities in Lesotho: opportunities and challenges. *Education and Information Technologies 25*(5), 3593-3612.

Vanichvasin, P. 2022. Impact of Chatbots on Student Learning and Satisfaction in the Entrepreneurship Education Programme in Higher Education Context. *International Education Studies 15*(6), 15-26.

Webb, H. 2019. A governance framework for algorithmic accountability and transparency. *Policy Document*. European Union, Brussels. https://doi.org/10.2861/59990.

Wissman, K. T. 2023. Student perceptions on the use of flip in higher education: learning strategies and classroom community. *International Journal of Technology in Education and Science 7*(1), 1-17.

Xia, Q., Chiu, T. K., Zhou, X., Chai, C. S. and Cheng, M. 2022. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4, p100118. https://doi.org/10.1016/j.caeai.2022.100118

Yang, C., Huan, S. and Yang, Y. 2020. A practical teaching mode for colleges supported by artificial intelligence. *International Journal of Emerging Technology and Learning (IJET). 15(17)*, 195-206.

Yu, E. 2023. Intelligent enough? Artificial intelligence for online learners. *Journal of Educators Online*, *20*(1), p.n1.

Zouhaier, S. 2023. The impact of artificial intelligence on higher education: an empirical study. *European Journal of Educational Sciences*, *10*(1), 17-33.

A Literature-based Evaluation of the Importance of Adopting Artificial Intelligence in Higher Education

O. S. Ogunleye and N. Wayi-Mgwebi
University of Mpumalanga
olalekan.ogunleye@ump.ac.za

Abstract

The proliferation of information technology and the speedy adoption of artificial intelligence systems in the last decade have made using artificial intelligence more common in many areas of life, including in education. As these technologies advance rapidly, artificial intelligence holds immense potential to transform the education landscape and address its unique challenges, particularly in developing countries. However, there are risks associated with overreliance on digital technology. Hence, the question is: to what extent will higher education benefit from integrating artificial intelligence in higher education? This paper provides an answer to this question. The paper reports on a comprehensive evaluation of literature reporting on empirical studies on using and integrating artificial intelligence in higher education settings for the period from 2010 to 2023. Multiple methods were applied, including selected bibliometrics, categorical meta-trend analysis and content analysis. The evaluation provides information on the current state of artificial intelligence in education research, selected artificial intelligence technologies and their adoption in education and reviews their potential benefits, the benefits of artificial intelligence in the educational process, and possible concerns associated with its use.

Keywords: Artificial intelligence, education, information technology, higher education, ethical considerations

Introduction

In 1950, Alan Turing pioneered the concept of "thinking machines" and its many potential applications. He envisaged machines capable of automating systems, and that is what, today, is referred to as artificial intelligence. As a combination of machines that can detect, recognize, learn, react, and solve complex problems (Kumar and Thakur, 2012; Spector and Martha, 1993) is emerging as the next disruptive innovation. Predictably, such innovative technology will bring about significant change in future workplaces (Horakova *et al*., 2017). Therefore, many research studies have viewed artificial intelligence as a driver that will form an essential part of the Fourth Industrial Revolution, and the possibility that it will usher in the Fourth Revolution in education, is high (Dai *et* al., 2020; Knox, 2020).

Over the past eight decades, artificial intelligence has seen increasing application across various sociocultural domains, reflecting its status as a key technology driving the future (Chaudhry and Kazim, 2022). In higher education, artificial intelligence has been used and studied with the help of various algorithms to develop the application of different technologies, including biometric recognition, machine learning, and virtual reality (Kumar and Thakur, 2012). It has prompted and induced significant changes in the substance of education techniques, management mode, education theory, teaching methods, and learning methods of higher education. Consequently, schools have started incorporating artificial intelligence education into their curricula (Dai *et al*., 2020; Knox, 2020).

Earlier the emergence of television and computers was once touted as game changers in education. However, they have enhanced access to information without substantially changing the core educational practices. In the same way that the introduction of computers and the Internet was initially heralded as a transformation for the educational system, it has been demonstrated that the innovations in these areas promoted access to information without fundamentally altering the foundational core of academic practices. Despite this, educators are obligated to investigate the current capabilities of artificial intelligence and investigate the various paths that could be taken to improve learning through artificial intelligence.

As a result of the increased attention given to the adoption of artificial intelligence in the educational sector, it is high time that recent research on the application of artificial intelligence in education is explored, examined, and evaluated. It is believed that this will allow educators to have an up-to-date accent on the trending information and understanding of the happenings around artificial intelligence implementation in education, preparing them for any potential paradigm shifts.

Loeckx (2016) argues that artificial intelligence could be a valuable educational tool that lightens the teaching load for students and educators while providing them with more beneficial educational experiences. This means there are many prospects for implementing artificial intelligence in education, particularly when these opportunities are combined with recent reforms such as digitizing educational resources, individualized learning experiences, and gamification. For instance, the modelling potential of artificial intelligence techniques has been exploited systematically to develop reactive and adaptive tutorials for constructing individualized learning environments (Boulay, 2016). This has been done to recompense the teacher shortage through intelligent tutoring systems (ITSs). These ITSs can provide personalised and tailored learning experiences in four primary methods: monitoring the learner's input, giving appropriate activities, providing helpful feedback, and implementing interfaces for human-computer communication (Seldom and Abidoye, 2018).

However, there is an upward possibility that the role of instructors will be impacted when more ITSs are built for more subjects and themes. As a result, schooling may need to be redefined to accommodate these changes. There are fears and misgivings among educators that artificial intelligence would threaten their careers (Lacity and Willcocks, 2017).

Furthermore, there is increasing awareness that the professional duties of educators will need to evolve in response to the development of artificial intelligence, which will result in the emergence of new organisational structures (Fenwick, 2018). In contrast, students as digital citizens can, to some extent, make better use of artificial intelligence to improve their learning outcomes (Flogie and Aberšek, 2015).

However, they tend to fail to employ artificial intelligence methods in acceptable and appropriate ways for a particular learning setting, which might result in negative attitudes toward learning (Ijaz *et al*., 2017).

This paper reports on a comprehensive evaluation of literature generated from several empirical studies on using and integrating artificial intelligence in educational settings. It focuses on literature held by the Web of Science Scopus that were published between 2010 and 2023. This specific thirteen-year period was selected because it corresponds to the period during which the second and third generations of artificial intelligence began to progress in education. The "first generation" of artificial intelligence was known to support human intellectual work, and the "second generation" may locate the ideal answer using statistical/search models. On the other hand, the "third generation," based on a brain model, will significantly improve recognition performance (Kumar and Thakur, 2012).

The research question adopted to guide the evaluation study is "*To what extent will the integration of artificial intelligence in higher education benefit the major stakeholders, i.e. the educators and students?*" However, to answer this question, the following sub-questions were taken into account: (a) What are the benefits and challenges of adopting artificial intelligence within the higher education sector? (b) What will be the key research interest areas regarding artificial intelligence in the higher education sector in the next decade? (c) What trends may be observed in recently published studies in the areas of application of artificial intelligence in higher education?

Objectives of the evaluation

The objectives of the evaluation are to:

- Identify and analyse the benefits of using artificial intelligence in higher education;
- Assess the current state of AI within the higher education sector; and
- Explore the trends in recent research on AI applications in higher education.

Methodology

The evaluation used a qualitative-interpretative methodology, using documentary type in the selection procedure and data recording suggested by Stokes and Urquhart (2013) and Wiesner (2022). This methodology was selected because it focuses on the understanding the complex and nuanced phenomena of artificial intelligence (AI) integration in higher education (Creswell and Poth, 2018). This method allows for an in-depth exploration of the subject matter, providing a comprehensive view of the existing literature and practices (Merriam and Tisdell, 2015). It is particularly suited for studies aiming to interpret patterns, themes, and the significance of AI in educational settings, where qualitative nuances are essential to comprehend the broader impacts and implications. Furthermore, Qualitative-interpretative research provides the depth needed to understand the intricate dynamics of AI in education, capturing the subtleties and complexities that quantitative methods might overlook. Also, this approach accommodates a broad range of empirical studies, supporting a comprehensive analysis of varied methodologies, findings, and theoretical contributions within the AI education domain. At the same time, it assists in contextualising findings within the broader educational and technological landscape, ensuring a richer and more meaningful analysis of the research study (Creswell and Poth, 2018).

The Web of Science and the Social Science Citation Index (SCSI) for the desired articles published between 2010 and 2023 were used because articles published in the SCSI database are perceived to be of high-quality publication by researchers in the education field. One hundred and twenty-three (123) articles on empirical studies were selected for use in the evaluation. Artificial intelligence and education were the search keywords in this study. After completing reading each article, the relevant information was input into the bibliographic matrix. This matrix is then used to catalogue the documents into categories.

Following the suggestion by Wu *et al*. (2013) the evaluation adopted a two-step approach. These are identification and coding. In the identification step, an article was included in the potential pool if it fulfilled one of the following two criteria: (a) the research utilised a particular artificial intelligence technique as an intervention in supporting learning or teaching, and (b) the article provided either empirical proof or an in-depth analysis. Studies that solely did not focus on the application of artificial intelligence were excluded. Second, regarding the research that was analysed, the only studies considered for inclusion were those that focused on artificial intelligence technologies' impact on the educational system. All the texts of all identified papers were examined and screened. The filtering procedure removed 59 articles from the initial list of 123.

The concept of thematic analysis was discussed in the second step. A coding system based on how artificial intelligence has been implemented in education was developed. The investigation mainly focused on research questions and the adoption of various technologies.

In terms of research issues, past research has established three fundamental models of artificial intelligence in knowledge processing. These models are information representation, knowledge acquisition, and knowledge derivation (Horakova *et al*., 2017). Based on this foundation, the research questions from the articles were organised into three categories: (a) development, which centred on the knowledge presentation model; (b) extraction, which centred on how to obtain knowledge from data mining; and (c) application, which emphasized the human-computer interaction through information derivation. Second, regarding the incorporation of technology, the emphasis was placed on the various forms of technology utilized by the research project. These forms of technology were further subdivided into software (such as algorithms and programmes) and hardware, including sensors and devices such as virtual reality).

The identification process involved three related activities. The first was using criteria for inclusion to select articles based on their direct application of AI in teaching and learning or providing empirical evidence or thorough analysis of AI's impact on educational systems. The second was the use of exclusion criteria to exclude studies focusing solely on AI as a learning subject, without practical application in education. The third was the screening process under which all articles meeting the inclusion criteria underwent a comprehensive review, with irrelevant or less pertinent studies filtered out.

The coding and thematic analysis involved establishing systematic coding framework to categorise how AI is implemented in the educational context. It involved five related activities. The first was development, which involved examining how AI contributes to the presentation and structuring of knowledge. The second was extraction, which involved investigating the role of AI in deriving insights from educational data. The third was application, which entailed assessing how AI facilitates interaction between humans and computers in learning. The fourth was technology categorisation, which entailed differentiating between AI's software (algorithms, programmes) and hardware (sensors, VR devices) aspects in education. The fifth was trend analysis which entailed conducting a comparative analysis of the research aims, AI technology adoption, and periods to predict future trends and challenges in AI education.

Results and discussion

The evaluation has revealed a multi-faceted landscape of artificial intelligence (AI) application in higher education. The key findings underscore AI's transformative potential in three primary areas: development, extraction, and application, each contributing uniquely to the educational ecosystem. These are discussed in the ensuing paragraphs.

Development

In the development dimension, AI emerges as a pivotal tool for innovating educational systems, where intelligent tutoring systems (ITS) and electronic assessments represent significant strides towards creating dynamic learning environments. These systems leverage AI's analytical capabilities, using algorithms for classification, matching, recommendation, and deep learning to craft personalized educational experiences. This not only aids in streamlining the teaching and learning processes but also facilitates a more nuanced understanding of academic content, enhancing both the delivery and absorption of knowledge. Sixteen of the articles focused on developing education systems, such as intelligent tutoring systems (ITS) and electronic assessment. In most cases, the development technique was carried out using an induction-deduction strategy, in which past experiments and data were evaluated to forecast the variables, followed by algorithm testing to generate the final modelling equation (Zipitria *et al*., 2013). An instructional system generally consists of presentations, logical modelling, and scope of data (Ge *et al*., 2018). Modelling techniques were the foundation of artificial intelligence methodology and fundamentally penetrated throughout the system development procedure. In this scope, research was typically undertaken in computer science or information science, and domain knowledge was imported as the source material into an algorithm frame with just a few pedagogical designs documented. For example, Horakova *et al.* (2017) used three classification strategies to investigate a text-mining machine's categorization capacity. The results reveal that artificial neural networks (ANNs) outperformed regression and decision trees in separating educational materials or text fragments. Bayesian networks, association rules, clustering, fuzzy C-means, and fuzzy and genetic algorithms were well-accepted algorithms for modelling individual student properties.

These strategies can be used to investigate the formation of homogenous and heterogeneous groups in an educational setting (Magnisalis *et al*., 2011).

Furthermore, the increasing amount of data challenges educators to interpret qualitative data efficiently. Natural language processing (NLP) simplifies and accelerates identifying what lies within the data, allowing it to diagnose and recommend the problem (Tierney, 2012). However, assessing a complicated educational system necessitates more in-depth information retrieval. It was proposed that smarter computer-aided systems be built in which agents may be educated automatically by integrating numerous methodologies, such as benchmarks in the NLP/Semantic Web sector (Malik *et al*., 2017).

Hierarchical structures were investigated as prospective methods to model the educational system to optimize modelling in the learning environment. This is because education is a complex system with subsystems and components, and invisible causal processes between subsystem/component behaviours will causally affect each other (Vattam *et al*., 2011). In education, systematic modelling should examine three dimensions: learner variation, learning domains, and learning activities (Casamayor *et al*., 2009; Gogoulou *et al*., 2008). Some scholars, for example, developed a higher-order item response theory framework, including overall ability on the first dimension and numerous domain abilities on the second dimension, which has been widely utilized in the automatic problem-solving process (Yang et al., 2011).

Based on the previous suggestions of Nguyen and Yang (2012), the goals of constructing an artificial intelligence-integrated system in education can be divided into four categories: classification (5 studies), matching (3 studies), recommendation (5 studies), and deep learning (10 studies). (1) Classification refers to the reconstruction of knowledge bases in which items are classified based on various qualities. Classification delineates knowledge content, which improves text analysis accuracy (Horakova *et al*., 2017). For example, some researchers created an ITS to categorize motion issues, allowing learners to access various motion problems in Mathematics conveniently (Nabiyev *et al*., 2016). (2) Matching is a conversion method in which various classification sets are linked to a certain learning aim.

For example, a text-to-diagram system for blind students was designed to link geometry words to an underlying diagram on the Braille printer, and it has been verified as an effective teaching/learning tool at a Blind school (Mukherjee et al., 2014). (3) The suggestion is viewed as an intelligent authoring tool. It might use natural language processing to automatically generate new topics, theories, and instructional elements in response to learner feedback, saving Educators time and effort (Liu *et al*., 2017). It built a human-computer connection and is frequently used to provide real-time and intelligent feedback based on student input. It is recognized as a dependable feature in modern assessment systems (Malik and Ahmad, 2017). (4) Deep learning, often known as machine learning, is a broad method for huge data processing and learning behaviour analysis. The system might self-adjust to fulfil users' dynamic requirements by upgrading its algorithms based on the growth of big data in education, such as learning or teaching behaviour (Williamson *et al*., 2018).

Some researchers have found that there is no meaningful impact on improving instruction. The difficulty was mostly attributable to poor educational design and a lack of acceptable assessment criteria (Loeckx, 2016). Therefore, future research should be anchored in learning theories for artificial intelligence to become a more satisfactory, accessible, and effective component of learners' lives.

Extraction

Educators have begun investigating appropriate applications of artificial intelligence techniques in their teaching. Some artificial intelligence applications have achieved the integration of technique, domain knowledge, and pedagogical design. This evaluation identified three types of pedagogical artificial intelligence applications: feedback (16 studies), reasoning (10 studies), and adaptive learning (9 studies). While these applications may be interconnected, they were categorized as such based on the classification explained by the authors of the reviewed articles.

Feedback or response

One of the obstacles to personalised learning is improper content sequencing. The restructuring of presentation sequences attempts to redefine knowledge organization based on the student's reaction. Feedback is important for meeting learners' proximal learning patterns (Melo *et al*., 2014). Using an artificial neural network, the system provides immediate feedback based on student input to assist them in gradually gaining access to abstract concepts and performing practical exercises. Furthermore, researchers noticed a positive trend toward the system, which could be attributed to two factors. The first one, according to Ohlsson's theory, students can learn from feedback generated due to an error (Tufekçi and Kose, 2013). In a physical teaching environment, the teacher could interact with students immediately if problems arose. However, such just-in-time interaction is difficult to achieve in an online setting. The situation necessitates the use of intelligent algorithms to provide automated feedback. For example, the intelligent virtual laboratory was created with the help of pedagogical agent-based cognitive architecture to provide appropriate feedback to students who encounter difficulties in the laboratory (Munawar *et al*., 2018). A learning website, Jutge.org, was also created with the features of a rich and well-organized problem repository. The website provides immediate feedback and assists students in solving problems and learning from their mistakes (Petit *et al*., 2018).

The second one is that immediate feedback encourages active training in interactive learning environments, which benefits learner comprehension diagnosis (Zipitria *et al*., 2013). The previous study combined speech recognition, natural language processing, and machine learning to measure the quality of classroom talk. New forms of interaction were created to provoke thoughts and further shape the effective interaction of the learning environment (Kelly *et al*., 2018). Another artificial intelligence system used path traversal algorithms to create causal chains, which provided students with elaborate feedback and hints rather than correct answers. The learning-by-teaching context was built through learners' self-organization of interactions and their interpretation of feedback (Chin *et al*., 2010).

Although many benefits were reported about automated feedback of domain knowledge, no article in this evaluation established a link to pedagogical theories. The majority of the authors in the development dimensions were from the computer science domain, which led to a technical focus on the presentation of source data (domain knowledge) with little pedagogical consideration.

AI-based reasoning

Because human-computer interaction can instil in students a sense of responsibility for improving the construction of knowledge repositories, recursive feedback may foster learners' abilities to reason in specific ways (Chin *et al*., 2013). However, some researchers discovered that novices, such as students and preservice Educators, had a poor understanding of the system's invisible causal behaviours compared to experts and experienced Educators (Vattam et al*., 2011)*. Another study came to a similar conclusion: students could learn the relevant facts and pairwise relations, but they could not reason with them very well (Chin *et al*., 2013). One possible explanation is that reasoning is largely invisible, making it difficult to induce reasoning processes through observation of behaviour. Artificial intelligence techniques, such as visualization, could improve learners' reasoning.

The graph structure (Nabiyev *et al*., 2016) and learners' engagement (Vattam *et al*., 2011) techniques have been studied to assist learners in improving their reasoning. Intelligent systems could be created for the graph structure to make thinking visible. The artificial intelligence simulation approach mimics thoughts, visually tracking reasoning in real-time. The argument-mapping tools, for example, were created to help students visualise the premises and conclusions of arguments. The findings revealed that a chain of connected ideas was chained together for learners to reach an ultimate conclusion (Rapanta and Walton, 2016). Vattam *et al*. (2011) reported that engaged learners could better understand the multiple levels of organization in complex systems by drawing on sociocultural learning theories when designing artificial intelligence to support students' reasoning. As a result, student engagement is critical when planning a learning system that aims to support reasoning.

The intelligent system's hierarchical reasoning aided students' learning. Firstly, it may assist learners in elucidating the relationships between the subcomponents of a specific topic. The intelligent reasoning system can evaluate whether the student has captured enough concepts for the given topic (Jain *et al*., 2014). Second, the system could provide an argumentative interaction, which was critical in creating a collaborative learning environment. Because of their peers' reasoning, learners tend to externalise their arguments and improve their premises. Jain *et al*. (2014) combined a visualised mapping tool with collaboration scripts. The design successfully assisted students in analysing and evaluating opposing viewpoints on contentious issues. Researchers generally regarded reasoning visualisation tools as valuable scaffolds for developing learners' critical thinking and writing skills (Rapanta and Walton, 2016).

However, artificial intelligence techniques like visualisation and hierarchical reasoning modelling may not support reasoning. The four studies examined focused on modelling to support general reasoning, whereas the reasoning model should be largely domain-specific (Chin *et al*., 2013; Rapanta and

Walton, 2016; Vattam *et al.*, 2011; Wegerif *et al.*, 2010). Furthermore, coding learners' behaviours is an unresolved challenge regarding artificial intelligence-supported reasoning. The reasoning process may be more effective when learners' personalized performance is considered. Although the visualized reasoning tools may perform well in a small-scale group setting, obtaining adequate data analysis from a large population is difficult because the reasoning system fails to adjust itself automatically. As a result, the requirements of dealing with increasingly large and diverse data sets necessitate self-adaptive solutions (Melo *et al.*, 2014).

Flexible learning

According to the new decentralised theories of artificial intelligence and social cognition, the apparent complexity of learners' behaviour largely reflected the complexity of the learning environments. Educators are prompted to provide adaptive scaffolds for diverse learning environments with various types of learners. Unlike a feedback system that provides stock responses, an adaptive educational system is a formative and corrective automated system that can adjust itself (target of intervention) to suit individual learners' characteristics, needs, and preferences (pedagogical objective) (Jones, 2011). Although only three empirical studies were identified in this review, some researchers were very optimistic about the future promotion of adaptive systems in teaching and learning. Intelligent speech recognition and automated writing evaluation (Kessler, 2018) technologies have been tested with promising results. Furthermore, there is substantial evidence that adaptive intelligence improves learning by automatically allowing learners to locate and access proximal educational resources regarding navigation and presentation support (Jonassen, 2011).

Previous research has emphasised the importance of the design dimension in applying adaptive systems (Walker *et al.*, 2009). To create successful adaptive systems in education, curriculum designers and system designers must incorporate modelling of the problem-solving process in the specific domain knowledge and big data (Kessler, 2018; Magnisalis e*t al.*, 2011). The adaptive system's mechanism connects learners' prior domain knowledge and the evaluation of their current domain performance to scaffold their problem-solving (Samarakou *et al.*, 2018). In adaptive intelligent contexts, pedagogical design is especially important. It entails the selection of adaptive algorithms and considerations for the compatibility of the learning style and the intelligence-supportive methods. In this sense, the assumption that artificial intelligence would endanger Educators' jobs may be unfounded due to Educators' critical role as curriculum designers. Second, big data is used to empower the adaptive system. Because personalization is the primary feature of the adaptive learning system, the collection of big data, such as a wide range of diverse individual characteristics, learning styles and preferences, is required for intelligent personalization to be realized. However, research on personalization in the context of adaptive systems has been limited to user characteristics related to domain knowledge. The deeper internal factors, such as human mental status and creativity, were barely noticed and studied (Magnisalis et al., 2011). This, however, has significant research potential with the development of advanced artificial intelligence techniques such as biofeedback techniques.

Application

The scope of application technology adoption emphasises the significance of including human affection in applying artificial intelligence in education. According to recent research, preference is increasingly being reported to significantly influence decision-making, perception, and learning (Ammar *et al*., 2010). Previous research on measuring learning performance only focused on two dimensions: learning outcomes and perceptions, with other aspects receiving less attention. With the maturation of biofeedback techniques such as eye tracking and EEG, affection computing has been increasingly used to investigate students' internal motivations for learning, such as creativity and responsibility (Arroyo *et al*., 2009).

According to the content analysis of the selected papers, five typical artificial intelligence techniques are preferred in computing and research in the education sector. They are complex algorithms, visualization, virtual/augmented/mixed reality, wearable techniques, and neuroscience. They supported each other in many situations to construct a smart learning environment and system. (1) Complex algorithms were designed considering human factors rather than the simple combination of functional blocks. From the human-computer interaction perspective, learners should be treated as knowledge creators rather than receivers, which helps generate a positive preference. From the standpoint of presentation modes, the traditional declarative statements in a computer system should be replaced by more diversified verbal presentations such as dialogue, coaching, and generality. (2) Visualisation was seen as an optimal method for solving complex conceptions. One of the benefits of visualization is making complex knowledge entertaining, such as game-based learning, in which learners' motivation will be greatly generated. (3) Virtual/augmented/mixed reality, including virtual/augmented/mixed reality, provides a highly simulated learning context, which may be challenging to realize in physical classrooms. For example, to help learners understand complex landforms in geography, virtual/augmented/mixed reality indulges students into a lively and creative status. (4) the wearable technique, such as Google glasses, helps to integrate learning activity into somatosensory moves. Although it was still in an exploratory period, it has great potential to advance domain knowledge in a practical context in daily life. (5) Modern neuroscience exploits how the brain works and expands the learning research to include the learners' physiological state. Research in this area would enrich the understanding of individual variations and provide additional avenues to match instruction with the most optimal guidance.

Four categories of the scope of application or learning could be identified. These are biofeedback, roleplaying, immersive learning, and gamification. Affection computing analyses human emotions and feelings captured by physical sensors and affective algorithms, which have received much attention in recent years. Affection computing improved human-computer interaction. Based on facial recognition, some researchers improved an intelligent tutoring system that detected students' emotional states to provide them with timely, dynamic feedback (Lin *et al*., 2012). Two essential aspects are required to optimize the affection computing technique: first, Educators must make timely appropriate instructional adjustments based on learners' affective status; second, comprehensive operation of multimode affection sources is required because a single source is unlikely to provide accurate affection analysis. For example, the eye-tracking technique could capture learners' eye fixation to track the attended area. Still, the foci could be attributed to different affections such as interest, anxiety, or even distraction. An additional data source, such as EEG, could aid in making a more accurate assessment (X. Zhai *et al*., 2018).

Roleplaying is a learning method that encourages students to think about problems while assuming different roles. Some algorithms were created by integrating roleplay into the pedagogical design, where students are taught by an intelligent agent rather than the learning system (Chin et al., 2013). Roleplaying can increase learners' investment in their interactions with computers. Furthermore, learners' sense of responsibility was directed toward the intelligent agent, consistent with Chase et al.'s research, which demonstrated that students might work harder on behalf of their agents than they would for themselves (Chase *et al*., 2009). Additionally, the politeness presentation mode was used in intelligent tutoring systems to motivate students to act as a companion to an intelligent agent, which was observed to benefit needy students (McLaren *et al*., 2011). Future roleplay research may allow students to customize their roles and target agents.

Immersive learning is a method that allows students to customize scenes of characters engaging in full-view learning environments. The advancement of virtual/augmented/mixed reality, 3D graphics, and wearable devices could promote learning performance, and these are strongly related to immersive affection, which generates students' academic performance and positive perceptions, such as excitement, enthusiasm, and creativity. Learners, for example, may experience a high level of excitement in the immersive learning environment. Immersive environments can also be combined with immersive collaboration using gestures, emotions, and nonverbal communication (Ijaz *et al*., 2017). Immersive learning may also reduce students' fear of complex topics and technical concepts when exposed to simulated technological and computing issues (Ngai *et al*., 2010).

Most importantly, many immersive learning tools encourage learners' enthusiasm to create and change environments, which may foster creativity (Albin-Clark *et al.*, 2011). However, few studies have found domain knowledge to be a variable. One possible explanation is that many immersive learning tools were in the exploratory stage. More research in specific domains is desperately needed.

Gamification has emerged as an important theoretical concept in the education sector. The most successful educational games tightly integrate pedagogical design, domain knowledge, and preferred elements with gameplay. Artificial intelligence has aided in integrating the game and knowledge domain, and the game's future potential is to dynamically adapt to the learners' behaviours and preferences (Thomas andYoung, 2010). Minecraft Edu is one example of properly integrating domain knowledge with affection. This is a historical simulation game in which students can learn about historical figures and events or gain insight into the spread of epidemics. Learners could gain access to historical events with authentic emotions in real-time interaction, and the collateral feeling would help them better understand the specific content knowledge (Loeckx, 2016).

Another example used a game reward system as a motivational mechanism to promote voluntary and proactive learning. The findings indicated that the reward system was a good fit with the pedagogical design, and future educational algorithms may be better associated with the field of artificial intelligence to motivate emergent learning (Moon *et al*., 2011).

Qualitative research on AI

According to selected qualitative research, the evolution of artificial intelligence in education went through a process of moving from theoretical analysis to a specific practice field and then back to review. Qualitative research supported the development of technical development research throughout the process. For example, in 2011 and 2012, qualitative research on decentralised theory (Jones, 2011) and collective intelligence (Wong and Looi, 2012) appeared, and then real artificial intelligence research began. Artificial intelligence algorithms were not mature initially, whereas advanced intelligent algorithms are usually based on big data technology and can constantly learn and improve in massive data. Big data must be decentralised and group-oriented. As a result, it is believe that early theoretical research played a significant supporting role.

The exploration of artificial intelligence (AI) in education has evolved significantly, with qualitative research playing a crucial role in bridging theoretical frameworks and practical applications. Early studies, such as those by Jones (2011) on decentralised theory and Wong and Looi (2012) on collective intelligence, laid the groundwork for understanding how AI could be integrated into educational contexts. These studies emphasised the importance of a decentralised approach to educational AI, where learning is collaborative and intelligence is distributed across networks rather than centralized in a single entity.

Jones (2011) investigated how decentralised educational systems could foster a more participatory learning environment where AI aids in distributing learning resources and collaborative problem-solving. This research was pivotal in demonstrating the potential of AI to support a more democratized form of education, aligning with the principles of collective intelligence.

Wong and Looi (2012) delved into collective intelligence within educational settings, showcasing how AI can facilitate group learning and knowledge sharing, thereby enhancing collective learning outcomes. Their work provided empirical evidence of AI's capacity to improve collaborative learning experiences through intelligent support systems that adapt to the group's dynamic learning needs.

As AI algorithms matured, they became more adept at handling big data, leading to more sophisticated applications in education. The transition from basic AI tools to advanced systems capable of learning and adapting from vast data sets marked a significant evolution in educational AI research. This progression underscored the importance of big data in developing AI tools responsive to the diverse and dynamic nature of academic environments.

These foundational studies underscore the iterative process of AI research in education, where early theoretical insights pave the way for practical applications, informing further theoretical exploration. This cyclical process has been instrumental in advancing the field of AI in education, leading to more nuanced and effective AI applications responsive to the complexities of educational contexts.

Emerging areas of research and developments on Artificial Intelligence in education

The exploration of artificial intelligence (AI) in education is progressing towards more integrated, interactive, and adaptive systems. The shift towards Internet of Things (IoT) technologies, collective intelligence, and advanced computational methods like neural networks, machine learning, and deep learning is set to redefine educational landscapes.

Existing research has primarily focused on virtual online systems, with the Internet of Things (IoT) receiving less attention. Future educational research should also look into learners' biofeedback. According to the papers reviewed, most artificial intelligence technology in education (103 out of 109) focused on online information technology or systems, such as intelligent tutoring systems, intelligent virtual laboratories, and assessment systems. Only one study (Ngai *et al*., 2010) examined learners' biofeedback using a wearable circuit. This could be attributed to the well-established smart online system, which is easier to build on and cost-effective.

On the other hand, the Internet of Things holds great promise for catering to diverse learning contents and skills. It can improve students' spatial and mechanical understanding of physical construction processes in science classes. In two qualitative studies (Seni, 2012; Williamson *et al*., 2018), the Internet of Things technology optimises human cognition and performance by simulating brain functions in a physical context to sense and understand human cognitive behaviours. Although no empirical studies were found in the selected papers to test the effect of IoT techniques on education, IoT with low costs and wearable computing devices could be a future area of artificial intelligence development in education. This is consistent with the 2019 Horizon report.

The IoT is increasingly recognized for its potential to transform educational environments. Its ability to enhance spatial and mechanical understanding through interactive and immersive learning experiences is notable. Integrating IoT in science education, for instance, can provide hands-on learning through simulations of physical construction processes, thus enriching the learning experience. As Ngai *et al*. (2010) demonstrated, wearable IoT devices can be pivotal in analysing learners' biofeedback, offering insights into the physiological aspects of learning.

Collective intelligence has emerged as a critical artificial intelligence development direction in which the roles of Educators and students will be fundamentally altered. According to the selected papers, the decentralized theory was first investigated in education in 2011 (Jones, 2011), followed by the introduction of Collective intelligence in education in 2012 (Wong and Looi, 2012). However, no empirical study has examined how Educators and students deal with the challenges of Collective intelligence. According to the characteristics of Collective intelligence, the following two topics are predicted to become research trends. Firstly, collective intelligence does not rely on centralized control of individual behaviours. In this situation, learners transition from knowledge absorbers to creators. They actively constructed knowledge by interacting with the system in various contexts. Educators' "authorities" may be challenged by a group of experienced practitioners, such as engineers and farmers, and a collaborative curriculum design would be built by a Collective intelligence system (Jonassen, 2011).

Furthermore, collective intelligence may shift educators' responsibilities from knowledge transmission to knowledge organization. Previous research has suggested that educators investigate the impact of crowdfunding or crowdsourcing on education and how educators will perform their organizing abilities in the future (Dai *et al.*, 2020). However, the investigation from the Educators' perspective is still insufficient and requires further research. Second, Collective intelligence aided adaptability in dynamic or unstable environments. Collective agents typically exchange information by leaving marks and observing the activities of their peers. For example, the best solution in the present moment may become unavailable. As a result, it is suggested that further research be conducted into how artificial intelligence performs dynamic recommendations for students at various stages of learning (Wong and Looi, 2012).

Collective intelligence emphasizes collaborative learning, where knowledge is co-created rather than transmitted unidirectionally from educators to students. The transformation from centralized to decentralized learning environments supports dynamic and context-aware educational models, promoting adaptability and personalized learning experiences. This trend, initially explored by Jones (2011) and furthered by Wong and Looi (2012), indicates a significant shift towards leveraging collective knowledge and expertise in educational settings.

Neural Networks, Machine Learning, Deep Learning and Advanced Computational Methods will reshape human-machine interactions in the future. Human-computer interaction trends will no longer be based on a human's perspective of machine operation. Instead, the machine can improve predictions by learning from big data without being explicitly programmed. Deep learning studies were first mentioned in the selected papers in 2017 (Malik *et al*., 2017; Malik and Ahmad, 2017). One empirical study by Kelly *et al*. (2018) focused on deep learning technology modelling scoring-based data. However, data based on human physical characteristics were less noticeable. Pearson and IBM have proposed investigating neurocomputation brain-based educational technologies based on the neuroscientific understanding of the brain (Williamson *et al*., 2018). However, only two qualitative studies (Seni, 2012; Williamson *et al*., 2018) proposed the integration of neuroscience and artificial intelligence in the education sector. Future research trends in integrating brain function with deep learning techniques to optimize human-computer interaction could be expected. It will impact the application and integration of artificial intelligence in education, such as adaptive learning and roleplaying. This viewpoint was reported in the Horizon report in 2018. The report predicts that adaptive learning techniques will be further generalized in two to three years.

Integrating neural networks, machine learning, and deep learning into educational technologies is advancing rapidly (Anthony, 2012). These computational methods enhance educational systems' capability to provide personalized learning experiences. Predictive analytics, powered by these advanced methods, can tailor educational content to meet individual student needs, fostering a more engaging and effective learning environment. This aligns with the predictions made in the Horizon report (2018) that adaptive learning techniques underpinned by AI will become more prevalent.

All empirical studies showed that artificial intelligence techniques positively impacted education. However, the interview and review paper both brought to light the challenges or misunderstandings of artificial intelligence in education (Lawler and Rushby, 2013; Magnisalis *et al*., 2011). A holistic evaluation criterion must be developed to assess artificial intelligence's effectiveness in education. To ensure the validity and reliability of the evaluation, a multidimensional model that incorporates technique, pedagogical design, domain knowledge, and human factors should be used. Woolf's Roadmap for Education Technology predicted that in the age of artificial intelligence Educational Data Mining, students' knowledge, progress, learning environments, and the success and failure of teaching strategies could be tracked chronologically (Woolf, 2010).

Furthermore, current research is overly focused on specific educational contexts and a small number of variables. Most studies included students as participants. Educators received less attention; most researchers considered science, humanity, and social science subjects, while sports, arts, and special education received less attention. Only one study, for example, discovered text-to-diagram conversion as a novel teaching aid for blind learners *(Mukherjee et al.*, 2014).

Empirical research has consistently demonstrated the positive impact of artificial intelligence (AI) on education, yet it has illuminated several challenges and misunderstandings within this domain (Lawler and Rushby, 2013; Magnisalis *et al*., 2011).

To navigate these complexities, future research must aim to develop a holistic evaluation framework that assesses AI's effectiveness across diverse educational settings. This framework should integrate various dimensions such as technological innovation, pedagogical integration, domain-specific knowledge, and human-centric factors. Incorporating these aspects will enable a more comprehensive assessment of AI's role and efficacy in education, ensuring that the benefits of AI are fully realized and accessible across all learning disciplines. Woolf's (2010) roadmap emphasizes the potential of Educational Data Mining in tracking and analysing students' learning trajectories, suggesting a growing need for sophisticated AI tools that can provide nuanced insights into the educational process.

Moreover, the current focus of AI research predominantly on conventional academic disciplines highlights a gap in exploring AI's potential in sports, arts, and special education. Future studies should diversify the scope of AI applications to include these underrepresented fields, investigating how AI can enhance learning experiences, support educators, and cater to the unique needs of diverse learner populations. For instance, the development of AI-driven assistive technologies can revolutionize teaching methods for special education, offering more personalized and accessible learning solutions for students with disabilities (Mukherjee e*t al*., 2014). Similarly, AI can play a pivotal role in sports education by analysing performance data to tailor training programs and in arts education by fostering creativity through adaptive learning environments. Research in these areas should focus on technological advancement and pedagogical strategies aligning with each domain's educational goals and requirements.

Challenges encountered in adopting artificial intelligence in higher education

Technical, pedagogical and ethical challenges are often encountered un the process of AI in higher education These challenges are encountered at various points along the value chain, from system design to user experience and data handling. One technical challenge is integrating AI into existing educational infrastructures. For example, the case of the Los Angeles Unified School District's attempt to implement an AI-based scheduling system resulted in significant disruptions due to system incompatibilities and a lack of user training (Norris and Soloway, 2011). This situation underscores the importance of ensuring that AI systems are compatible with existing educational infrastructures and that adequate training is provided to all users.

The pedagogical challenges related to AI's inability to fully adapt to the diverse needs of individual learners. A study by Lewin (2013) found that AI systems often lack the nuanced understanding required to tailor learning experiences to individual student needs, leading to a one-size-fits-all approach that fails to maximise the potential benefits of personalised learning. Another pedagogical challenge is based on the role of educators in an AI-integrated educational environment. The NMC 2018 Horizon Report (Becker *et al.*, 2018) noted that educators might resist AI integration due to fear of being rendered redundant, or for fear of gradually becoming totally dependent on AI. Kessler (2018) highlights the need for professional development to prepare educators for the effective use of AI in teaching and learning processes.

An ethical challenge that is often encountered relate to maintaining privacy when handling or process student data privacy. Aoun (2017) discusses the ethical dilemmas encountered when using student data in AI applications, where the balance between personalised learning and privacy rights becomes critical. Moreover, the case of AI translators providing students with immediate answers, as discussed by Kessler (2018) illustrates how AI can undermine the learning process by under developing critical thinking and inquiry competencies. Similarly, the gamification of the learning through AI raises concerns about compromising the core educational values. Thomas and Young (2010) highlight the challenge of maintaining academic integrity while engaging students through game-based learning.

Conclusions

The categorisation of AI research into development, extraction, and application domains, as well as the identification of emerging trends like the Internet of Things (IoT), Collective Intelligence, deep learning, and neuroscience, provides a comprehensive overview of the field's evolution and potential trajectory. The findings suggest that AI can significantly contribute to educational success when its applications are carefully integrated with pedagogical practices. However, the challenges identified, particularly in aligning technical capabilities with academic needs and ethical standards, highlight the necessity for a collaborative approach between educators and AI developers. Educators must actively design develop and use AI systems to ensure these technologies are pedagogically sound and aligned with educational goals (Luckin et al., 2016; Woolf, 2010). Educators should leverage AI to personalise learning experiences, thereby addressing the diverse needs of students. AI developers, on the other hand, should focus on creating adaptable, user-friendly, and ethical AI tools that enhance teaching and learning processes. For instance, integrating AI in formative assessments can provide immediate, personalized feedback to students, aiding their learning journey (Roll and Wylie, 2016; Woolf, 2010). Furthermore, the collaboration between educators and AI engineers is crucial in navigating the ethical landscape of AI in education, ensuring that student data are handled responsibly and that AI applications are used to augment rather than replace human interactions in the learning process (Aoun, 2017; Thomas and Young, 2010).

In conclusion, the symbiotic relationship between AI and education needs to be nurtured with a focus on developing AI tools that are both technologically advanced, pedagogically relevant, and ethically sound. Future research should continue to explore this interface, with a particular emphasis on developing AI applications grounded in educational theory and practice, thereby truly enhancing students' academic success.

Limitation and future research

By relying on literature on one database, the evaluation excluded relevant literature on other databases. This selective approach limited the range of perspectives captured in the evaluation and thereby potentially skewing the perceived trends (Borgman, 2017). The constrained keyword selection, limited to terms directly related to the Social Science Citation Index, narrowed the scope significantly. Expanding the keyword list to include terms like "adaptive learning" and "tutor systems" could have revealed a broader range of AI applications and methodologies in education, thereby providing a more comprehensive view of the field (Settouti *et al*., 2016).

Future evaluation should use a multi-database approach to mitigate these limitations. Future research should also consider historical analyses to track the evolution of AI in education, identifying shifts in trends, methodologies, and applications since the advent of AI. This longitudinal perspective could offer valuable insights into the progress and transformation of AI technologies in educational contexts, highlighting areas of significant change or potential for future development (Weller, 2020).

References

Albin-Clark, A., Howard, T. L. J. and Anderson, B. 2011. Computer graphics simulation of blockplay in early childhood. *Computers and Education*, *57*(4), 2496–2504.

Ammar, M. B., Neji, M., Alimi, A. M. and Gouarderes, G. 2010. The affective tutoring system. *Expert Systems with Applications*, *37*(4), 3013–3023.

Anthony, G. P. 2012. The evolution of big data and learning analytics in american higher education. *Journal of Asynchronous Learning 16*(3).

Aoun, J. E. 2017. *Robot-proof: Higher Education in the Age of Artificial Intelligence*. MIT Press.

Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K. and Christopherson, R. 2009. Emotion sensors go to school *Artificial Intelligence in Education*, *200*, 17–24.

Becker, S. A., Brown, M., Dahlstrom, E., Davis, A., DePaul, K., Diaz, V. and Pomerantz, J. 2018. *NMC Horizon Report: 2018 Higher Education Edition*.

Borgman, C. L. 2017. *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press.

Boulay, B. 2016. Artificial Intelligence as an effective classroom assistant. *IEEE Intelligent Systems*, *31*(6), 76–81.

Casamayor, A., Amandi, A. and Campo, M. 2009. Intelligent assistance for teachers in collaborative e-learning environments. *Computers and Education*, *53*(4), 1147–1154.

Chase, C. C., Chin, D. B., Oppezzo, M. A. and Schwartz, D. L. 2009. Teachable agents and the protege effect: increasing the effort towards learning. *Journal of Science Education and Technology,*, *18*(4), 334–352.

Chaudhry, M. A. and Kazim, E. 2022. Artificial intelligence in education (AIEd): a high-level academic and industry note. *AI Ethics*, *2*, 157–165. https://doi.org/https://doi.org/10.1007/s43681-021-00074-z

Chin, D. B., Dohmen, I. M., Cheng, B. H., Oppezzo, M. A., Chase, C. C. and Schwartz, D. L. 2010. Preparing students for future learning with teachable agents. *Educational Technology Research and Development*, *58*(6), 649–669.

Chin, D. B., Dohmen, I. M. and Schwartz, D. L. 2013. Young children can learn scientific reasoning with teachable agents. *IEEE Transactions on Learning Technologies*, *6*(3), 248–257.

Creswell, J. W. and Poth, C. N. 2018. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (5th ed.). Sage Publications.

Dai, Y., Chai, C. S., Lin, P., Jong M., Guo, Y. and Qin, J. 2020. Promoting students' wellbeing by developing their readiness for the artificial intelligence age. *Sustainability*, *12*(16), 1–15.

Fenwick, T. 2018. Pondering purposes, propelling forwards. *Studies in Continuing Education*, *40*(3), 367–380.

Flogie, A. and Aberšek, B. 2015. Transdisciplinary approach of science, technology, engineering and mathematics education. *Journal of Baltic Science Education*, *14*(6), 779–790.

Ge, X., Yin, Y. and Feng, S. 2018. Application research of computer artificial intelligence in college student sports autonomous learning. *Kuram Ve Uygulamada Egitim Bilimleri,* , *18*(5), 2143–2154.

Gogoulou, A., Gouli, E. and Grigoriadou, M. 2008. Adapting and personalising the communication in a synchronous communication tool. *Journal of Computer Assisted Learning*, *24*(3), 203–216.

Higgins, S., Xiao, Z. and Katsipataki, M. 2019. The impact of digital technology on learning: a summary for the Education Endowment Foundation. *Education Endowment Foundation*.

Horakova, T., Houska, M. and Domeova, L. 2017. Classification of the educational texts styles with the methods of artificial intelligence. *Journal of Baltic Science Education*, *16*(3), 324–336.

Ijaz, K., Bogdanovych, A. and Trescak, T. 2017. Virtual worlds vs books and videos in history education. *Interactive Learning Environments*, *25*(7), 904–929.

Jain, G. P., Gurupur, V. P., Schroeder, J. L. and Faulkenberry, E. D. 2014. Artificial Intelligence-based Student Learning Evaluation: A concept map-based approach for analyzing a student's understanding of a topic. *IEEE Intelligent Systems Transactions on Learning Technologies*, *7*(3), 267–279.

Jonassen, D. H. 2011. Ask systems: Interrogative access to multiple ways of thinking. *educational technology research and development*, *59*(1), 159–175.

Jones, A. 2011. Philosophical and Socio-cognitive foundations for teaching in higher education through collaborative approaches to student learning. *Educational Philosophy and Theory*, *43*(9), 997–1011.

Kelly, S., Olney, A. M., Donnelly, P., Nystrand, M. and D'Mello, S. K. 2018. Automatically measuring question authenticity in real-world classrooms. *Educational Researcher*, *47*(7), 451–464.

Kessler, G. 2018. Technology and the future of language teaching. *Foreign Language Annals*, *51*(1), 205–218.

Knox, J. 2020. Artificial Intelligence and education in China. *Learning Media and Technology*, *45*(3), 1–14.

Kumar, K. and Thakur, G. S. M. 2012. Advanced applications of neural networks and artificial intelligence: a review. *International Journal of Information Technology and Computer Science*, *4*(6), 57 - 68.

Lacity, M. and Willcocks, L. P. 2017. *Robotic Process Automation and Risk Mitigation: The Definitive Guide*. SB Publishing.

Lawler, R. W. and Rushby, N. 2013. An interview with Robert Lawler. *British Journal of Educational Technology*, *44*(1), 20–30.

Lewin, T. 2013. After setbacks, online courses are rethought. *The New York Times*.

Lin, H. C. K., Wang, C. H., Chao, C. J. and Chien, M. K. 2012. Employing Textual and facial emotion recognition to design an affective tutoring system. *Turkish Online Journal of Educational Technology-TOJET*, *11*(4), 418–426.

Liu, M., Rus, V. and Liu, L. 2017. Automatic Chinese factual question generation. *IEEE Transactions on Learning Technologies*, *10*, 194–204.

Loeckx, J. 2016. Blurring boundaries in education: context and impact of MOOCs. *The International Review of Research in Open and Distributed Learning*, *17*(3), 92–121.

Luckin, R., Holmes, W., Griffiths, M. and Forcier, L. B. 2016. *Intelligence Unleashed: An argument for AI in Education*. Pearson Education.

Magnisalis, I., Demetriadis, S. and Karakostas, A. 2011. Adaptive and Intelligent systems for collaborative learning support: a review of the field. *IEEE Transactions on Learning Technologies*, *4*(1), 5-20.

Malik, Mir, R. R., Farhan, M., Rafiq, T. and Aslam, M. 2017. Student query trend assessment with semantical annotation and artificial intelligent multi-agents. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(7), 3893–3917.

Malik, K. R. and Ahmad, T. 2017. E-Assessment data compatibility resolution methodology with bidirectional data transformation. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(7), 3969–3991.

McLaren, B. M., DeLeeuw, K. E. and Mayer, R. E. 2011. Polite webbasedintelligent tutors: can they improve learning in classrooms? *Computers and Education*, *56*(3), 574–584.

Melo, F. R., Flores, E. L., Carvalho, S. D., Teixeira, R. A. G., Loja, L. F. B. and de Sousa Gomide, R. 2014. Computational organisation of didactic contents for personalised virtual learning environments. *Computers and Education*, *79*, 126–137.

Merriam, S. B. and Tisdell, E. J. 2011. *Qualitative Research: A Guide to Design and Implementation* (4th ed.). Wiley.

Moon, M. K., Jahng, S. G. and Kim, T. Y. 2011. A computerassisted learning model based on the digital game exponential reward system. *Turkish Online Journal of Educational Technology-TOJET*, *10*(1), 1–14.

Mukherjee, A., Garain, U. and Biswas, A. 2014. Experimenting with automatic text-to-diagram conversion: a novel teaching aid for the blind people. *Journal of Educational Technology and Society*, *17*(3), 40–53.

Munawar, S., Toor, S. K., Aslam, M. and Hamid, M. 2018. Move to smart learning environment: exploratory research of challenges in computer laboratory and design intelligent virtual laboratory for e-learning technology. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(5), 645–1662.

Nabiyev, V. V., Çakiroglu, U., Karal, H., Erumit, A. K. and Çebi, A. 2016. Application of graph theory in an intelligent tutoring system for solving mathematical word problems. *Eurasia Journal of Mathematics, Science and Technology Education*, *12*(4), 687–701.

Ngai, G., Chan, S. C. F., Chan, J. C. Y. and Lau, W. W. Y. 2010. Deploying a wearable computing platform for computing education. *IEEE Transactions on Learning Technologies*, *3*, 45–55.

Nguyen, B. A. and Yang, D. L. 2012. A Semi-automatic approach to construct vietnamese ontology from online text. *The International Review of Research in Open and Distributed Learning*, *13*(5), 148–172.

Norris, C. and Soloway, E. 2011. Tips for a Successful implementation of mobile learning in schools. *Educational Technology and Society*, *51*(5), 13-15.

Petit, J., Roura, S., Carmona, J., Cortadella, J., Duch, J., Gimnez, O., Mani, A., Mas, J., Rodriguez-Carbonell, E., Rubio, E., de San Pedro, E. and Venkataramani, D. 2018. Jutge.org: Characteristics and Experiences. *IEEE Transactions on Learning Technologies*, *11*(3), 321–333.

Rapanta, C. and Walton, D. 2016. The Use of Argument Maps as an Assessment Tool in Higher Education. *International Journal of Information Technology and Computer Science of Educational Research.* , *79*, 211–221.

Roll, I. and Wylie, R. 2016. Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, *26*, 582-599.

Samarakou, M., Tsaganou, G. and Papadakis, A. 2018. An e-learning system for extracting text comprehension and learning style Characteristics. *Journal of Educational Technology*, *21*(1), 126–136.

Seldom, A. and Abidoye, O. 2018. *The Fourth Education Revolution: Will Artificial Intelligence Liberate or Infantilise Humanity?* University of Buckingham Press.

Seni, D. A. 2012. Do the modern neurosciences call for a new model of organisational cognition? *Science and Education*, *21*(10), 1485–1506.

Settouti, L. S., Prié, Y., Marty, J. C. and Mille, A. 2016. A trace-based approach to identifying users' engagement and qualifying their engaged-behaviours in interactive systems: application to a social game. *User Modeling and User-Adapted Interaction*, *26*(5), 413-451.

Spector, J. M. and Martha, C. P. 1993. Automating instructional design: concepts and issues. In D. J. Muraida (Ed.), *Automating Instructional Design: Concepts and Issues*. Educational Technology Publications.

Stokes, P. and Urquhart, C. 2013. Qualitative interpretative categorisation for efficient data analysis in a mixed methods information behaviour study. *Information Research*, *18*(1). http://InformationR.net/ir/18-1/paper555.html

Thomas, J. M. and Young, R. M. 2010. Annie: automated generation of adaptive learner guidance for fun serious games. *IEEE Transactions on Learning Technologies*, *3*(4), 329–343.

Tierney, P. 2012. A Qualitative analysis framework using natural language processing and graph theory. *International Review of Research in Open and Distributed Learning*, *13*(5), 173–189.

Tufekçi, A. and Kose, U. 2013. Development of an artificial intelligence based software system on teaching computer programming and evaluation of the system. *Hacettepe Universitesi Egitim Fakultesi Dergisi*, *28*(2), 469–481.

Vattam, S., Goel, A. K., Rugaber, S., Hmelo-Silver, C. E., Jordan, R. C., Gray, S. A. and Sinha, S. 2011. Understanding complex natural systems by articulating structure-behaviour function models. *Educational Technology and Society*, *14*(1), 66–81.

Walker, E., Rummel, N. and Koedinger, K. R. 2009. Beyond Explicit feedback: new directions in adaptive collaborative learning support. 9th International Conference on Computer Supported Collaborative Learning, Maratea, Italy.

Wegerif, R., McLaren, B. M., Chamrada, M., Scheuer, O., Mansour, N., Miksatko, J. and Williams, M. 2010. Exploring creative thinking in graphically mediated synchronous dialogues. *Computers and Education*, *54*(3), 613–621.

Weller, M. 2020. *25 Years of Ed Tech*. Athabasca University Press.

Wiesner, C. 2022. Doing qualitative and interpretative research: reflecting principles and principled challenges. *Political Research Exchange*, *4*(1). https://www.tandfonline.com/doi/full/10.1080/2474736X.2022.2127372

Williamson, B., Pykett, J. and Nemorin, S. 2018. Biosocial Spaces and Neurocomputational Governance: Brain-based and Brain-targeted Technologies in Education. *Discourse: Studies in the Cultural Politics of Education*, *39*(2), 258–275.

Wong, L. H. and Looi, C. K. 2012. Swarm Intelligence: New Techniques for Adaptive Systems to Provide Learning Support. *Interactive Learning Environments*, *20*(1), 19–40.

Woolf, B. P. 2010. *A Roadmap for Education Technology*.

Wu, Y. T., Hou, H. T. and F.K., H. 2013. A Review of Intervention Studies on Technology-assisted Instruction from 2005–2010. *Journal of Educational Technology and Society*, *16*(3), 191–203.

Zhai, Q., Fang, Y., Dong, Y., Wei, Z., Yuan, J., Cacciolatti, L. and Yang, Y. 2018. The effects of biofeedback-based stimulated recall on self-regulated online learning: a gender and cognitive taxonomy perspective. *Journal of Computer Assisted Learning*, *34*(6), 775–786.

Yang, C. W., Kuo, B.C. and Liao, C. H. 2011. A HO-IRT based diagnostic assessment system with constructed response items. *Turkish Online Journal of Educational Technology-TOJET*, *10*(4), 46–51.

Zipitria, I., Arruarte, A. and Elorriaga, J. 2013. Discourse measures for basque summary grading. *Interactive Learning Environments*, *21*(6), 528–547.

Enhancing Teaching and Learning in South African Higher Education Institutions through the Adoption and Use of Generative Artificial Intelligence

N. Wayi-Mgwebi
University of Mpumalanga
N.Wayi-mgwebi@ump.ac.za

Abstract

This paper examines how generative artificial intelligence is changing higher education, particularly in universities in South Africa. Using a systematic literature review approach, the paper explores how generative artificial intelligence can enhance teaching and learning experiences in South African higher education institutions. It examines the state of higher education in South Africa today, highlighting developments and obstacles, and also gives examples of generative AI tools that are currently in use in the higher education sector. Challenges affecting teaching and learning in higher education institutions are presented and suggestions on how generative artificial intelligence could assist in addressing the challenges. The potential of generative artificial intelligence in this regard is demonstrated using generative artificial intelligence tools like Python-Bot and Jill Watson.

Keywords: Chatbots, Fourth Industrial Revolution, generative artificial intelligence, higher education, learning management system, teaching and learning

Introduction and background

Human life has fundamentally changed as a result of the Fourth Industrial Revolution, which is driven by rapid technological developments such as artificial intelligence (AI), virtual reality, blockchain, robotics, cloud computing, data science, and 3D printing (Al-Maskari *et al*., 2022). AI, in particular, has become important across industries leading to some people believing that computers are going to replace human power in diverse jobs. AI is a branch of computer science focused on the theory and creation of computer systems that can carry out operations that would typically need human intellect, like speech recognition, visual perception, decision-making, and language translation (Luebering J, 2023). Despite the fact that AI has been around for decades, from as far back as the 1940s. The field was, however, officially formalised in the 1950s with the proposal of what is now known as the Turing Test to assess a machine's ability to exhibit intelligent behaviour indistinguishable from that of a human being (Turing, 1950). Within AI, generative AI is a rapidly growing field within AI that has seen significant expansion in recent years, causing disruptions across various industries (Mannuru *et al*., 2023). Generative AI is a type of AI that is proficient in generating diverse data forms (for example, images, videos, audio, text, and 3D models) by learning patterns from existing data and applying this knowledge to produce novel and intricate outputs, exhibiting a capacity for creating remarkably realistic and complex content akin to human creativity (Marr, 2023).

A paradigm shift has occurred with the introduction of AI into a number of fields, enabling computers to perform tasks jobs and activities on their own. AI systems are able to evaluate data, identify patterns, and make choices without continual human supervision by utilising sophisticated algorithms and machine learning. This capability has several uses, ranging from the automation of repetitive tasks in many industries to the facilitation of complex problem-solving in a variety of domains. The Fourth Industrial Revolution is still happening, and AI technology is making great progress in enabling machines to handle difficult jobs autonomously.

The higher education sector is one of the fields that has been deeply impacted by technological developments, particularly generative AI. For example, generative AI is expanding access to education for those who would otherwise be left out due to geographical location (Guerrero-Quiñonez *et al*., 2023). Other beneficial impacts of AI in higher education include enabling efficiency in teaching and learning, provision of student support service, and administration (Bond and De Laat, n.d.; Kumar *et al*., 2021; Venugopal and Mamatha, 2023). These have been welcome developments in higher education until the advent of Chat Generative Pre-trained Transformer (ChatGPT), a new generative AI tool, at the end of 2022.

The introduction of ChatGPT marked a pivotal moment in the discourse surrounding generative AI, prompting widespread contemplation about how it would impact various aspects of people's lives (Gulya Jason, 2023). Since then, ChatGPT has grown to be a massively popular individual and commercial tool. A recent study conducted among undergraduate students from South African higher education institutions revealed that there is generally a positive perception about the role that generative AI plays towards the achievement of study outcomes, including using the tools as a simpler reading aid compared to lengthy textbooks (Bosch et al., 2023). Academics from South African institutions believe that the key issue has been the threat of tools like ChatGPT to assessments, which calls for innovative and improved assessment approaches (Tarisayi, 2024).

Problem statement

A few weeks following the introduction of ChatGPT, there were widespread concerns about how it and similar technologies could significantly disrupt sectors of the economy including higher education. The possibility that students could use generative AI technologies to cheat or plagiarise in their written assignments and tests was one of the key concerns (Chan, 2023a). According to a survey conducted among students, thousands of university students confirmed that students were aware of the improper use of generative tools like ChatGPT, but they were committed to continuing to use them, nevertheless (Chan, 2023b; Sullivan *et al*., 2023). ChatGPT significantly disrupted the status quo with some suggesting a complete ban of it and similar technologies (Chan, 2023b; Gimpel *et al*., 2023) because they were perceived as a threat to the integrity of higher education integrity (Gimpel *et al*., 2023; Sullivan et al., 2023). These actions have triggered a debate on whether ChatGPT should be adopted or renounced by higher education institutions.

At the same time, it is believed that AI and related technologies have the potential to bridge or worsen the existing education disparities (Bosch *et al*, 2023; Krish *et al*. 2023). This is because of disparities that still exist in terms of unequal access to education and levels of technological access (Bosch *et al*., 2023).

The objective of this paper is to explore how generative AI can enhance teaching and learning experiences in South African higher education institutions. Specifically, the paper analyses key challenges associated with teaching and learning in South African higher education; identifies and discuses AI tools that are deployed in higher education; and assess how generative AI will benefit South African higher education institutions by way of enhancing teaching and learning.

Methodology

This paper is an outcome of systematic literature review (SLR). SLR is a systematic research methodology used to gather and assess the body of knowledge about a specific topic (Okoli and Schabram, 2012). The primary objective of SLR is to thoroughly review all previous research on a given subject, and assess, and compile the knowledge that is currently accessible on a certain subject or issue. A literature review can address research topics with a power that no one study can match by incorporating the conclusions and points of view from numerous empirical findings (Snyder, 2019). It assists the researchers in identifying significant gaps and unanswered questions in the existing literature. It is a rigorous and time-intensive methodology (Okoli, 2015).

Okoli (2015) and Siddaway (2019) describe specific stages of an SLR to be followed and while each author's steps differ; a combined approach can be adapted to include planning, searching, extracting, and execution. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is applied to help systematic reviewers transparently report why the review was done, what the authors did, and what they found (Page *et al*., 2021).

The primary source of the literature used was the Google Scholar. Although this is not a traditional database, it still provides a broad search across various disciplines and sources within the academic community. Keywords were used to extract the required information. Additionally, Boolean operators were applied to broaden or narrow search results as needed. The following keywords were used to search for relevant sources for the study: Generative Artificial Intelligence in Higher Education, Generative Artificial Intelligence for Teaching and Learning, and AI and South African Higher Education. Artificial Intelligence was included because during initial screening it was discovered that the papers also touched on aspects relating to Generative AI.

According to Okoli (2015), one of the key steps when conducting an SLR is screening for inclusion, which requires that the reviewers be explicit about what studies they considered for review and which ones they eliminated. The majority of papers used were published between 2015 and 2023. There are, however, few papers published prior to 2015 which are used mainly for historical comparison, and some assist in establishing the background and context.

Below is the visual representation of the data extraction procedure in the shape of a PRISMA Flowchart.

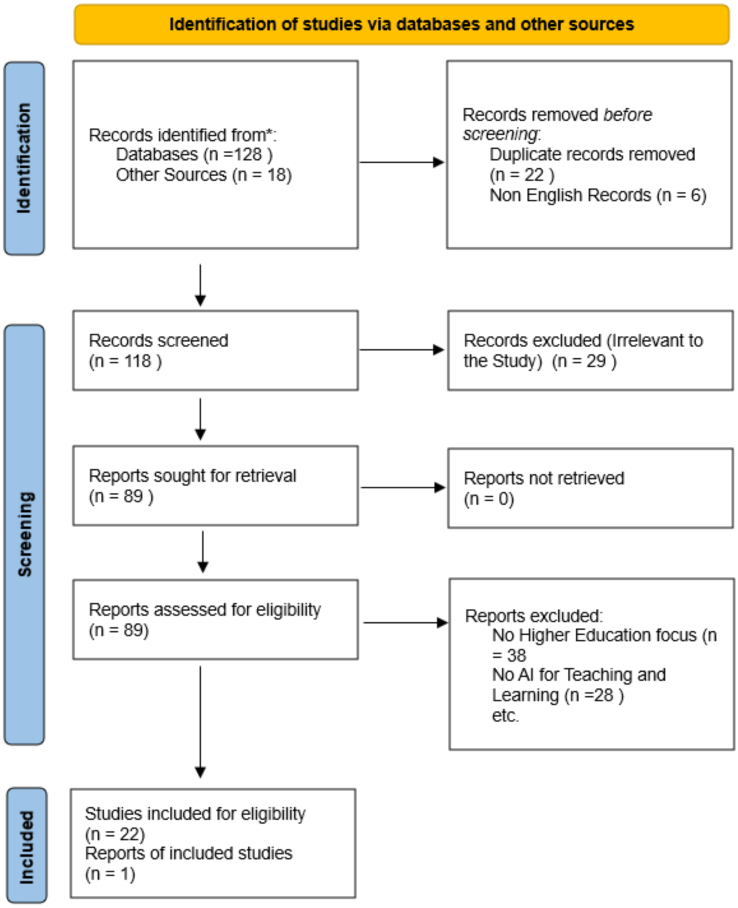

Figure 1: The PRISMA flow chart (Page *et al*., 2021)

The ensuing section of the paper presents the findings from the study, which cover the objectives of the study.

Challenges that the South African higher education faces

The higher education sector in South Africa has transformed towards greater inclusivity, featuring 26 public universities and 124 private higher education institutions, collectively enrolling 1.3 million students in undergraduate and postgraduate programs as of 2021 (DHET, 2023). This notwithstanding, the sector faces numerous challenges. Examples of such challenges include scarcity of resources, high dropout rates, and unequal access to high-quality instruction (Sibanda and Fon, 2022). Furthermore, many higher education institutions in South Africa are still struggling to keep up with the demands of a rapidly changing educational landscape (Landa *et al*., 2021). This became evident during the Covid-19 induced lockdown. As was the global case, higher education institutions in South Africa were compelled by Covid-19 lockdowns to navigate the balance between meeting students' requirements and adhering to health guidelines, which became a disadvantage for those students with limited or no access to the internet (Azionya *et al.,* 2021). This highlighted the critical role that technology plays in ensuring continuity in education. Some higher education institutions in South Africa discontinued teaching and learning for extended periods of time due to their inability to successfully make the shift to online instruction (Lubinga *et al*., 2023). What has been of great benefit was the fact that the majority of South African higher education institutions already had existing Learning Management Systems (LMSes), although they were not being used extensively. The lockdown provided a platform for wide usage of LMSes (Landa *et al*., 2021).

There has been rapid developments in technology-based teaching and learning in South African higher education institutions. However, due to inequality among the higher education institutions, some of them might easily adopt intelligent tutors, virtual assistants, and other benefits offered by generative AI whilst some institutions might not have the infrastructure and support systems to do so. An initial response for some South African institutions might be to support and enhance the current methods rather than replace them with generative AI technologies (Krish et al., 2023). The main challenges in relation to teaching and learning South African higher education institutions are briefly discussed below:

Poor preparation

A significant proportion of higher education students are from rural communities, and enrol for higher education underprepared mostly due to lack of resources in rural schools. These students battle with socio-economic hardships, including having no access to basic facilities such as electricity (Sibanda and Fon, 2022). This means that students are already at a disadvantage when they enter the higher education system.

Large classes

The challenge of large classes affects many institutions, though it can be argued that a large class is not merely about numbers but context. Teaching large classes presents multifaceted challenges, affecting educators and students alike, as individuals may become faces in a crowd, hindering meaningful connections, and hindering personalized guidance and leading to difficulties managing assignments and exams for academics (Reddy, 2017).

High dropout

High dropout rates in higher education are linked to academic struggles, financial challenges and disengagement.

Diverse learning needs

Creating an inclusive higher education environment necessitates acknowledging and addressing the disparities arising from diverse student backgrounds and preparation levels, as traditional teaching methods may fall short in accommodating varied understanding and proficiency levels. Students with disabilities are not always catered for under traditional teaching and learning system.

Language and social barriers

Language is one of challenges affecting students as English is not the first language for majority of students, particularly students from disadvantaged communities**.** Students enrolled at South African HEIs come from various racial, social, cultural and religious backgrounds (Singh, 2015). This contributes to unique teaching challenges, primarily arising from the need to address a wide range of perspectives, learning styles, and educational experiences.

Tedious assessment in teaching and learning

Traditional assessment methods, particularly manual grading, pose challenges by consuming valuable instructor time and introducing potential consistency issues in large-scale grading, impacting fairness and accuracy. In addition, the time-intensive nature of these assessments often leads to delayed feedback, hindering students' prompt identification and addressing of learning gaps.

Broader society challenges

Figure 2 below depicts general issues that are a reality for students in higher education. An example is the country's high unemployment rate (DHET, 2023), which is one of the challenges that students are always concerned about as it later affects their choices.

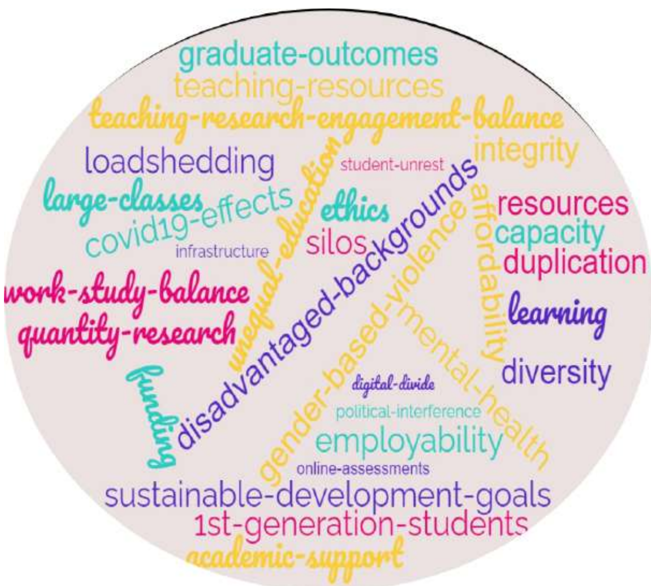

Figure 2. Mix of Reality for Higher Education Students

Generative artificial intelligence in higher education

It is commonly acknowledged that AI has the ability to significantly transform education by improving the processes of teaching and learning. When it comes to content generation, generative AI uses language models that have been trained to identify relationships between elements by analysing large data sets (Gimpel *et al*., 2023). This produces responses that fall within a probability range that has been set during training. However, the absence of explicit reasoning can produce results that are both logical and sometimes inaccurate. It becomes clear that a deeper comprehension of the tactics needed to manage the broad use of AI in higher education is important (Abgaryan et al., 2023). These chatbots have a variety of functions in higher education, including responding to students' questions, giving feedback and assessments, and fostering interaction and cooperation between students and teachers (Ilieva et al., 2023). For example, generative AI helps students identify areas of weakness and develop their skills in a highly adaptive way by giving them individualized feedback and support (Kasneci et al., 2023). In addition to providing students with adaptable tools catered to their specific talents, this personalised educational approach fosters a more positive and productive learning environment.

Through a variety of channels, educators are positively impacted by AI in education. One noteworthy feature of AI is its capacity to reduce administrative workloads, freeing up educators to concentrate more on the essential components of mentoring and instruction. For example, automated grading systems can effectively manage regular assessments, saving teachers a significant amount of time that would otherwise be required for manual grading. Because it saves time, teachers can devote more of their efforts to creating lesson plans that are interesting, giving each student personalized attention, and creating a lively, interactive learning environment.

Despite the implementation of the European Union (EU) AI Act, constituting the inaugural regulatory framework for artificial intelligence, the decision by Italy to restrict the application of generative AI, as exemplified by the case of OpenAI's ChatGPT (Yeralan and Lee, 2023), stands as a noteworthy and somewhat surprising development. This action encourages a closer examination of the complex issues and repercussions associated with the application of generative AI across various regions and reignites conversations about the ethical and legal implications of cutting-edge AI technologies.

Opportunities presented by generative AI to higher education

Incorporating AI into education enhances learning outcomes, increases access to educational resources, improves retention rates, lowers costs, and reduces the time required for course completion (Akinwalere and Ivanov, 2022; Bates *et al*., 2020). Overall, AI contributes to a more efficient, accessible, and effective teaching and learning environment. Several authors (Guerrero-Quiñonez *et al*., 2023; Nassuora, 2022; Venugopal and Mamatha, 2023) have discussed various opportunities that generative AI present to higher education to overcoming the challenges faced by higher education institutions. The main ones of these opportunities are listed below:

- ***Personalised learning experiences***: AI can be used to create personalized learning experiences for students that take into account their individual learning styles, preferences, and goals. For example, AI tools can be used to recommend learning materials and activities that are best suited to a student's individual needs.

- ***Intelligent tutoring systems***: AI can be used to create intelligent tutoring systems that provide students with personalized feedback and support. These systems can help students stay engaged and motivated by providing immediate feedback and guidance.
- ***Adaptive learning platforms***: AI can be used to create adaptive learning platforms that adjust to the needs and abilities of individual students. AI tools can help make global classrooms available to all, including those who speak different languages or who might have visual or hearing impairments (Akinwalere and Ivanov, 2022) these platforms can help students stay on track and succeed in their studies.
- ***Predictive analytics***: AI can be used to analyse student data and provide insights into student behaviour and performance. This information can be used to identify areas where students are struggling and to develop interventions to help them succeed.
- ***Chatbots and virtual assistants***: AI can be used to create chatbots and virtual assistants that can answer student questions and provide support. These tools can help students get the help they need quickly and easily.

Generative AI tools commonly used in higher education in South Africa

Generative AI tools used in higher education enhance teaching and learning by providing personalised, efficient, and adaptive support; fostering student engagement; and addressing diverse learning needs. These sections provides few examples of AI and generative AI tools commonly used in higher education in South Africa.

Advances in the development of generative AI tools in South Africa includes the development of a system at the University of Johannesburg that predicts transformer failures, enhancing efficiency and reducing electricity costs, alongside the creation of an AI system capable of restoring lost voices, as well as applications in diagnosing leukaemia and detecting epilepsy (Olaitan *et al*., 2021). Another example is the Python-Bot, also developed by the University of Johannesburg. Python-Bot belongs to the chatbot family of generative AI. Chatbots are conversational or interactive agents powered by various technologies, including artificial intelligence (AI) and natural language processing (NLP), enabling them to comprehend and react to user inputs conversationally and provide prompt responses (Okonkwo and Ade-Ibijola, 2021). The Python-Bot assists novice programmers to understand Python's basic syntactic structures and semantics. It starts an operation with the introduction of the Bot and the user. It describes itself and seeks the student's willingness to engage in conversation, as well as recognition of the student's name. The Bot receives the student's query, prepares it, and answers the student by sending him/her predefined textual insights into the subject." (Chinedu Eduvos *et al*., 2021).

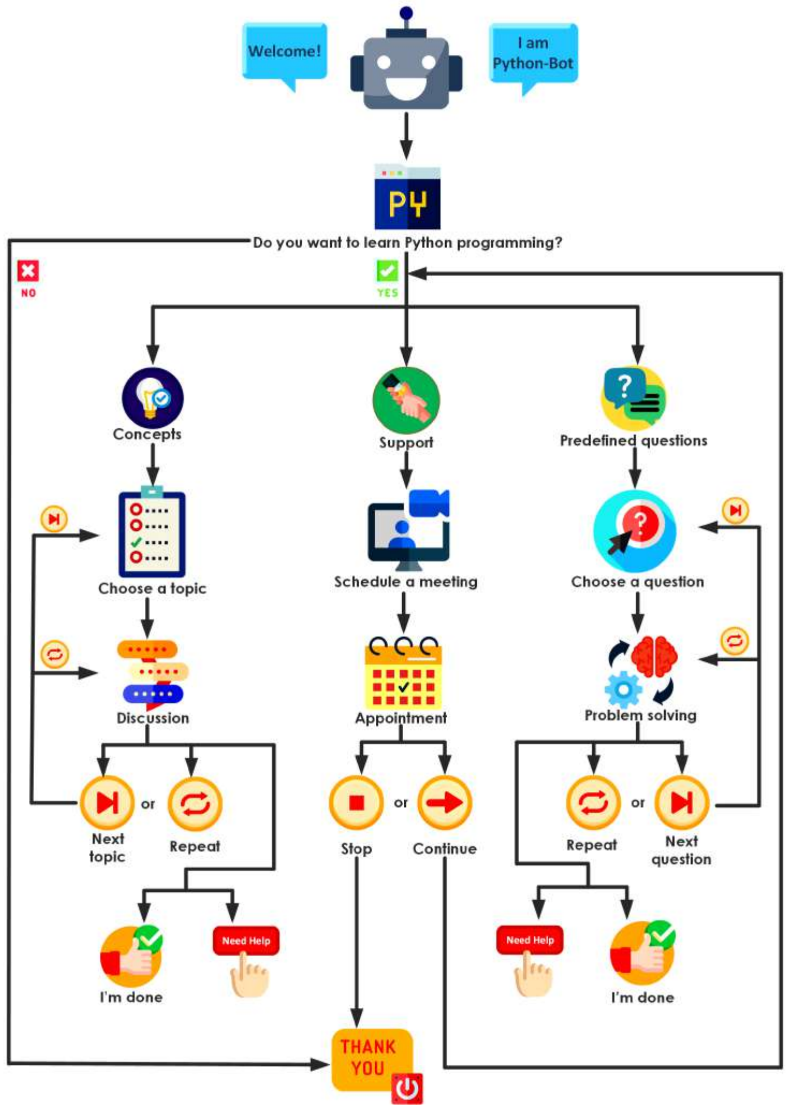

Figure 3: Python-Bot Flowchart (Okonkwo and Ade-Ibijola, 2021)

The basic flow of interactions with the tool are depicted on Figure 3 above and also the following steps (Chinedu Eduvos *et al.*, 2021):

i. Python-Bot starts an operation with the introduction of the Bot and the user.
ii. It describes itself and seeks the student's willingness to engage in conversation, as well as recognition of the student's name.
iii. The Bot receives the student's query, prepares it and answers the student by sending him/her predefined textual insights into the subject.
iv. Also, it provides some examples of the implementation of Python algorithms and outputs showing the willingness of the user to participate in a learning conversation and the ability of the Bot to provide the correct answers.

Another example is Jill Watson, an AI-powered chatbot developed by researchers at the Georgia Institute of Technology (Goel Ashok *et al.*, 2019). The conception of Jill Watson was prompted by Georgia Tech's introduction of a massive online course for a Master's degree in Computer Science, which attracted thousands of global learners leading to a surge in questions on discussion forums; which was a big challenge to the limited number of instructors (Wang *et al*., 2017).

Watson's introduction was a cutting-edge experiment that demonstrated the potential of artificial intelligence (AI) in education, specifically in terms of bolstering and improving the learning process (Goel Ashok *et al*, 2019). The interesting fact is that some students initially thought that they were chatting to a human teaching assistant (Goel Ashok *et al*., 2019). Figure 4 below presents a summary of the evolution of Jill Watson and some of its characteristics.

Figure 4: Jill Watson Snippets (Wang et al., 2017)

Enhancing teaching and learning in higher education using generative AI tools

Higher education institutions ought to make a concerted effort to establish relationships with other higher education institutions both locally and internationally in order to foster knowledge sharing and the formation of alliances in academic programs, research, and training that are pertinent to the modern technology landscape (Chan, 2023a). Collaborations the mutual benefit of academics and students, this initiative is essential.

Another aspect that has become necessary with the proliferation of generative AI tools for higher education is curriculum review. South African higher education institutions must critically assess their curricula to make sure that they provide students with the knowledge, skills, and attitudes necessary for navigating the complexities of the modern world if they want to be relevant, competitive, and responsive (Nqabeni, 2023). This calls for a sustained dedication to curriculum development that integrates cutting-edge ideas and interdisciplinary viewpoints to give students a broad range of skills for taking on today's opportunities and challenges. Furthermore, it is essential to place equal emphasis on creative methods of assessment in order to fairly represent the range of knowledge, abilities, and eliminate potential dishonest use of generative AI tools.

Making a reference to the existing tools introduced above, Table 1 below enlists how generative AI can be used to address the challenges facing higher education as discussed earlier in this paper.

Table 1: Contribution of Generative AI Tools towards Addressing the Higher Education Challenges

Teaching and learning challenges	**Contribution of Generative AI tools towards addressing the identified challenges**
High dropout rates	AI algorithms and predictive analytics can analyse student data to identify early signs of academic or personal issues that may lead to dropout, allowing for timely interventions. To improve prediction accuracy, they focused on non-cognitive variables such as time management and self-esteem, unlike many other studies (Nassuora, 2022).
Diverse learning Needs	Personalised Learning Platforms: AI-driven platforms can adapt to individual learning styles, pace, and preferences, providing customized learning experiences. Educational institutions using assistive technology find intelligent assistants to be very beneficial (Algabri Hayder Kareem *et al*., 2021). Adaptive Assessments: AI can create assessments that adjust difficulty based on a student's demonstrated proficiency, ensuring a more accurate measure of knowledge.
Language and social barriers	Language Support Tools: AI-powered language translation tools can assist students and faculty in overcoming language barriers, facilitating better communication.

Assessments	The main advantage of AI in the classroom is that it prevents teachers from making a calculation error, resulting in a more objective and accurate assessment (Algabri Hayder Kareem *et al*., 2021). There are proctoring tools that are employed to ensure integrity of assessments.

Conclusion

The adoption and use of generative AI in South African higher education signifies a transformative shift, promising improved teaching and learning experiences. The Fourth Industrial Revolution has brought technological advancements, and Generative AI stands out as a disruptive force with the potential to address challenges faced by institutions. While the benefits include enhanced efficiency, expanded access, and personalised learning, caution is warranted, particularly with tools like ChatGPT, which have sparked concerns about academic integrity.

The paper acknowledges the existing challenges in South African higher education, including poor preparation, large classes, high dropout rates, diverse learning needs, language barriers, and societal challenges. The emergence of generative AI tools like Python-Bot and Jill Watson offers promising solutions, demonstrating the technology's potential to overcome educational obstacles.

It is evident from this paper that generative AI tools will be of much value in improving teaching and learning in South African higher education institutions. According to recent studies there is positive reception of these technologies from both students and academics, although academics have reservations about potential threat to integrity of teaching and learning. Further research is needed to establish and make recommendations on policy directions on use of AI within the South African higher education sector.

References

Abgaryan, H., Asatryan, S. and Matevosyan, A. 2023. Revolutionary changes in higher education with artificial intelligence. *Main Issues of Pedagogy and Psychology 10*(1), 76–86. https://doi.org/10.24234/miopap.v10i1.454.

Akinwalere, S. N. and Ivanov, V. 2022. Artificial intelligence in higher education: challenges and opportunities. *Border Crossing 12*(1), 1–15. https://doi.org/10.33182/bc.v12i1.2015.

Algabri Hayder Kareem, Kharade K. G. and Kamat K. 2021. *Promise, Threats, And Personalization In Higher Education With Artificial Intelligence*. https://www.researchgate.net/publication/359894024.

Al-Maskari, A., Al Riyami, T. and Ghnimi, S. 2022. Factors affecting students' preparedness for the fourth industrial revolution in higher education institutions. *Journal of Applied Research in Higher Education*. https://doi.org/10.1108/JARHE-05-2022-0169.

Bates, T., Cobo, C., Mariño, O. and Wheeler, S. 2020. Can artificial intelligence transform higher education? *International Journal of Educational Technology in Higher Education* 17(1). https://doi.org/10.1186/s41239-020-00218-x.

Bond, M. and De Laat, M. (n.d.). *Artificial Intelligence in Higher Education Insight from a Systematic Review of Reviews*. https://doi.org/10.13140/RG.2.2.24571.69921.

Bosch, T., Jordaan, M., Mwaura, J., Nkoala, S., Schoon, A., Smit, A., Uzuegbunam, C. E. and Mare, A. 2023. South African university students' use of AI-powered tools for engaged learning. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4595655.

Chan, C. K. Y. 2023a. A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education 20*(1). https://doi.org/10.1186/s41239-023-00408-3

Chan, C. K. Y. 2023b. A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education 20*(1). https://doi.org/10.1186/s41239-023-00408-3.

Chinedu Eduvos, O., Ade-Ibijola, A. and Okonkwo, C. W. 2021. *Python-Bot: A Chatbot for Teaching Python Programming*. https://www.abejide.org.

DHET. 2023. *Statistics on Post-School Education and Training In South Africa: 2021* Department of Higher Education and Training. www.dhet.gov.za.

Gimpel, H., Hall, K., Decker, S., Lämmermann, L., Mädche, A., Röglinger, M., Ruiner, C., Schoch, M., Schoop, M., Urbach, N. and Vandirk, S. 2023. *Unlocking the Power of Generative AI Models and Systems such as GPT-4 and ChatGPT for Higher Education A Guide for Students and Lecturers Torsten Eymann 2, 6 Executive Summary*.

Goel Ashok, K., Lalith, P., Dede, C., Richards, D. and Saxberg, B. 2019. *Jill Watson: A Virtual Teaching Assistant for Online Education*. Georgia Institute of Technology.

Guerrero-Quiñonez, A. J., Bedoya-Flores, M. C., Mosquera-Quiñonez, E. F., Mesías-Simisterra, Á. E. and Bautista-Sánchez, J. V. 2023. Artificial Intelligence and its scope in Latin American higher education. *Ibero-American Journal of Education and Society Research 3*(1), 264–271. https://doi.org/10.56183/iberoeds.v3i1.627.

Gulya Jason J. 2023. The Humanities in the age of AI: A Professor’s reflection. *Higher Education Digest.*

Ilieva, G., Yankova, T., Klisarova-Belcheva, S., Dimitrov, A., Bratkov, M. and Angelov, D. 2023. Effects of Generative Chatbots in Higher Education. *Information 14*(9), 492. https://doi.org/10.3390/info14090492.

Krish, C., Nothando, N. and Sharlene, S. 2023. *Navigating the impacts of generative AI in South Africa: challenges, opportunities and ethics*. Human Sciences Research Council.

Kumar, R., Sehgal, K. and Meena, A. L. 2021. Artificial intelligence: Encouraging students in higher education to seize its potential. In *Artificial Intelligence, Machine Learning, and Data Science Technologies: Future Impact and Well-Being for Society 5.0* (pp. 155–171). CRC Press. https://doi.org/10.1201/9781003153405-8.

Landa, N., Zhou, S. and Marongwe, N. 2021. Education in emergencies: Lessons from COVID-19 in South Africa. *International Review of Education 67*(1–2), 167–183. https://doi.org/10.1007/s11159-021-09903-z

Lubinga, S., Maramura, T. C. and Masiya, T. 2023. The Fourth Industrial Revolution Adoption: Challenges in South African Higher Education Institutions. *Journal of Culture and Values in Education*, *6*(2), 1–17. https://doi.org/10.46303/jcve.2023.5.

Luebering J. 2023. *Higher Education. Britannica Encyclopaedia*. Https://Www.Britannica.Com/Topic/Higher-Education.

Mannuru, N. R., Shahriar, S., Teel, Z. A., Wang, T., Lund, B. D., Tijani, S., Pohboon, C. O., Agbaji, D., Alhassan, J., Galley, J., Kousari, R., Ogbadu-Oladapo, L., Saurav, S. K., Srivastava, A., Tummuru, S. P., Uppala, S. and Vaidya, P. 2023. Artificial intelligence in developing countries: The impact of generative artificial intelligence (AI) technologies for development. *Information Development*. https://doi.org/10.1177/02666669231200628.

Nassuora, A. B. 2022. *Applied Artificial Intelligence Applications In Higher Education Institutions: A Systematic Review*. http://www.webology.org.

Nqabeni, N. 2023. Rethinking South African Higher Education Institutions Role Towards Curriculum Restructuring. *E-Journal of Humanities, Arts and Social Sciences*, 1253–1263. https://doi.org/10.38159/ehass.20234108

Nwosu, L. I., Bereng, M. C., Segotso, T. and Enebe, N. B. 2023. Fourth Industrial Revolution tools to enhance the growth and development of teaching and learning in higher education institutions: a systematic literature review in South Africa. *Research in Social Sciences and Technology*, *8*(1), 51–62. https://doi.org/10.46303/ressat.2023.4

Okoli, C. 2015. A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*. http://aisel.aisnet.org/cais

Okoli, C. and Schabram, K. 2012. A guide to conducting a systematic literature review of information systems research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1954824

Okonkwo, C. W. and Ade-Ibijola, A. 2021. Chatbots applications in education: A systematic review. In *Computers and Education: Artificial Intelligence* (Vol. 2). Elsevier B.V. https://doi.org/10.1016/j.caeai.2021.100033.

Olaitan, O. O., Issah, M., Wayi, N. and Olaitan, O. 2021. A framework to test South Africa's readiness for the fourth industrial revolution. *South African Journal of Information Management*. https://doi.org/10.4102/sajim.

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., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1). https://doi.org/10.1186/s13643-021-01626-4.

Reddy, M. M. 2017. Teaching to enhance quality learning in large classes within the South African higher education context. *PONTE International Scientific Research Journal 73*(7). https://doi.org/10.21506/j.ponte.2017.7.22.

Sibanda M and Fon F. 2022. Prospects and challenges of integrating technology into pedagogy in South African higher education institutions in the context of the Fourth Industrial Revolution. *Kagisano 13*, *13*.

Singh, R. J. 2015. Current trends and challenges in South African higher education. *South African Journal of Higher Education 29*, 1–7.

Snyder, H. 2019. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039.

Sullivan, M., Kelly, A. and McLaughlan, P. 2023. ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching 6*(1), 31–40. https://doi.org/10.37074/jalt.2023.6.1.17.

Tarisayi, K. S. 2024. ChatGPT use in universities in South Africa through a socio-technical lens. *Cogent Education 11*(1). https://doi.org/10.1080/2331186x.2023.2295654.

Venugopal, R. and Mamatha, V. 2023. impact of artificial intelligence (AI) on teaching and learning in india's higher education sector. *Issue 5 Ser. II*, *13*. https://doi.org/10.9790/7388-1305020106.

Wang, I., Mujumdar, Q., Goel, R., Gregori, A. and Jill Watson, V. 2017. *Jill Watson: A Suite of Online Learning Tools*. https://dilab.gatech.edu/a-suite-of-

Yeralan, S. and Lee, L. A. 2023. Generative AI: Challenges to higher education. *Sustainable Engineering and Innovation 5*(2), 107–116. https://doi.org/10.37868/sei.v5i2.id196.

Developing a Framework for Enhancing Critical Thinking in Student Assessment in the Era of Artificial Intelligence

A. Makina
University of South Africa
makina@unisa.ac.za

Abstract
Artificial intelligence has far-reaching implications on student assessment in higher education. Students are able to use material generated by AI tools to complete their assessments. They are therefore able to do well in assessments without having applied their minds critically to the issues required by the assessments. This compromises the attainment of learning outcomes such as critical thinking and problem solving, which is a serious concern. This paper presents a framework that educators can use to enhance critical thinking among students while allowing them to use artificial intelligence in completing their assessments. A developmental research design focusing on the creation of the initial prototype of a framework to enhance the quality of assessment with AI to develop critical thinking skills was used. The Bloom's Taxonomy and the levels of critical thinking provided the theoretical framework that guided the process of designing and developing a framework for enhancing critical thinking among students when they are allowed to use artificial intelligence in assessments. Educators can improve on the framework depending on the specific contexts that they apply it to.
Key words: Artificial intelligence, critical thinking, higher education, educators, assessment

Introduction

In the digital era, artificial intelligence (AI) has emerged as a powerful tool, revolutionising diverse sectors, including education. The integration of AI into education and some industries testifies to its potential. Maintaining a healthy scepticism about AI while recognising its dependence on human intelligence is crucial (Stanford University, 2023). Students are improving at finding information online rather than memorising it, and human critical thinking has become vital to the evolving AI. With the growing sophistication of AI and the ease with which it can be used by students across all disciplines and subjects, questions are asked about its impact on human critical thinking (Fang, Su and Xiao. 2018).

AI is transforming education, but as with every new technology its benefits must be scrutinised. Whilst useful, chatbots like ChatGPT are only as good as the datasets of text and statistical patterns that they are trained on (Spector and Ma, 2019). These chatbots models may be incomplete, may misinterpret meanings, or only reflect certain cultural viewpoints (Spector and Ma, 2019). Therefore, relying on answers generated by chatbots can be dangerous. For example, students can save time using chatbots like ChatGPT to find information for an essay, but if a chatbot will do the whole thing for them, they may think that there is no need to put any effort in the work at all. However, the educator will need ways to scrutinise the chatbot's answers by finding alternative ways of preparing teaching and assessment so as to improve the students' critical thinking skills. Educators need to be aware of the dangers of only using AI content. This will enable them to gain confidence in navigating AI critically to make the best use of it in their learning and beyond. Understanding how AI shapes the future of critical thinking provides valuable insights into how educators can leverage AI to enhance decision-making processes and, problem-solving abilities, and overall human intelligence (Stanford university, 2023).

Critical thinking is the capacity to question, evaluate, and reflect on previously held beliefs (TEAM International, 2022). It is the ability to recognise ambiguity and make informed judgements and decisions (TEAM International, 2022). It involves examining, interpreting, and reasoning to clarify and justify positions (University of Louisville (2023). Critical thinking involves scepticism, reasoning, logic, correlation, and causality to separate what's necessary from what's irrelevant in a discussion like what you must do when chatting with Generative AI. Because, critical thinking enables people to solve problems, make informed decisions, enhance communication skills, form arguments, or reach conclusions its benefits include better communication, increased creative thinking, enhanced problem-solving and improved decision making. On the other hand, critical thinking refers to higher levels of thinking that students need to enable them to think effectively and rationally (Aithal and Silver, 2023). Critical thinking helps individuals see beyond their point of view, to manage a problem and make the best decision possible pragmatically (TEAM International. (2022). The problem is that this ability cannot be digitised or only learnt in a class. It requires practice and experiences where one is compelled to debate and confront diverse opinions (Aithal and Silver, 2023). Unlike other skills, critical thinking is learnt primarily from failure, especially from solving complicated problems, which teaches one to experiment with different solutions (TEAM International, 2022). Therefore, only educators willing to give their students the freedom to learn from their mistakes will gain from this experience.

Since educators still need to guide learners in the solution of complex problems, interpret data, and address ethical concerns, managing AI content become key. To make the most of AI in the classrooms, educators must change the way they teach and especially the way they deal with assessment to sift through AI-generated content. Generative AI was designed to understand and generate human-like text and is trained by feeding it examples and prompts (Cambridge, 2023). It is important to use the information it provides only as a starting point to refine the writing to fit your situation.

Understanding critical thinking

Though there several definitions of critical thinking, a concise definition of critical thinking remains elusive. Critical thinking, broadly speaking, is a multi-dimensional and multifaceted human capability that has been interpreted from three perspectives, namely education, psychology, and epistemology (Ennis, 2018). In a developmental approach to critical thinking, Spector (2019) argues that it involves a series of cumulative and related abilities, dispositions and other variables (for example, motivation, criteria, context, knowledge). This approach proceeds from experience (for example, observing something unusual) and then to various forms of inquiry, investigation, examination of evidence, exploration of alternatives, argumentation, testing conclusions, rethinking assumptions, and reflecting on the entire process. Experience and engagement are ongoing throughout the process, which proceeds from relatively simple experiences (for example, direct and immediate observation) to more complex interactions (for example, manipulation of an actual or virtual artifact and observing effects (Spector, 2019).

On the other hand, the developmental approach to critical thinking involves a diversity of mental processes and non-cognitive states, which help a person's decision making to become purposeful and goal directed (TEAM International, 2022). The associated critical thinking skills enable individuals to achieve a desired outcome in a challenging situation. In the process of critical thinking, there are two additional cognitive capabilities essential to critical thinking namely, metacognition and self-regulation (Schraw, Crippen and Hartley, 2006). Many researchers believe that metacognition has two components. The first one is awareness and understanding of one's own thoughts, and the second one is the ability to regulate one's own cognitive processes (Schraw *et al*., 2006). For example, Davies (2015) described metacognition as the capacity to monitor the quality of one's thinking process, and then to make appropriate changes. There are also some other abilities such as communication, collaboration and creativity, which are now essential in current society (Schraw *et al*., 2006). Those abilities along with critical thinking are called the 4Cs. They are individually monitored and regulated through metacognitive and self-regulation processes. The abilities involved in critical thinking are categorised in Bloom's taxonomy into higher order skills (for example, analysing and synthesising) and lower level skills (for example, remembering and applying) (Anderson and Krathwohl, 2001). The developmental stages of critical thinking are presented in Figure 1. Critical thinking is broken down into three areas, namely 'understanding and analysing ideas and arguments'; 'evaluating ideas and arguments'; and 'solving problems and making decisions' (Table 1). It should be noted that the most in demand skills in the Fourth Industrial Revolution for 2030 are problem-solving, self-management, working with people (TEAM International, 2022). These rightly align with principles that drive critical thinking.

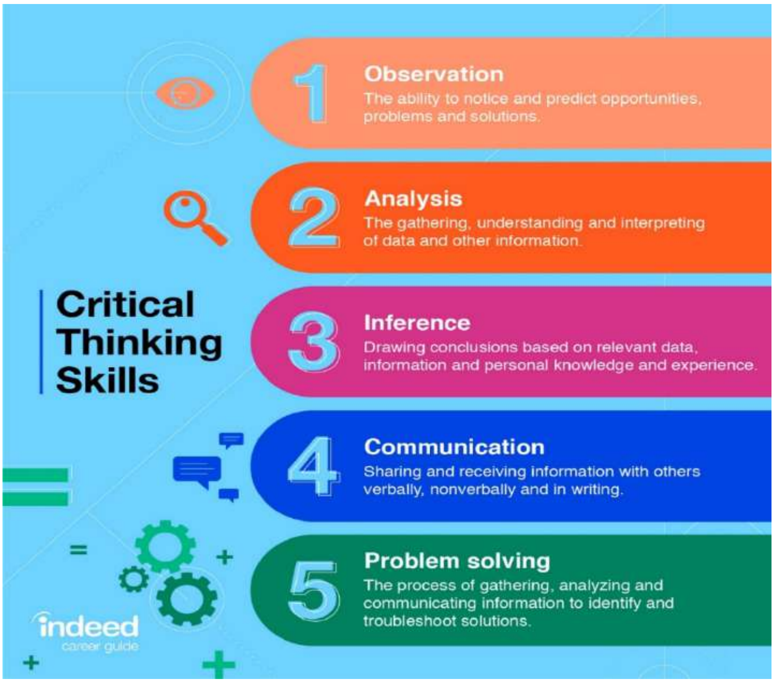

Figure 1: Developmental Stages of Critical Thinking

Critical thinking enables educators and students to solve problems, make informed decisions, enhance communication skills, form arguments, or reach conclusions. Critical thinking is the base for better communication, increased creative thinking, enhanced problem-solving and improved decision making. A critical thinker embraces challenges and problems as they emerge because they can make quick and assertive decisions. Full of resources, they consider different opinions, and offer a list of solutions. Becoming an influential critical thinker does not happen overnight. Therefore, educators need to understand the stages of critical thinking, to make it easier for them assess learners in a continual practice and eventually make it part of their mental process. So, the essential steps of the critical thinking are outlined in Table 1.

Table 1: The three main areas of critical thinking skills

No	**Main levels of critical thinking skills**	**Detailed explanation**
1.	**Understanding and analysing ideas and arguments**	Refers to a learner's ability to identify and analyse information in order to recognize patterns and relationships. This helps learners to gain a deeper understanding of ideas and arguments, as well as to interpret and draw inferences about the information they are presented with.

2.	**Evaluating ideas and arguments**	Is related to a learner's ability to judge which arguments or ideas they can rely on and which they should be sceptical about. This includes evaluating evidence presented in an argument, as well as the argument's overall logic. Mastering this competency helps learners draw appropriate conclusions and construct strong arguments themselves.
3.	**Solving problems and making decisions**	Involves many skills such as identifying and analysing problems, gathering appropriate information, evaluating a range of options, making decisions about which options to implement and finally, evaluating those decisions to further refine solutions.

Source: (Aithal and Silver, 2023)

The problem statement

The central problem is a conceptual one, concerning the link between the use of AI in assessment and the development of critical thinking in assessment. Building on contemporary understandings of assessment within courses, it is argued that a paradigm shift is needed from viewing poorly set assessment to, highly adaptable, unbounded, assessment with AI that enables the development of critical thinking. It is further argued that the issues are less to do with the type of assessment choices and more to do with the adoption of a broader conceptualization of quality assessment that develops critical thing within the use of AI. A framework is developed to guide universities and educators on how to develop critical thinking in students while using Artificial intelligent content.

The paper seeks to provide answers to two fundamental questions below.

- How can we align AI content with assessment for the development of critical thinking in the students?
- Would it be possible to develop a model that can guide educators to be conscious of developing assessment for critical thinking whilst they are using AI on different aspects of their teaching?

The paper seeks to motivate for the development of a framework for enhancing critical thinking in assessment with AI. The focus is to determine the development of critical thinking skills by transforming the assessment strategies to suit the use of artificial intelligence.

Methodology

The material reported in this paper has been generated following a developmental research design, focusing on the creation of the initial prototype of a framework to enhance the quality of assessment with AI to develop critical thinking skills for teaching and learning. This process draws insights from extensive literature reviews, reflective practice literature, and the researcher's own experiences.

Development research, in this context, is viewed as a pragmatic approach where the researcher synthesizes knowledge from literature and personal expertise to design, and develop a prototype (van den Akker, 1999:5). This approach views development research as an iterative process, guided by formative evaluations during the prototype's development, trial testing, and implementation. The framework presented in this paper serves as an educational tool, empowering universities to effectively utilise the strategies of assessment with AI to enhance students' critical thinking. A well-informed product of this nature can transform and guide lecturers' practices, thereby benefiting the teaching and learning process.

The theoretical framework

Bloom's Taxonomy and the stages of critical thinking processes served as the theoretical framework guiding what is presented in the paper. Educators must be made aware of the stages of critical thinking before they can start to assess students in an environment that they are using AI. It is also important to understand the specific thinking skills that are valuable in the context of AI and bind them around Bloom's taxonomy before they can carry out valuable and quality assessment of students in the classroom. A structured assessment with AI framework can help educators to guide students into analysing AI-related issues, making informed decisions, and addressing complex problems.

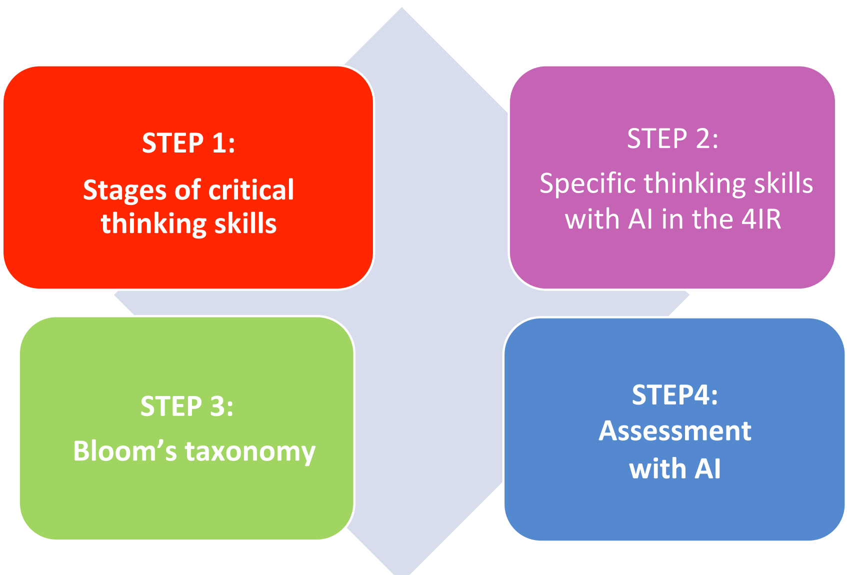

Figure 2: The Theoretical Framework

Applying Bloom's Taxonomy in assessment with AI

The purpose of education is to change the thoughts, feelings and actions of students (Bloom 1952). This quote encapsulates the essence of Bloom's Taxonomy, which aims to promote higher levels of thinking and learning that are important in the development of critical thinking skills. Bloom's Taxonomy divides the cognitive domain into six distinct categories, each representing a different level of understanding in the process of learning (Table 1). These categories serve as the building blocks for crafting effective learning objectives and designing appropriate assessment activities. It is a hierarchical model designed to classify educational learning objectives into levels of complexity and specificity. It was developed by educational psychologist Benjamin Bloom and his colleagues in the 1950s. The original taxonomy consists of three domains: cognitive, affective, and psychomotor. The cognitive domain, which is most commonly used, includes six levels: remembering, understanding, applying, analysing, evaluating, and creating. These levels are often depicted as a pyramid, with the lower-level skills forming the base and the higher-level skills at the top.

The taxonomy is a valuable tool for assessment with AI as it provides a framework to structure learning objectives against assessments that develops critical thinking. For example, in a history lesson, a teacher might start by asking students to remember key dates (remembering), then move on to understanding the significance of these dates, applying this knowledge to different historical contexts, analysing different sources of information, evaluating different interpretations, and finally, creating their own interpretation of events. The bloom's taxonomy objectives are further divided into four knowledge dimension represents factual, conceptual, procedural and metacognitive (Table 1).

Assessment systems can play a significant role in improving critical thinking through Bloom's Taxonomy incorporated, profoundly to improve critical thinking. This paper discourages lower domains which promote cramming and memorization while promoting the higher domains of learning comprising analysis, evaluation and creativity which focus on critical thinking skills. Therefore, by transcending assessment pattern from the lower level domains of remembering, understanding and applying to the higher domains of analysing, evaluating and creating, the teaching-learning process in higher education can be improved to profound degree and level. It is therefore, important to use assessment as an effective means to improve critical thinking and discourage practices that tend to lower critical thinking behaviours. Aligning assessment with the Bloom's taxonomy ultimately inculcate critically analytical and problem-solving approaches among students and teachers, and thus moulding the teaching pedagogy. The six cognitive levels of Bloom's Taxonomy serve as a roadmap for educators to design learning experiences that target a range of intellectual skills and promote critical thinking (Adams, 2015). To ensure a comprehensive learning experience, educators should focus on incorporating a variety of learning objectives that address higher-order thinking and mental processes. By doing so, they can create engaging learning environments that foster the development of critical skills through conceptual knowledge and metacognitive strategies.

It follows from the foregoing that it is critical to incorporate Bloom's Taxonomy into assessment practices; that assessment approaches should include both lower and higher domains of learning; and critical, analytical and problem-solving approaches should be added to assessment system to positively improve and uplift critical thinking.

Table 2: Describing Bloom's Revised levels

Cognitive categories	**Expected knowledge**	**Knowledge dimensions**	**Expected knowledge**
Remember	• Recognizing • Recalling	• Factual Knowledge	• Knowledge of terminology • Knowledge of specific details and elements
Understand	• Interpreting • Exemplifying • Classifying • Summarising • Inferring	• Conceptual Knowledge	• Knowledge of classifications and categories • Knowledge of principles and generalizations • Knowledge of theories, models, and structures
Apply	• Executing • Implementing	• Procedural Knowledge	• Knowledge of subject-specific skills and algorithms • Knowledge of subject-specific techniques and methods • Knowledge of criteria for determining when to use appropriate procedures
Analyse	• Differentiating • Organising • Attributing	• Meta-cognitive Knowledge	• Strategic Knowledge • Knowledge about cognitive tasks, including appropriate contextual and conditional knowledge • Self-knowledge
Evaluate	• Checking • Critiquing	• Judgments in terms of internal evidence • Judgments in terms of external criteria	• Make judgments based on criteria and standards (for example, detect inconsistencies or fallacies within a process or product,
Create	• Generating • Planning • Producing	• Production of a unique communication • Production of a plan, or proposed set of operations • Derivation of a set of abstract relations	• Put elements together to form a new coherent or functional whole; reorganise elements into a new pattern or structure

Educators must be made aware of both Bloom's Taxonomy and stages of critical thinking before they can start to prepare assessment for students in an environment that they are using AI. It is also important to understand the specific thinking skills that are valuable in the context of AI and bind them around Bloom's taxonomy before they can carry out quality assessment for students in the classroom. Structured assessments with AI can help educators to guide students into analysing AI-related content, making informed decisions, and addressing complex problems.

Understanding artificial intelligence and its role in education and assessment

The advancement of technology including AI is revolutionising assessment in education. Students' learning environments are shifting from technology-enhanced learning environments to smart learning environments that rely more on the support of AI technology such as neural networks, learning analytics and natural language processing (Spector, 2016). The smart learning or personalised learning is better supported and realised in a smart learning environment. In short, in the current era, personalised learning is essentially about using AI (Spector, 2016).

AI presents potential competitive advantages and innovative opportunities in the context of critical thinking, which are crucial for understanding ahead in today's rapidly evolving education landscape (Spector and Ma, 2019). One way of assessing students with AI is to bring more interactivity into the classroom. It could be bringing teaching methods that get students to be creative, to role-play, or to think critically leading to a deeper kind of learning than rote memorisations (Cambridge, 2023; Spector, 2019). AI like ChatGPT can play the role of a debate opponent and generate counterarguments to students' positions. For example, by exposing students to an endless supply of opposing viewpoints, chatbots could help them to look for weak points in their own thinking. In addition, if English is not a student's first language, chatbots can be a big help in drafting text or paraphrasing existing documents, doing a lot to level the playing field (Cambridge, 2023). Chatbots also serve students who have specific learning needs, to aiming for bringing the benefits of AI methods into life in a controlled, reliable and stable environment.

Maintaining a healthy scepticism about artificial intelligence while recognising its dependence on human intelligence is crucial. Despite its capabilities, AI is not infallible and may only sometimes provide accurate information, but it must be noted that AI cannot replace human critical thinking. The fear of artificial intelligence threatening critical thinking is not new. As AI continues to evolve, developing and nurturing critical thinking skills will be essential to shaping a future where AI benefits the students at large. Understanding how Artificial Intelligence shapes the future of critical thinking provides valuable insights into how education can leverage AI to enhance decision-making processes and problem-solving abilities,

The effect of AI on critical thinking

Technology has impacted cognitive development. An example of this is the 'Google effect' (Sparrow, Betsy *et al*., 2023) which refers to people tending to forget things they know they can easily find on the internet but remember stuff they know they cannot find online. This is not necessarily a negative effect but a change in how memory operates. Instead of remembering the actual information, people remember where they can find it (Stanford University, 2023). It could also improve learning and assessment quality by facilitating personalised conversations between educators and students, rapidly assessing students' skills, and matching them with suitable roles in the workforce (Cambridge, 2023). It allows individuals to approach AI technologies with careful consideration and a deep understanding, which can help promote AI's positive impact on society while minimizing potential risks. As AI continues to evolve, developing and nurturing critical thinking skills will be essential to shaping a future where AI benefits all.

AI and the educator

It is not necessary for educators to be experts in AI. However, it is critical for them to have basic understanding of the AI tools. The question is how can the educators be equipped with the right skills so that they are relevant to issues of today. Policies that relate to teacher development remain the same that have been in existence for decades as they have not been adapted notwithstanding the recent technological changes that affect teaching and assessment (Seufert, Guggemos and Sailer, 2020). Even those that develop teachers are not well acquainted with the new developments (Chai, Jong and Yan, 2020). Often the focus of interventions is on students and not on the educators. The influence of AI in education demands that this state of affairs should change and the change is urgent, especially in the context of a generative tool like AI that can write things from scratch and improve what is already improved. By integrating the stages of critical thinking and how people think in assessment with AI, educators can help students approach formative or summative assessment with a structured and analytical mindset. This approach can improve their comprehension of subject areas and equip them to make informed decisions and participate in meaningful and ethical discussions within AI technologies use.

The role AI in assessment

New demands on higher education have had a disruptive effect on the trusted conventional assessment strategies (Bloxham, 2013). Assessment is a critical component of higher education, as it makes students’ learning visible and evaluates their progress towards specific learning outcomes (Brown, 2013; Brown, 2019; Brown, 2020). AI enhanced assessments can provide new opportunities for assessing student learning. However, it is important to ensure that educators use these AI content in a way that enhances critical thinking. Educational assessments must align closely with curricular goals and values, while also being applicable within classroom settings (Bennett, 2018). AI should complement human involvement in education, rather than replace teachers with automated teaching or testing machines. Though AI uses detailed approaches in utilising organised learning models and dynamic learning maps, combining cognitive and curriculum insights with psychometric tools to measure students' status and progress, they should be closely connected to the curriculum and teaching objectives (Brown and Hattie, 2012; Brown *et al*., 2018). Kingston *et al*., 2022). Shin, Guo and Gierl (2021). Assessments must delve deeper into the insights provided by educational psychology, considering aspects such as human behaviour, attitudes, interpersonal relationships, and emotions (Shin *et al*., 2021). AI should work collaboratively with teachers in the classroom, offering valuable insights. The focus should be on assessments that support teachers, not replace them, offering a promising prospect for the future of education (Bennett, 2018).

With the advancement of AI technology, like ChatGPT, assessments should focus on higher-order thinking skills such as applications, analyses, and problem-solving. Lecturers should design assignments that pose issues with no straightforward correct answers to encourage students to think critically and creatively. Incorporating authentic assessments as part of a module's assessment strategy can provide students with the opportunity to develop critical thinking through real-life problems through the execution of tasks. By adopting a more holistic approach to assessments and encouraging deeper learning, lecturers can create an environment that supports students' academic success and prepares them for future challenges (Cambridge, 2023). This paper delves on how AI-driven insights can be used to monitor and improve classroom assessment dynamics to improve critical thinking skills.

The effectiveness of assessment to build critical thinking relies significantly on the delicate balance between assessment strategies and productive learning (Boud, 2018). Productive learning aims to empower students to apply knowledge and skills in new contexts, forge connections within a learning community and its resources, engage in meaningful interactions, and grasp the intricacies of knowledge construction which are the required aspects in the development of critical thinking. Assessing for productive learning is intricate and demands meticulous planning. Productive learning in the development of critical thinking hinges on the concept of constructive alignment, where desired learning outcomes, actual achievements, and assessment strategies harmonise. This alignment encompasses aspects like goals, learner outcomes, interactions, context, feedback processes, and communication. However, the way assessment tasks articulate productive learning expectations shapes students' critical thinking. Assessment tasks play a pivotal role in motivating meaningful learning for critical thinking. But unless meticulously planned, assessments can inadvertently encourage lower-level cognitive skills (James, McInnis and Devlin, 2002). Framing assessment tasks is insufficient and therefore tasks must be meticulously planned and designed to foster critical thinking.

A framework for enhancing critical thinking in student assessment in the era of AI

This framework advocates for an inclusive approach that integrates perspectives from the traditional way of assessment by educators, to a conscious way that advocates critical thinking. This is intended to equip future generations with the analytical ability necessary to navigate the evolving landscape of AI-generated content effectively. This framework underscores the importance of cultivating a diverse skill set among students, encompassing not only technical proficiencies but also ethical reasoning, creativity, and adaptability. Ethical considerations regarding AI-generated content, including bias and transparency, are also addressed within the context of the assessment practices described in the framework.

Table 3: A Framework for Assessment with AI to Develop Critical Thinking

	Category (Revised Bloom's Taxonomy)	Assessment knowledge expected	Appropriate assessment events
1.	**Remembering**:	Use AI to get information or solve problems with or without learners. Find answers/feedback from AI (for example using ChatGTP). Feed data/information into Artificial intelligence (GIGA). Ask questions that can directly use AI. Retrieved information is sifted for its correctness and credibility.	**AI prompts** (**AI Technical solutions) (GIGA)**
2.	**Understanding**:	Provide assessment that describe and demonstrate steps, practice performing the steps (for example, getting aware of formulas and procedures) Supportive knowledge' components (concepts, facts, processes, and principles that must be known or understood to perform tasks). *Verbs that can be used are Classify, identify and distinguish*	**Knowledge identification and retrieval (Ethical and credibility issues)**
3.	**Applying**:	Relating AI to real classroom and real-life activities (for example case studies) and relating it from real life. Finding ways to generally solve problems from both (for example, from first principles or from formulars).	**Integrate declarative and procedural knowledge (for example Making decisions and predict about events or outcomes)**
4.	**Analysing**:	Distinguishes between facts and inferences. Understanding processes and explaining why. For example assessing material or concepts through component parts so that the whole concept/structure can be understood. Assessing for part processes of solutions and not the final solutions. Providing opportunities for learners to explain when, how, and why. Explain distinctions among concepts. *Supporting assessment with group discussion or interaction.*	**Provide reflection. Declarative knowledge**

5.	**Evaluating**:	Assessing judgments about the value of ideas or materials. Comparing ideas/concepts/processes/solutions. Prompting learners to reflect, self-explain, and discuss what they are learning. For example, using YouTube/videos to compare the understanding Explain actions and decisions or events in terms of principles. Let learners create videos /flowcharts /mind maps etc. *Supporting assessment with group discussion or interaction*	**Using the learned knowledge Procedural knowledge (Distinguishing among) (Troubleshoot a process/concept/**idea)
6.	**Creating**:	Improve a concept or a process. Build a structure or pattern from diverse elements. Put parts together to form a whole, with emphasis on creating a new meaning or structure. Use AI to explain the solution of a problem (Using AI as your co-teacher) (for example, maximising profits with linear programming/Let students prepare assessment of part/whole concept/s for other students with AI use in mind.	**Knowledge generated by learners (for example, using AI mega-prompts)**

Table 4: Using the Framework to Check the Quality of Assessment with AI

5%	**10%**	**15%**	**20%**	**25%**	**30%**
(1) AI prompts (AI Technical solutions) (GIGA)	(2) Particular Knowledge identification and retrieval (Ethical and credibility issues)	(3) Integrate declarative and procedural knowledge (for example Making decisions and predict about events or outcomes)	(4) Provide reflection. Declarative knowledge	(5) Using the learned knowledge Procedural knowledge (Distinguishing among) (Troubleshoot a process/concept/idea)	(6) Knowledge generated by learners (for example using AI mega-prompts)

The purpose that the framework is intended to serve is providing lecturers with the know-how to directly enhance desired learning outcomes in teaching and learning environments. Developed based on the theoretical framework of Bloom's Taxonomy and the conceptual framework of the stages of critical thinking, it evaluates assessment with AI against defined criteria.

Educators can nurture students' intellectual skills by incorporating higher order thinking skills into assessment with AI. This approach nurtures cognitive development and encourages students to engage in meaningful conversations about the subject matter(Brame, 2020). Educators should be self-aware and work on specific areas to help learners develop critical thinking skills faster. It makes a great difference when these practices of developing critical skills exist in most if not all assessment with AI in the modules. One way to apply Bloom's Taxonomy in assessment with AI to improve critical thinking is through questioning techniques that encourage students to think at different levels of complexity (Anderson, 2014). In addition, group discussions, visualising learning through concept maps, graphic organisers, and other visual aids can help students grasp complex ideas more effectively. In summary, applying Bloom's Taxonomy in assessment with AI through various questioning, engaging conversations, and visualising techniques helps educators create a fertile ground for nurturing learners' critical thinking and fostering the development of higher-order thinking skills (García, Pacheco and Aguilar, 2018).

Some of the things that the educator can pursue to inculcate higher-order critical thinking skills, among students are as follows:

- Incorporate problem-solving activities by designing learning experiences that require students to apply their knowledge and skills to solve real-world problems.
- Promote critical thinking through debate, discussion and analysis of various perspectives.
- Make use of open-ended questions in assessment to encourage students to think deeply, synthesise information, and generate original responses.
- Implement project-based learning assessment that require learners to analyse, evaluate, and create solutions, fostering higher-order thinking skills.
- Assist students to develop an awareness of their thought processes and teach them strategies to regulate and monitor their own learning.
- Use AI to provide opportunities for students to collaborate, create, and explore concepts in new and innovative ways.

Conclusion

This paper has presented a framework based on Bloom's Taxonomy that can be used to enhance critical thinking among students in assessments that involve the use of AI. Educators can use the framework as a guide to help students approach assessment related to AI with structured and analytical mindsets. This approach can improve their comprehension of AI content, equip them to make informed decisions, and participate in meaningful discussions about AI technologies' ethical and societal impacts.

The infusion of critical thinking skills into the way assessment is handled stimulates students thinking and improves their learning ability. The framework is grounded in questioning techniques or skills by the educator that are closely aligned to stages or types of critical thinking.

References

Aithal, V. and Silver, J. 2023 Enhancing learners' critical thinking skills with AI-assisted technology. Enhancing critical thinking skills with AI-assisted technology (cambridge.org)

Anderson L.W. and Krathwohl D.R. 2001 *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Boston: Allyn and Bacon.

Anderson, L. W. 2014. A comparison of national and transnational reports on the quality of educational research and development in Europe. *Teachers and Teaching* 20(2), 131-147.

Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J. Wittrock, M.C. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A revision of Bloom's Taxonomy of Educational Objectives*. New York: Pearson, Allyn and Bacon.

Bloom, B.S. (eds). 1956. *Taxonomy of Educational Objectives, Handbook I*: The Cognitive Domain. New York: David McKay Co Inc.

Boud, D. 2018. Assessment could demonstrate learning gains, but what is required for it to do so? *Higher Education Pedagogies* 3:1, 1-3.

Cambridge 2023. Enhancing learners' critical thinking skills with AI-assisted technology

Carroll, D. 2018. A review of revised Bloom's taxonomy and other learning taxonomies. *Research in Learning Technology*, 26.

Elsen-Rooney, M. 2023. NYC blocks access to ChatGPT on school networks as cheating fears swirl. [online] Chalkbeat New York. Available at: https://ny.chalkbeat.org/2023/1/3/23537987/nyc-schools-ban-chatgpt-writing-artificial-intelligence

Ennis, R.H. 2018 Critical thinking across the curriculum: A vision. *Topoi*. 37(1), 165–184

Fang, J., Su, H. and Xiao, Y. 2018. Will Artificial Intelligence Surpass Human Intelligence? https://doi.org/10.2139/ssrn.3173876.

Forehand, M. 2010. Bloom's Taxonomy. In M. Orey (ed.) *Emerging Perspectives on Learning, Teaching, and Technology*. Global Text Project.

James, R., McInnes, C. and Devlin, M. 2002. *Assessing Learning in Australian Universities*. Canberra, Australia: Australian Universities Teaching Committee.

Schraw, G., Crippen, K. J. and Hartley, K. 2006. Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*, 111–139.

Spector, M. 2019. Complexity, Inquiry Critical Thinking, and Technology: A Holistic and Developmental Approach. In *Mind, Brain and Technology* (pp. 17–25). Springer, Cham.

Spector.M. and Ma, S. 2019. Inquiry and critical thinking skills for the next generation: from artificial intelligence back to human intelligence. *Smart Learning Environments* Volume 6(8)

Stanford University .2023. AI Will Transform Teaching and Learning. Let's Get it Right.

TEAM International. 2022. Critical thinking is the most valued skill among innovative companies. [online] Available at: https://www.teaminternational.com/critical-thinking/.

The Reboot Foundation. 2018. The State of Critical Thinking: A New Look at Reasoning at Home, School, and Work. Retrieved from https://reboot-foundation.org/w content/uploads/_docs/reboot_foundation_white_paper.pdf .

University of Louisville 2023. What is Critical Thinking? — University of Louisville Ideas to Action. [online] Louisville.edu. Available at: https://louisville.edu/ideastoaction/about/criticalthinking/what.

Van den Akker, J. 1999. Principles and methods of development research. In J. van den Akker, R. M. Branch, K. Gustafson, N. Nieveen and Plomp, T (eds). *Design Approaches and Tools in Education and Training* (pp. 1–14). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Van den Berg, G. and du Plessis, E. 2023. ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking, and openness in teacher education. *Education Sciences 13*(10).

Walter, Y. 2024. Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education* 21, 15.

Wu, D., Xing, D. and Lu, C. 2019. The effects of learner factors on higher-order thinking in the smart classroom environment. *Journal of Computers in Education* 6(4), 483–498.

Zaphir, I. and Lodge, J. M. 2023. Is critical thinking the answer to generative AI? *Times Higher Education*.

Zaretskii, V. K. 2009. The zone of proximal development: What Vygotsky did not have time to write. *Journal of Russian and East European Psychology, 47*(6), 70–93.

The Use of Artificial Intelligence Tools in Research

S. Pramjeeth and P. Ramgovind
The Independent Institute of Education
spramjeeth@varsitycollege.co.za

Abstract
The recent rapid introduction of artificial intelligence tools and their prevalence in higher education has resulted in many higher education institutions being reactive rather than proactive. This paper reports on a study undertaken to explore the use of artificial intelligence tools and large language models by researchers and supervisors of postgraduate research projects in a private higher education institution in South Africa. The purpose is to understand whether these tools simply make researchers more efficient, or they make them become lazy. The results of the study indicated that many researchers and supervisors of postgraduate research projects were making use of artificial intelligence tools and large language models. The institution itself has not yet developed a position either in favour of, or against the use of these tools in research. Based on the findings of the study, it is recommended that the institution should develop guidelines to direct researchers and supervisors of postgraduate research projects towards ethical use of these tools in research. Another recommendation is that there should be a community of practice on the use of these tools in research.
Keywords: Artificial intelligence, large language models, private higher education, research, tools

Introduction and background
Industries such as manufacturing, healthcare, logistics and supply chain, music, food, and travel, have, over time, experienced disruption as a result of technological developments (Afrianto *et al*., 2020; Ali *et al*., 2021; Kurniasih *et al*., 2022). Ordinarily, any form of disruption is viewed negatively, given that there is stasis in the operations and activities of organisations. However, the disruption caused by technological developments, when properly understood, can provide organisations with countless benefits, from sustainability to competitive advantage and customer satisfaction through applications such as mobile technology, simulations, drones and nanotechnology (Bai *et al*., 2020; Mondal *et al*., 2023). Although this is the case, the educational system has steadfastly resisted the disruption despite all of this change (Popenci, 2023). Despite claims to the effect that "disruptive" educational technologies, such as the massive open online courses (MOOCs), would revolutionise both the accessibility and quality of education for all, these claimed benefits have not been realised to their fullest, with low rates of student retention and low levels of engagement in learning activities among enrolled students (Hengtao and Yingxiao, 2023; Jones and Sharma, 2019). Regarding artificial intelligence (AI) as a component of the latest technological developments, Chen *et al*., (2020) noted that the incorporation of AI into the education sector must be seen as more than supercomputers embedded into a computer system, but understanding the true reach of AI in education from smart buildings to robotics and enhanced personalised learning. AI tools into are viewed by many beneficial in education for addressing challenges and accelerating progress made towards achieving Sustainable

Development Goals as outlined by the United Nations. Nomerovska (2023) notes that the role of AI in education is to optimise the learning and teaching process through personalised learning algorithms.

The Covid-19 pandemic has been a catalyst for educational institutions to relook at their teaching and learning strategies by adopting online and hybrid learning and increasingly integrating the use of technology in the classrooms (Zhao, 2020). Through a systemic literature review of over forty papers that focused on education during Covid-19, Abu Talib *et al* (2021) found that higher education institutions and the government were grossly underprepared for the immediate migration to technology-driven teaching, learning and working. However, that abrupt pivoting to technology in education has brought about changes in curriculum, pedagogy, the dynamics between lecturers and students, facilitation of assessments, and the location and time of lecturers (Mbhiza, 2021; Zhao and Watterston, 2021).

While most higher education institutions are coming to terms with the lasting impact and teachings brought about by Covid-19, a new technology-driven disruption has come to the forefront of education in the form of AI. November 2022 saw the entrance of ChatGPT, a generative AI intelligence language model that has the potential to disrupt the education landscape significantly, leaving many educators and institutions fearing that students will use the AI tools to complete their assessments, resulting in no learning and non-attainment of competency (Popenici, 2023). ChatGPT catapulted AI-driven tools into the spotlight and prompted higher education institutions to respond either positively or suspiciously towards the incorporation of AI-driven tools in teaching, learning and assessment. Those that have responded positively have incorporated AI tools into their academic and non-academic operations, while those that have responded suspiciously continue to debate the merits and ethics of AI and the its impacts on the traditional values of higher education learning, teaching and research (Singh and Hiran, 2022). They are faced with the challenge of understanding how to leverage the strengths AI, and mitigate their weaknesses, the evidenced gaps in digital competencies, resistance to change, access to technological devices, and disparities in socio-economic conditions that drive inequality (Mhlanga *et al*., 2022).

Statement of the problem/ issue addressed

AI-driven technologies are have disrupted the higher education sector. Understanding how higher education institutions should prepare academics, students, researchers, and supervisors to use AI tools to succeed in the era of the Fourth Industrial Revolution is crucial. Operating within a world in which the use of AI is becoming a norm, it is important for higher education institutions to begin to relook at their models of teaching and learning, as well as research. Failure to do that results in a *laissez faire* situation in which students, educators and research will use these tools in unregulated manner, which could have repercussions on quality of teaching and learning, as well as research.

Aim and objectives

This paper reports on a study undertaken to explore the use of artificial intelligence tools and large language models by researchers and supervisors of postgraduate research projects in a private higher education institution in South Africa. The main objectives of the study to (a) explore the perceived influence AI tools and large language models like ChatGPT and BING's Copilot on students' research skills, writing abilities, referencing, and expectations for doing research in postgraduate degrees; (b) understand how the use of AI tools in research can allow students and researchers to attain the best learning and research experiences; and (c) understand whether these tools simply make researchers more efficient, or they make them become lazy.

Theoretical framework

A constructivist theoretical framework and the technology acceptance model (TAM) underpinned this study. Constructivism, postulated by Piaget Jean and developed over time by Lev Vygotsky as a teaching and learning theory, has, as its fundamental premise the understanding that instruction is most effective when it is centred on the needs of the individual students, allowing them to instinctively "construct" their own knowledge through their experiences and interactions with the world over time (Biggs and Tang, 2011). In the context of the study reported in this paper, constructivism holds that researchers are accountable for their learning, comprehension, and creation of new knowledge, while the AI's role is to support the researcher, facilitate information gathering and synthesis, and act as a co-explorer based on the prompts created by the researcher. Thus, the AI tool serves as a tool to aid and augment learning and knowledge creation. However, the researcher/student also requires institutional support to gain relevant knowledge, skills and competencies in using the AI tools.

The TAM framework (Davis, 1993) describes how users come to embrace and employ new technologies. According to the paradigm, perceived usefulness (PU) and perceived ease of use (PEOF) are the two main criteria influencing consumer acceptance of a technology. A technology is deemed useful (PU) by the users if they believe it will make it simpler for them to achieve their goals. The belief of the users that using the technology will be easy and simple to use is referred to as perceived ease of use (PEOF) (Davis, 1993).

Constructivism offers valuable insights into how researchers engage with AI tools in their work. It highlights the importance of active learning, adaptability, and social interactions. This framework was useful in the study by engendering the understanding of how the use of AI tools in research can allow students and researchers to attain the best learning and research experiences. However, it has limitations when applied to the complex and rapidly evolving field of AI research, particularly in objective data analysis, the complexity of AI algorithms, and ethical considerations. The TAM framework was useful in understanding the perceptions of researchers and supervisors regarding the usefulness (PU) of the AI tools as regards completing their research, and how useful the tool can be to their students. In terms of PEOF, this can be used to assess if the users of the AI tools found it simple and easy to use. The easier and less complex technology is, the higher and quicker the adoption rates.

Benefits of using AI in research

One of the main benefits of AI for research in higher education is that it can enhance the efficiency, quality, and innovation of research processes and outcomes. AI can help researchers collect, analyse, and synthesise large amounts of data from various sources, such as academic publications, social media, online platforms, and sensors (Zhang *et al*., 2021). AI can help researchers to perform complex and tedious tasks faster and more accurately. For example, AI can be used to conduct literature reviews, synthesise information, identify gaps and trends, and generate hypotheses (Zawacki-Richter *et al*., 2019). The use of AI in research can be seen through the acceleration of fundamental research as outlined by Xu *et al*. (2021). The authors explain that new research and applications are expeditiously emerging as a result of AI assistance, AI infrastructure, algorithms and frameworks. AI can also be used to automate data collection, processing, analysis and visualisation, reducing human errors and biases and enabling researchers to handle large and diverse datasets (Liu *et al*., 2020).

Furthermore, AI can be used to generate novel and creative outputs, such as texts, images, sounds and videos, that can inspire new ideas and perspectives for research (Shneiderman *et al*., 2020). Zhang and Aslan (2021) reported a comprehensive review of selected empirical studies on AI in education (AIEd) using multiple methods, including bibliometrics, content analysis and categorical meta-trends analysis. They highlighted the current state of AIEd research, its proven and potential benefits for education, and its future directions.

Another benefit of AI for research in higher education is that it can facilitate interdisciplinary and transdisciplinary collaborations and communications among researchers, educators, and stakeholders from across geographical and disciplinary boundaries, as well as access and leverage diverse expertise, resources, and perspectives (Kovanović *et al*. 2019).

AI can support researchers in disseminating their findings and engaging with wider audiences, such as through chatbots, interactive visualisations, and open-access platforms (Kothari, 2023). AI can also enable new forms of research that were not possible before, such as natural language generation, computer vision, and deep learning (Zawacki-Richter *et al*., 2019). AI can also be used to improve the accessibility and visibility of research outputs by translating them into different languages, formats and media and by optimising them for search engines and social media platforms (Alshater, 2022). Moreover, AI can be used to engage various stakeholders and audiences, such as students, policymakers and practitioners, by disseminating their research findings and impacts more widely and providing personalised feedback, guidance and recommendations based on their needs and preferences (Holstein et al., 2019).

Risks and challenges associated with using AI in research

Using AI in research in higher education also poses some challenges and risks that need to be addressed. One of the main limitations of AI is that it may not be able to capture the complexity, nuance, and context of human phenomena and experiences (Surden, 2014). AI may rely on oversimplified assumptions or models that do not accurately represent the reality or diversity of human situations, genders, races, behaviours, and geographic regions (Han *et al*., 2021; Fenwick and Molnar, 2022).

Fabi, Xu and de Sa (2022) in Fenwick and Mohar (2022, p. 2) also mention the inherent biases of AI as well as "lack of transparency and explainability, power imbalance, and liability". As a result, AI may unintentionally amplify and perpetuate unwarranted human biases, limit usage based on diversity, and propagate potential bias in how AI functions (Daugherty *et al.*, 2018; Kiritchenko and Mohammad, 2018). According to Fenwick and Molnar (2022), if AI is humanised and a multi-layered strategy is taken into account, it might pave the way for its evolution to contain traits of "human intelligence, cognition, and behaviour" that support human limitations and defend human goals.

One of the challenges is to ensure the quality and validity of AI-generated outputs, as they may contain errors, inaccuracies or biases that can compromise the reliability and credibility of research (Dignum, 2019) due to the quality or representativeness of the data or algorithms used (Fenwick and Molnar, 2022). For example, Herman (2022) warned that ChatGPT chatbots might generate superficial or plagiarised essays from existing sources.

Another challenge is to balance the human and machine roles and responsibilities in research, as AI may replace or augment human skills and judgments, raising ethical, legal and social issues (Floridi *et al.*, 2018). A third challenge is to foster the trust and acceptance of AI among researchers and other stakeholders, as they may have different expectations, perceptions or concerns about the use and impact of AI on research (Alshater, 2022).

Another challenge of AI for research in higher education is that it may raise ethical and social issues that require careful consideration and regulation. AI may pose threats to the privacy, security, autonomy, and dignity of researchers and research participants. AI may also create ethical dilemmas or conflicts regarding the ownership, authorship, accountability, and responsibility of research outputs and outcomes. Moreover, AI may have unintended or unforeseen consequences for society and the environment at large. For example, Edwards and Cheok (2018) proposed replacing some teachers' roles with robots with AI, which may have implications for the quality of education and the relationship between teachers and students.

AI tools commonly used in research

To provide insight into the AI-driven tools that researchers can use, a summary has been provided in Table 1.

Table 1: Summary of Certain AI Tools Used in Academic Research

AI Tool	**Description**	**Use in Research**	**Limitation of Tool**
ChatGPT	Provides users with free access to basic AI content development, answers questions, translates languages and has personalised interactions to name a few (Kothari, 2023).	Useful tool for providing a list of applicable solutions	Lacks common sense; emotional intelligence; understanding of contexts; difficulty in creating long-form content in a structured format; managing multiple tasks simultaneously; potential for response bias;

			limitation in knowledge; accuracy and grammatical issues; computation costs and the need for fine-tuning (Marr, 2023).
BING's Copilot	This is an AI-generated tool that can be used across Windows 11, Edge browser, Bing searches and Microsoft 365 (Mehdi, 2023).	Useful tool to assist with getting more information from the web and web searches	Limited chats and sessions; lacks creativity in responses; shorter chat responses; limited fit-for-purpose responses; high system requirements; slower chat response time; and noticeable legal and ethical issues (Ivankov, 2024).
Scholarcy	This tool makes use of deep learning technology and reads content-related information, for example, reports, articles and books, and breaks it down into smaller bits of information that highlight key constructs (Warren-Jones, 2023).	Best tool for summarisation	Articles or text that are behind a paywall cannot be processed; the language translation is limited (especially within an African context); and varying result depending on the nature of text inputted (Scholarcy, 2024).
OpenRead	This tool is an AI-driven interactive platform that allows users to engage with, organise and examine various types and forms of content (Ingle, 2023).	Best tool to write literature reviews	It is still in the “upcoming” stage; requires a high-speed internet connection; a compatible device; response bias (Findmyaitool, 2024).
Consensus	This is a cutting-edge search engine that makes use of machine learning and natural language processing (NLP) to examine and evaluate content found on the web (Bello, 2023).	Best tool for generating answers	Limited scope and range; limited in knowledge; may not have the latest information; and limited in quality and reliability (Findmyaitool, 2024).
QuillBot	As an AI writing assistant, QuillBot assists individuals in creating high-quality content through the use of NLP algorithms. It provides the user with improved grammar and style, rephrasing of sentences and overall improved coherence in the work (Cointelegraph, 2023).	Best tool for paraphrasing	Character limitations for paid and free plans; manual paraphrasing is often required; no AI content detection and no Bot to converse with (Roza, 2024).
Elicit	This tool provides its user with the ability to gather and analyse data through NLP approaches. The tool allows the user to quickly analyse a significant amount of content to find patterns, sentiments, and trends (Lim, 2023)	Best tool for finding relevant papers	Cannot evaluate the trustworthiness of a paper; more efficient in some disciplines than others and researcher to confirm findings to avoid bias (Elicit Help Center, 2024)
Semantic Scholar	As an AI-powered academic search engine, the main focus of this tool is scientific content, whereby it analyses research papers, summarises key aspects and creates recommendations based on NLP algorithms and machine learning (Kundariya, 2023).	Best tool for academic search	Has not reached the development maturity or indexed a sufficient number of publications to compare to Google Scholar; primary language coverage is English; and limited in discipline range (Fricke, 2018).

Scite	This tool allows the researcher to determine how many times an article has been cited, together with the content of the citation and proposed other related searches (Warren-Jones, 2023)	Best tool for citation evaluation	A subscription fee is required and it is limited in research domains (Findmyaitool, 2024).
ChatPDF	This tool allows the user to quickly extract information from PDF files and provides solutions to questions. Furthermore, it allows the user to engage in a conversational dialogue with the PDF file (Ingle, 2023).	Best tool for analysing journal articles	May not answer all questions completely or accurately; unable to manage PDFs that are too large; may not support all languages, formats or features; cannot always guarantee originality or authenticity; and may not present misuse or abuse of AI technology (Findmyaitool, 2024).
Trinka	This AI-powered tool is a writing assistant and is designed to identify grammar, spelling and vocabulary errors. It also provides the user with referencing suggestions to ensure the user is compliant with specific referencing guidelines (Warren-Jones, 2023).	Best tool for grammar checks when writing research papers	Limited API access; language limitations; and dependency on AI Technology (Toolify, 2024)
SciSpace	It allows the users to look through, understand and submit scientific articles, serves as a plagiarism checker and provides a number of different options for paper templates (Ingle, 2023).	Best tool for decoding research papers	May not support all citation styles and formations required; and may not be compatible with certain PDF readers (Findmyaitool, 2024).

It is important to note that Table 1 summarises the most commonly used AI tools in research and it is by no means an exhaustive, given the rapid introduction of various AI tools over the last five years. To further contextualise the vastness of AI-driven tools that can be used in research, some examples are PDFgear; Bit.ai; Research Rabbit; Zotero; Mendeley; Tableau; IBM Watson; Knewton; Google Scholar; Gradescope; MagicSchool.ai; ReadCube Papers; (Harshini, 2023; Bello; 2023; Lim, 2023). Each of these AI-driven tools has specific functionalities that when understood and utilised effectively, can provide the user with a significant understanding of content in a shortened period of time. Therefore, it is critical that before engaging with AI-driven tools, the user is aware of the advantages, disadvantages and functionality of the tools to ensure that the tools are utilised appropriately (Klimova *et al*., 2023). Often, it is out of a lack of understanding and unconscious practice that users may unintentionally misuse an AI-driven tool, thus rendering their research unethical (Harris, 2023).

Implication of using AI in academic research

Perkins (2022) highlights the importance for higher education institutions to carefully develop an academic integrity policy that outlines how students and staff may use AI tools such as ChatGPT and specifies what would constitute academic misconduct. Further to this, examples of what is ethical and unethical must be provided, statements must be unambiguous, and the policy should also ensure that training is provided to students and staff (Perkins, 2022).

The research ethics committees in higher education institutions generally address the impact of AI on human participants before research is conducted; however, there are broader implications therefore, adjustments are typically made to the research design and not necessarily from unanticipated risks that can arise from the application or use of new AI (Pournaras, 2023). As a means of addressing these challenges in response to AI models used in research ethics applications, Pournaras (2023) make recommendations that include that (a) individuals must be held accountable for every component of scientific practice; (b) the research ethics panel must comprise of inter-disciplinary reviewers when reviewing applications that have elements of AI; (c) when presenting planned research, the AI tool used must include information pertaining to the version used, prompted entered and responses received; (d) applications that have research hypotheses and questions that seek to address the scope of specific AI tools but are posed out of the scope of the intended AI tool are subject to research integrity and ethical issues and must be treated as high-risk applications by the review panel; (e) the ethics review of the application must develop guidelines for identifying research designs that present either low or high-integrity risks in relation to AI; (f) researchers that engage in research that includes AI must include countermeasures to address plagiarism, biases and inaccuracies. Each ethical review application must address these areas; (g) the motivation and aim of research based on AI or using AI must have merit and should extend further than merely testing AI tool prompts as this lacks scientific rigour and academic inquiry; (h) auditing protocols are necessary when there is an input to an AI tool where it is closed and proprietary. The purpose of this is to ensure that the personal and sensitive information of the participants or the researcher is not unintentionally shared; (i) it is the responsibility of the researcher/s to moderate any output generated by AI tools so that it is not harmful to the designated participants or any special groups; and (j) research ethics panels and regulatory bodies must develop and maintain an agreement on how AI tools can and cannot be used in research.

The South African higher education landscape is no different to its international counterparts in that it too faces the same generic challenges and opportunities of AI and academic research. However, there are nuances to the South African higher education landscape that present specific challenges and opportunities when integrating AI tools in research.

The South African higher education sector is rich and diverse and is a reflection of the melting pot of cultures, traditions and ethnicities that is modern South Africa (Mathibe and Zhou, 2023). In the past, the higher education sector was one of the tools for discrimination. In the post-apartheid dispensation, higher education is an instrument for change and redress of past injustices (DHET, 2022). Given that a significant number of students entering higher education face the challenge of digital literacy, a study by Bosch *et al*. (2023), highlights some of the opportunities that AI brings to South African students, in that first-year students who require assistance with enhancing their writing skills utilise applicable AI tools to strengthen the quality of their writing as well as well as to assist them in adhering to writing conventions. The study also found that AI assisted in making the material more accessible and understandable, especially where students are not first-language English speakers. Therefore, the presence of AI tools seeks to assist entering students to gain the necessary confidence in their academic ability. Further to this Intelliverse (2023) notes that the use of AI in higher education can be utilised to develop multilingual educational tools that will make academic material more accessible to individuals from diverse language groups, which is necessary for South Africa where there are twelve official languages. The use of AI tools in teaching, learning and research can bridge the gap of student success in higher education. However, the challenge is that students are using AI to generate ideas for assessments. As a result, it removes the authenticity and originality of the work produced. As the South African student is rich with cultural and heritage roots, the removal of student authenticity in the academic learning journey is a disservice to the uniqueness of South African graduates (Bosch *et al*. 2023).

Ethical consideration

A discussion on AI in academic research cannot be held in isolation from ethics. The use and application of AI in higher education, especially in research, will have a significant impact both positively and negatively on citizens and society. Therefore, there must be a deliberate understanding of AI in research and ethics to mitigate the potential dangers (Ebell *et al*., 2021). As a result, to understand what a global consensus is on what constitutes ethical AI, five ethical principles have emerged, namely justice and fairness, transparency, non-maleficence, privacy and responsibility (Jobin *et al*., 2019).

Research Methodology

The study reported in this paper was exploratory, guided by the interpretivist paradigm (Creswell and Creswell, 2018). The justification for using this methodology is that there is little or no previous research available to serve as a reference point on the perceived influence of AI tools and large language models like ChatGPT and BING's Copilot on research within the South African higher education landscape (Saunders *et al*., 2019; Rahi, 2017). An open-ended online questionnaire informed by the literature on AI was developed to explore the views of the institution, academics, researchers and students on the use of AI tools in research. It was deemed suitable taking into consideration the geographical spread of the campuses of the institution. The instrument was anonymous, and the confidentiality of the participant's responses was assured. The questionnaire consisted of eight unstructured questions. Participants were also asked to share their opinion on what they believed the perceived impact AI tools and large language models like ChatGPT and BING's Copilot would have on students' research skills, writing abilities, referencing, and expectations for doing research in postgraduate degrees; as well as on how the use of AI tools in research may allow students and researchers to attain the best learning and research experiences and their opinions. The research instrument was assessed and approved by the institutional research ethics committee, and two academics at the institution reviewed the instrument for clarity, with minor changes being recommended, namely, the removal of the demographical questions and minor language errors.

Data was collected through the online survey platform Microsoft Forms from 03 April 2023 – 15 April 2023. The survey link was distributed to the targeted population via email using a non-probability sampling methodology, namely purposive sampling. Participation in the study was voluntary. The survey was closed after 12 questionnaire responses due to data saturation being attained. The software Nvivo 12 Pro was used to analyse the data, with three themes emerging and subthemes.

Results and discussion

The results of presented and discussed under three themes, namely views regarding permissibility of the use of AI, the official position of the institution regarding use of AI, and unsanctioned use of AI.

Permissibility of the use of AI at the institution

Majority of the participants expressed the view that the institution should permit students and lecturers to use AI tools in their research. Their reasons were that globally, education is changing with greater dependence on integration of technology to prepare students to work with technology, especially AI, as AI is not going to fade but continue to evolve and become more entrenched in education and society. Students need to be equipped with the skills to use these tools correctly and productively for the betterment of themselves, the organisation and society. The views of the participants in this regard are in line with the TAM's focus on attitudes and intentions (Davis,1993). The participants' strong preference indicates a positive attitude toward AI, which is a key driver of technology adoption, according to the TAM.

The notion of including generative AI tools such as ChatGPT in teaching and learning in higher education was also positively received by undergraduate and postgraduate students in Hong Kong (Chan and Hu, 2023). While supporting the notion of the inclusion of AI in higher education, students from the medical profession go on to further explain that AI will not replace doctors but can radically transform the healthcare system (Buabbas *et al*., 2023). This is also supported by Kim and Kim (2022), who indicate that educators expressed a positive response towards using AI for learning given that it served as an expert model and provided valuable resources and qualified examples of scientific writing.

The participants who believed that AI should be allowed held the view that it is a good tool for brainstorming and generating ideas to inform their thinking and has good research assistant capabilities to help speed up the research process. Participants who had used AI tools indicated that they found them to assist in providing a basis for their thinking and research questions. Furthermore, they served as good research assistants, improved participants' writing abilities, assisted in their literature review by finding appropriate sources and summarising the articles, and assisted in data analysis, saving the researcher time. However, the participants noted that the output generated by the AI tool must be reviewed by the researcher for accuracy. Nonetheless, there was consensus that the AI tools have ultimately saved them time and made the research process much easier and simpler.

Majority of the participants felt that the tools would expedite the research process and contribute to a richer research output, with a few participants noting that it could make researchers or students lazy if the intent and purpose are questionable. Constructivism emphasises that learning is an active process where individuals engage with and construct knowledge based on their experiences (Vygotsky, 1978). This perspective can be applied to researchers using AI tools. Researchers who perceive AI tools as useful in providing reliable and relevant information will actively engage with AI systems to brainstorm ideas, design experiments, analyse data, and draw conclusions. According to the TAM if one believes the use of technology will enhance their job performance and simplify their job, the acceptance of the technology is quicker (Davis, 1993). Thus, understanding the cognitive processes involved in this interaction is crucial for improving the effectiveness of AI tools in research.

It is important to note that while advocating for the use of AI tools to be permitted at the institution, the majority of the participants were aware of the need to use these tools ethically. To this end, they also advocated for students and academics to be trained on how to use these tools ethically to ensure the integrity of the research and the quality. Some AI tools are often employed to process vast amounts of data and generate insights. Constructivism's focus on subjective interpretation and individual experiences (Vygotsky, 1978) may not fully address the need for objective data analysis and decision-making. Thus, researchers must balance the subjectivity of constructivist approaches with the objectivity required in data-driven research. Ethical issues surrounding AI, such as bias, privacy, and accountability, require careful analysis and action. Constructivism's emphasis on individual interpretation may not adequately address researchers' collective responsibility in ensuring ethical AI use in research.

Official position of the institution on the use of AI

Regarding the official position on the institution on the use of AI in research, all participants indicated that no formal position has been adopted or communicated to the students, academics and researchers. The general consensus among the participants was that there was no clear formal communication on the use of AI by the institution thus placing the supervisors and researchers in a very difficult predicament. This has generated concerns among the students, academics and researchers in the institution.

Social constructivism, an extension of constructivism, highlights the role of social interactions in knowledge construction (Vygotsky, 1978). In the realm of AI tools, collaboration and discussions among researchers play a vital role in shaping their collective understanding and usage of these tools. Examining how researchers interact with each other and with AI systems and their PU and PEOU can offer insights into the social dynamics of research communities. However, the policies and regulations of higher education institutions may affect use of these tools (Bagozzi, 2007), which is not factored into the TAMs framework.

Unsanctioned use of AI within the institution

Majority of the participants confirmed that they used the AI tools when undertaking their research and/or during supervision of postgraduate students. The commonly used AI tools were Quilbot, Grammarly, Research Rabbit, BING's Copilot, ChatGPT and Taguette, with participants indicating they served as valuable research assistants and saved them a lot of time. The participants found these tools beneficial in various ways. For brainstorming ideas, suggesting topics, focus areas and research questions, summarises, and peer-reviewing, CHATGPT and Bing were used. To improve one's writing, Quilbot and Grammarly were used. To assist in finding suitable literature sources and summarises of the paper, Research Rabbit was used. In terms of assisting in data analysis, mainly qualitative analysis, tools such as Otter and Taguette were used. Based on the responses below, participants believe these tools have a positive impact on making it easier to complete their research, and the tools' perceived ease of use will further contribute to the acceptance and adoption of the tools (Davis, 1993).

Participants felt that if their students used the various AI tools correctly, it would assist them in thinking critically, summarising lengthy documents, improving their writing skills and assisting in how to structure their papers, expediting the research process time, improving research skills and assisting in referencing. Participants believed that these tools would enhance students' research and writing skills if they were taught how to use them correctly and responsibly. AI tools can source and summarise the articles speedily and provide the context; however, students are still required to critically evaluate the output and provide their own thoughts and voice. Thus, the education and training of the supervisor is essential so they can equip their students with the correct skills on how to use these AI tools effectively.

Majority of the participants felt that by allowing students to use AI tools in their research, they would be able to attain a better learning and research experience as they would be able to brainstorm ideas/topics, refine their research problems and research questions, evaluate their writing style and shorten the research literature search time and assist in analysing the data quickly. They will be able to easily review and compare different authors views helping them formulate better arguments and apply for critical thinking. However, this would be only achievable if the tools are used correctly, responsibly and ethically. Thus, the appropriateness of the prompts and how these tools and the outputs are being used are very important.

AI tools often employ complex algorithms that are beyond the comprehension of most users. It is important to note that researchers and students may lack the cognitive capacity to construct a deep understanding of these algorithms, especially in fields where AI is not the primary focus. This challenges constructivism's notion of active construction of knowledge. As AI technologies evolve rapidly, it makes it difficult for researchers to construct a stable and deep understanding of AI tools. The constructivist approach may struggle to keep up with the pace of technological change, leading to gaps in researchers' understanding and usage of AI. As technology advances and changes quicker, users, based on their varying degrees of technology acceptance and comfortability, may not be able to learn and adopt these technologies at the same pace.

There is an enhanced awareness of AI and its application, and while students are demonstrating positive attitudes toward the adoption thereof, Almaraz-López *et al*. (2023) note that despite there being a lack of current knowledge and training, there remains the need to expand upon current practices and uses of AI such that real-life cases are presented to students, and the benefits and real limitations of technology are demonstrated.

Interpretation of the findings

Although there is strong evidence of the need to adopt AI in higher education, there is very little direction from the private higher education institution at which the study was conducted. Whilst there is overwhelming support for the use of AI, its use has not been mandated or sanctioned by the private higher education in question. This understanding is supported by Slimi (2021) who explains that while there are obvious benefits to AI in education from an assessment and quality of service stance, it remains the responsibility of the higher education institution to investigate the shortcomings of AI in terms of assessment and ethical limitations. Despite this, there is consensus that training is required – especially regarding its ethical use and extent of use, suggesting that not many have a conscious experience of the use of AIs and its uses/applications. Ocaña-Fernández *et al*., (2019), have noted that one of the most significant challenges facing higher education institutions is the planning, development, and implementation of digital skills to facilitate the training of professionals that are needed to understand the contemporary technological educational and world of work landscape. The importance of writing into policy all the regulatory aspects of AI in the research environment within the higher education institution is evident as it delimits and sets expectations regarding the use of AI and the allocation of resources required to monitor and the time spent needed for the upskilling of researchers, supervisors and students alike in the regulation of activities and the integration of AI therein.

Higher education institutions can minimise the risks to institutional, reputational and academic integrity by training and raising awareness of what AI involves, its uses, and how to implement interventions to prevent its use and abuse in ways that undermine academic integrity. Further, leveraging this knowledge, academics, researchers, supervisors and higher education institutions need to manage this risk by reconceptualising the research space and how AI tools can be used responsibly to improve the efficiency of the research process and serve as a platform to stimulate further thinking and discussion. Thus, the manner in which the human and logical elements of research completion within the ethical parameters need to be brought to the fore when engaging with the subject matter. Thus, if the higher education institution wishes to keep up with technological innovation to remain relevant in a changing education landscape, the adoption and integration of AI tools within the teaching and learning, and research frameworks of the institution, is paramount. This is supported by de Almeida, do Santos and Farias (2021) who indicate that the use of an AI framework will serve as a starting point for the regulation of AI use in higher education. If AI tools are used correctly, integrity will remain intact. Knowledge and training are key in the management of students and the use of AI. There also needs to be increased admission of the use of AI, thus, being transparent in its use (like having users to sign a declaration) is required by students and researchers. In terms of software detection and reporting of the use of AI, there are tools available; however, their accuracy in the detection results has not been very evident; thus, greater reliance on the supervisor and reviewer skills and knowledge of the content is still required.

AI is a powerful tool that can offer many opportunities and advantages for research in higher education. AI is revolutionising academic research by automating research processes such as literature surveys, text generation, and editing (Liu *et al*., 2020). AI tools can help researchers analyse large volumes of literature and summarise relevant information (Zhang *et al*., 2021). AI can assist in identifying patterns and trends in research data, which can help researchers make more informed decisions and draw accurate conclusions (Zawacki-Richter *et al*., 2019). Additionally, AI-powered virtual assistants can also help researchers manage their time and tasks more efficiently, allowing them to focus on more important aspects of their work (Lazar, 2021). Moreover, AI can also aid in identifying potential research gaps and suggesting new research directions based on the analysis of existing literature (Zawacki-Richter *et al*., 2019). This can help researchers focus their efforts on areas that have not been explored yet and contribute to advancing knowledge in their field.

However, AI also has some drawbacks and risks that need to be acknowledged and mitigated. Therefore, it is essential for researchers, supervisors, and institutions to adopt a strategic and holistic approach that integrates education, planning and research when using AI for research in higher education. It is also important to engage in campuswide discussions about the impact of AI on research practices and ethics (Zhang *et al*., 2021), as well as to collaborate with other stakeholders from different fields and sectors. By doing so, researchers can harness the potential of AI while ensuring its quality, integrity, and sustainability. One of the challenges of conducting AI research is anticipating and mitigating AI's potential harm to society. To address this challenge, some AI researchers and organisations have adopted ethics review processes (Srikumar *et al*., 2022) that require them to reflect on their work's ethical and societal implications before or after publication. However, these processes are not standardised or widely adopted, and their effectiveness and impact are not well understood. By doing so, a culture of ethical awareness and responsibility among AI researchers can be fostered. It can also contribute to the development of best practices and norms for ethics review in AI research.

Conclusion and recommendations

The main conclusion is that higher education institutions and their students hold the responsibility to ensure that technology-driven tools such as AI are used in a responsible and ethical manner. Higher education institutions should establish guidelines to encourage the use of AI in education but within acceptable parameters. Only once acceptable guidelines are developed, and students are actively working within these parameters can AI truly become an effective tool in research. When users of AI-driven tools do so with knowledgeable and ethical insights, the research process can be enhanced and deliver superior results. Furthermore, higher education institutions should guard against placing unreasonable restrictions to the use of AI.

The main recommendations are that, firstly, higher education institutions should prioritise continuous training and development focused on the use of the various tools, as well as training on how to use AI ethically, and how to establish if it has not been used ethically. Secondly, higher education institutions should adopt a well-considered position on the use and abuse of AI in research. Such position should be expounded on in institutional academic policies, processes and procedures. Thirdly, researchers at the specific private higher education institution should form a community of practice that should meet monthly to deliberate challenges they are experiencing in using these tools (as well as input from their students) and new ways they have found to use the tools optimally or additional uses that have been established as well as demonstrate new AI tools that they have found to be useful. Fourthly, research aimed at acquired deep insights into AI, its beneficial effects to research, its flip sides, and other aspects, should be institutionalised, because there is limited understanding of AI and its ramifications.

References

Abu Talib, M., Bettayeb, A. M. and Omer, R. I. 2021. Analytical study on the impact of technology in higher education during the age of Covid-19: systematic literature review. *Education and Information Technologies* 26. https://doi.org/10.1007/s10639-021-10507-1

Afrianto, I., Djatna, T., Arkeman, Y., Sukaesih Sitanggang, I. and Hermadi, I. 2020. Disrupting agro-industry supply chain in Indonesia with blockchain technology: current and future challenges. *2020 8th International Conference on Cyber and IT Service Management (CITSM)*, Pangkal, Indonesia, 2020, 1-6, doi: 10.1109/CITSM50537.2020.9268872.

Ali, I., Arslan, A., Khan A. and Tarba, S. Y. 2021. The role of industry 4.0 technologies in mitigating supply chain disruption: empirical evidence from the Australian food processing industry. *IEEE Transactions on Engineering Management*, doi: 10.1109/TEM.2021.3088518.

Almaraz-López, C., Almaraz-Menéndez, F. and López-Esteban, C. 2023. Comparative study of the attitudes and perceptions of university students in business administration and management and in education toward artificial intelligence. *Education Sciences* 13 (6), 609. https://doi.org/10.3390/educsci13060609.

Alshater, M. 2022. Exploring the role of artificial intelligence in enhancing academic performance: a case study of ChatGPT. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4312358.

Amini, L., Chen, C.-H., Cox, D., Oliva, A. and Torralba, A. 2020. Experiences and insights for collaborative industry-academic research in artificial intelligence. *AI Magazine* 41(1), 70-81. https://doi.org/10.1609/aimag.v41i1.5201.

Bagozzi, R. P. 2007. The legacy of the technology acceptance model and a proposal for a paradigm shift. *Journal of The Association For Information Systems* 8 (4), 3. DOI: 10.17705/1jais.00122.

Bai, C., Dallasega, P., Orzes, G. and Sarkis, J. 2020. Industry 4.0 technologies assessment: A sustainability perspective. *International Journal of Production Economics*, 229, 107776. https://doi.org/10.1016/j.ijpe.2020.107776.

Bello, C. 2023. *The best AI tools to power your academic research*. https://www.euronews.com/next/2023/08/07/best-ai-tools-academic-research-chatgpt-consensus-chatpdf-elicit-research-rabbit-scite.

Berente, N., Gu, B., Recker, J. and Santhanam, R. 2019. Managing AI, call for papers. *MIS Quarterly*, 1–5.

Biggs, J. and Tang. C. 2011. *Teaching for Quality Learning at University (4th ed.).* Berkshire: Open University Press.

Bosch, T., Marenet Jordaan, Mwaura, J., Sisanda Nkoala, Schoon, A., Smit, A., Uzuegbunam, C. E. and Mare, A. 2023. South African university students' use of AI-powered tools for engaged learning. *Social Science Research Network*. https://doi.org/10.2139/ssrn.4595655

Buabbas, A. J., Miskin, B., Alnaqi, A. A., Ayed, A. K., Shehab, A. A., Syed-Abdul, S. and Uddin, M. 2023. Investigating students' perceptions towards artificial intelligence in medical education. *Healthcare* 11(9), 1298. https://doi.org/10.3390/healthcare11091298

Casal, J. E. and Kessler, M. 2023. Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics* 2(3). https://doi.org/10.1016/j.rmal.2023.100068.

Chan, C. K. Y and Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20 (1). https://doi.org/10.1186/s41239-023-00411-8

Chen, L., Chen, P. and Lin, Z. 2020. Artificial intelligence in education: a review. *IEEE Access*, 8(2169-3536), 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Choi, S., Jang, Y. and Kim, H. 2022. Influence of pedagogical beliefs and perceived trust on teachers' acceptance of educational artificial intelligence tools*. International Journal of Human-Computer Interaction* 39 (4), 910-922. https://doi.org/10.1080/10447318.2022.2049145.

Cointelegraph. 2023, June 17, *5 AI tools for learning and research*. https://cointelegraph.com/news/5-ai-tools-for-learning-and-research.

Coetzee, J., Neneh, B., Stemmet, K., Lamprecht, J., Motsitsi, C. and Sereeco, W. 2021. South African universities in a time of increasing disruption. *South African Journal of Economic and Management Sciences* 24(1). https://doi.org/10.4102/sajems.v24i1.3739

Cox, A.M. 2021. Exploring the impact of artificial intelligence and robots on higher education through literature-based design fictions. *International Journal of Educational Technology in Higher Education,*18, 3. https://doi.org/10.1186/s41239-020-00237-8

Creswell, J.W. and Creswell, J.D. 2018. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). USA: Sage Publications.

Davis, F. D. 1993. User acceptance of information technology: system characteristics, user perceptions and behavioural impacts. *International Journal of Man-Machine Studies* 38(3), 475–487. https://doi.org/10.1006/imms.1993.1022

de Almeida, P. G. R., dos Santos, C. D. and Farias, J. S. 2021. Artificial intelligence regulation: a framework for governance. *ethics and information technology* 23(3), 505–525. https://doi.org/10.1007/s10676-021-09593-z

DHET. 2022a. *Annual Performance Plan*. https://www.dhet.gov.za/Strategic%20Plans/Department%20of%20Higher%20Education%20and%20Training%20APP%202022-23.pdf.

Dignum, V. 2019. HumanE AI: Toward AI Systems that Augment and Empower Humans by Understanding Us, our Society and the World Around Us. *HumanE AI.* https://www.humane-ai.eu/wp-content/uploads/2019/11/D13-HumaneAI-framework-report.pdf

Daugherty, P.R., Wilson, J. and Chowdhury, R. 2018. *Using Artificial Intelligence to Promote Diversity*. Boston: MIT Sloan Management Review.

Edwards A. and Cheok A.D. 2018. Why not robot teachers: artificial intelligence for addressing teacher shortage. *Applied Artificial Intelligence* 32 (4), 345-360, DOI: 10.1080/08839514.2018.1464286

Ebell, C., Baeza-Yates, R., Benjamins, R., Cai, H., Coeckelbergh, M., Duarte, T., Hickok, M., Jacquet, A., Kim, A., Krijger, J., MacIntyre, J., Madhamshettiwar, P., Maffeo, L., Matthews, J., Medsker, L., Smith, P. and Thais, S. 2021. Towards intellectual freedom in an AI Ethics Global Community. *AI and Ethics*. https://doi.org/10.1007/s43681-021-00052-5

Findmyaitool. 2024, March 13. Consensus. https://findmyaitool.com/tool/consensus.

Findmyaitool. 2024, March 13. OpenRead. https://findmyaitool.com/tool/openread.

Findmyaitool. 2024, March 13. SciSpace. https://findmyaitool.com/tool/scispace.

Findmyaitool. 2024, March 13. ChatPDF. https://findmyaitool.com/tool/chatpdf.

Findmyaitool. 2024, March 13. Scite. https://findmyaitool.com/tool/scite.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V. and Schafer, B. 2018. AI4People—an ethical framework for a good AI society: opportunities, challenges, rules, and principles. *Ethics and Information Technology*. https://pubmed.ncbi.nlm.nih.gov/30930541/

Fenwick, A. and Molnar, G. 2022. The importance of humanising AI: using a behavioural lens to bridge the gaps between humans and machines. *Discover Artificial Intelligence* 2, 14. https://doi.org/10.1007/s44163-022-00030-8.

Fricke, S. 2018. Semantic Scholar. *Journal of the Medical Library Association* 106(1). https://doi.org/10.5195/jmla.2018.280

Han, S., Kelly, E., Nikou, S. and Svee, E-O. 2021. Aligning artificial intelligence with human values: reflections from a phenomenological perspective. *AI and Society* 37, 1383–1395 https://doi.org/10.1007/s00146-021-01247-4 .

Harris, D. E. 2023, June 16. AI is already causing unintended harm. What happens when it falls into the wrong hands*? The Guardian*. https://www.theguardian.com/commentisfree/2023/jun/16/ai-new-laws-powerful-open-source-tools-meta

Harshini. C. 2023, April 23. *Top 10 AI Tools for Academic Research*. Analytics Insight. https://www.analyticsinsight.net/top-10-ai-tools-for-academic-research/

Hengtao, T. and Yingxiao, Q. 2023. Designing MOOCs with Little. *Cogent Education* 9:1, DOI: 10.1080/2331186X.2022.2064411

Holstein, K., McLaren, B. M., and Aleven, V. 2019. Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. *Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK'19)*, 383–392.

Herman, D. 2022, December 9. *The End of High-School English. The Atlantic*. https://www.theatlantic.com/technology/archive/2022/12/openai-chatgpt-writing-high-school-english-essay/672412/.

Ingle, P. 2023, April 16. Best AI tools to power your academic research. *MarkTechPost*. https://www.marktechpost.com/2023/04/16/best-ai-tools-to-power-your-academic-research-2023

Intelliverse.ai. 2023, March 13. The Challenges and Opportunities of AI in African Education. https://www.linkedin.com/pulse/challenges-opportunities-ai-african-education-intelliverse-ai-lab#:~:text=Access%20and%20Infrastructure%3A%20One%20of,crucial%20to%20ensure%20equitable%20access.

Ivankov, A. 2024, March 13. Advantages and Disadvantages of Microsoft Copilot. https://www.profolus.com/topics/bing-chat-review-pros-and-cons/

Jobin, A., Ienca, M. and 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

Jones, K. and Sharma, R. 2019. Sustaining or disruptive: future of mooc innovation in tertiary education still up for grabs. *2019 IEEE Global Engineering Education Conference (Educon)*, 900-909. DOI:10.1109/Educon.2019.8725046

Keleş, P. U. and Aydın, S. 2021. University students' perceptions about artificial intelligence. *Shanlax International Journal of Education* 9 (S1-May), 212–220. https://doi.org/10.34293/education.v9is1-may.4014

Kim, N. J., and Kim, M. K. 2022. Teacher's perceptions of using an artificial intelligence-based educational tool for scientific writing. *Frontiers in Education*, 7. https://doi.org/10.3389/feduc.2022.755914

Kiritchenko, S. and Mohammad, S. M. 2018. Examining gender and race bias in two hundred sentiment analysis systems. ArXiv.org. https://arxiv.org/abs/1805.04508.

Klimova, B., Pikhart, M. and Kacetl, J. 2023. Ethical issues of the use of AI-driven mobile apps for education. *Frontiers in Public Health*, 10. https://doi.org/10.3389/fpubh.2022.1118116.

Kothari, P. 2023, June 23. 7 New AI tools to increase your creativity and online success. *Medium*. https://soulscribed.medium.com/7-new-ai-tools-to-increase-your-creativity-and-online-success-6ac68d6cf8f6.

Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., de Vries, P., Hatala, M., Dawson, S., Siemens, G. and Gašević, D. 2019. Examining communities of inquiry in Massive Open Online Courses: the role of study strategies. *The Internet and Higher Education*, 40, 20-43. https://doi.org/10.1016/j.iheduc.2018.09.001.

Kurniasih, D., Setyoko, P. I. and Saputra, A. S. 2022. Digital transformation of health quality services in the healthcare industry during disruption and society 5.0 Era. International *Journal of Social and Management Studies* 3(5), 139–143. https://doi.org/10.5555/ijosmas.v3i5.225.

Kundariya, H. 2023, September 28. *Best AI Tools for Academic Research*. Enago Academy. https://www.enago.com/academy/guestposts/harikrishna12/best-ai-tools-to-empower-your-academic-research/.

Lazar, I. 2021. *Quantifying the Benefit of Intelligent Virtual Assistants to Improve Meeting Experiences AI-powered virtual assistants reduce costs and improve productivity Q1 2021*. Retrieved September 28, 2023, from https://www.cisco.com/c/dam/en/us/solutions/collateral/collaboration/the-benefit-of-intelligent-virtual-assistants.pdf.

Lim, A. 2023, June 3. The 6 best AI tools for researchers and teachers. MUO. https://www.makeuseof.com/the-6-best-ai-tools-for-researchers-and-teachers/.

Lubinga, S., Maramura, T, S. and Masiya, T. 2023. The Fourth Industrial Revolution adoption: challenges in south african higher education institutions. *Journal of Culture and Values in Education* 6(2), 1–17. https://doi.org/10.46303/jcve.2023.5

Mathibe, M. and Zhou, H. 2023, May 11. Restless Prejudice: Is higher education a right or privilege?https://mg.co.za/thoughtleader/opinion/2023-03-24-restless-prejudice-is-higher-education-a-right-or-privilege/.

Mbhiza, H. 2021. Shifting paradigms: rethinking education during and post-Covid-19 pandemic. *Research in Social Sciences and Technology*, 6(2), 279-289. https://doi.org/10.46303/ressat.2021.31

McCarthy, J. 1997. *What is Artificial Intelligence?* Www-Formal.stanford.edu. https://www-formal.stanford.edu/jmc/whatisai/.

Mehdi, Y. 2023, February 7. R*einventing search with a new AI-powered Microsoft Bing and Edge*, *your copilot for the web*. The Official Microsoft Blog. https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/.

Mhlanga, D., Denhere, V. and Moloi, T. 2022. Covid-19 and the key digital transformation lessons for higher education institutions in South Africa. *Education Sciences* 12(7), 464. https://doi.org/10.3390/educsci12070464.

Learnovate, M. I. M. 2023, June 8. *Best AI Tools for Academic Research*. MIM Learnovate. https://mimlearnovate.com/best-ai-tools-for-academic-research.

Marr, B. 2023, March 13. The top 10 limitations of ChatGPT. https://www.forbes.com/sites/bernardmarr/2023/03/03/the-top-10-limitations-of-chatgpt/?sh=92f48a08f355.

Mondal, S., Das, S. and Vrana, V. G. 2023. How to bell the cat? a theoretical review of generative artificial intelligence towards digital disruption in all walks of life. *Technologies* 11(2), 44. https://doi.org/10.3390/technologies11020044.

Nomerovska, I. 2023, September 12. *how ai is impacting on the education system. Keymakr's Blog Features the Latest News and Updates*. https://keymakr.com/blog/how-ai-is-impacting-on-the-education-system/.

Ocaña-Fernandez, Y., Valenzuela-Fernandez, L. and GarroAburto, L. 2019. Artificial intelligence and its implications in higher education. *Propósitos y Representaciones*, 7(2), 536-568. http://dx.doi.org/10.20511/pyr2019.v7n2.274.

Perkins, M. 2023. Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice*, 20(2), 1-24. https://doi.org/10.53761/1.20.02.07.

Popenici, S. 2023, March 25. Higher education lacks solutions to the challenges of the AI era. *University World News*. https://www.universityworldnews.com/post.php?story=20230321141632370.

Pournaras, E. 2023, May 24. *Science in the Era of ChatGPT, Large Language Models and AI: Challenges for Research Ethics Review and How to Respond*. ArXiv.org. https://doi.org/10.48550/arXiv.2305.15299.

Rahi, S. 2017. Research design and methods: a systematic review of research paradigms, sampling issues and instruments development. *International Journal of Economics and Management Sciences* 6 (2), 1-5.

Roza, N. 2024, March 13. Pros and Cons of Quillbot- Worth Getting in in 2024 (Advantages and Disadvantages Honestly Covered)? https://nikolaroza.com/quillbot-pros-cons/.

Saunders, M., Lewis, P. and Thornhill, A. 2019. *Research Methods for Business Students* (8th ed.). Harlow: Pearson Education Limited.

Scholarcy. 2024, March, 13. Questions, answered. https://www.scholarcy.com/faq#:~:text=Scholarcy%20needs%20access%20to%20the,to%20be%20processed%20by%20Scholarcy.

Singh, S. V. and Hiran, K. K. 2022. The impact of AI on teaching and learning in higher education technology. *Journal of Higher Education Theory and Practice* 22(13), p 135-148.

Slimi, Z. 2021. The impact of AI implementation in higher education on educational process future: A systematic review. *Research Square*. https://doi.org/10.21203/rs.3.rs-1081043/v1.

Srikumar, M., Finlay, R., Abuhamad, G., Ashurst, C., Campbell, R., Campbell-Ratcliffe, E., Hongo, H., Jordan, S. R., Lindley, J., Ovadya, A. and Pineau, J. 2022. Advancing ethics review practices in AI research. *Nature Machine Intelligence* 4(12), 1061-1064.

Surden, H. 2014. Machine learning and law. *Washington Law Review* 89 (1), 87.

Sutton, R. S. 2020. John McCarthy's definition of intelligence. *Journal of Artificial General Intelligence* 11(2), 66-67. DOI: 10.2478/

Toolify.ai. 2024, March 13. Enhance your writing with trinka: the ultimate ai tool for academic and technical writing. https://www.toolify.ai/ai-news/enhance-your-writing-with-trinka-the-ultimate-ai-tool-for-academic-and-technical-writing-1687210.

Vygotsky, L. S. 1978. *Mind in Society*: *The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Wahle, J. P., Ruas, T., Kirstein, F. and Gipp, B. 2022. How Large Language Models are Transforming Machine-Paraphrased Plagiarism. ArXiv Preprint ArXiv:2210.03568. https://arxiv.org/pdf/2210.03568.pdf.

Wang, T., Lund, B. D., Marengo, A., Pagano, A., Mannuru, N. R., Teel, Z. A. and Pange, J. 2023. Exploring the potential impact of artificial intelligence (ai) on international students in higher education: generative AI, Chatbots, analytics, and international student success. *Applied Sciences* 13(11), 6716. https://doi.org/10.3390/app13116716.

Warren-Jones, E. 2022, August 8. The 5 best ai tools for postgraduate research. Scholarcy | the long-form article summariser. https://www.scholarcy.com/the-5-best-ai-tools-for-postgraduate-research/

Xu, Y., Wang, Q., An, Z., Wang, F., Zhang, L., Wu, Y., Dong, F., Qiu, C.-W., Liu, X., Qiu, J., Hua, K., Su, W., Xu, H., Han, Y., Cao, X., Liu, E., Fu, C., Yin, Z., Liu, M. and Roepman, R. 2021. Artificial intelligence: a powerful paradigm for scientific research. *The Innovation*, 2(4), 100179. Sciencedirect. https://doi.org/10.1016/j.xinn.2021.100179.

Zawacki-Richter O., Marín V.I., Bond M. and Gouverneur F. 2019. Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education* 16(1), 39. https://doi.org/10.1186/s41239-019-0171-0.

Zhang, K. and Aslan A.B. 2021. AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, 2, 100029.

Zhang, B., Anderljung, M., Kahn, L., Dreksler, N., Horowitz, M. C., and Dafoe, A. 2021. *Ethics and Governance of Artificial Intelligence: Evidence from a Survey of Machine Learning Researchers*. arXiv preprint arXiv:2105.02117.

Zhao Y. 2020. Covid-19 as a catalyst for educational change. *Prospects* (Paris), 49 (1-2): 29-33. doi: 10.1007/s11125-020-09477-y.

Zhao, Y. and Watterston, J. 2021. The changes we need: education post-Covid-19. *Journal of Education Change* 22, 3-12.

Artificial Intelligence in Higher Education in South Africa: Some Ethical Considerations

T. de Villiers-Botha
Stellenbosch University
tdev@sun.ac.za

Abstract

There are calls from various sectors, including the popular press, industry, and academia, to incorporate artificial intelligence-based technologies in general, and large language models (such as ChatGPT and Gemini) in particular, into various spheres of the South African higher education sector. Nonetheless, the implementation of such technologies is not without ethical risks, notably those related to bias, unfairness, privacy violations, misinformation, lack of transparency, and threats to autonomy. This paper gives an overview of the more pertinent ethical concerns that may result from the deployment of various current artificial technologies in the South African higher education context. It provides a broad overview of the relevant literature on artificial intelligence and distils a few general ethics principles that can serve as guidelines for the ethical development, adoption, and use of artificial intelligence systems. Suggestions are made as to how these might be applied to mitigate the ethical concerns in the South African higher education context. Overall, it argues that artificial intelligence technologies should be adopted if they offer demonstrable benefits to end users and that care should be taken to ensure that any potential harm is adequately addressed.

Keywords: Artificial intelligence ethics, ethics principles, higher education, Large Language Models, risk

Introduction and background

The recent prominence of Large Language Models (LLMs), such as ChatGPT, has given fresh impetus to calls to integrate Artificial Intelligence (AI)-based technologies into various aspects of higher education in South Africa (Govender, 2023; Marwala, 2023). AI technologies are seen as one of the key drivers of the "Fourth Industrial Revolution" (World Economic Forum, 2016; Schwab, 2016) and there is concern that South Africans and Africans as a whole suffer a skills deficit that leaves them ill-prepared for the future job market (Mkansi and Landman, 2021; World Economic Forum, 2023). There are assertions about how such technologies will benefit teaching, learning, and administration in higher education (Waghid, Waghid, and Waghid, 2019; Cele, 2021; Evans, 2021; Sedola, Pescino and Greene, 2021; Marwala, 2023) to the extent that there are calls to realign university administration, teaching, learning, and curricula accordingly (Butler-Adam, 2018; Xing, Marwala, and Marwala, 2018; Obi, 2022; Nwosu, 2023). However, while many of the benefits mentioned could be forthcoming, there is no guarantee that all of them will be realised. Much will depend on how such technologies are developed, deployed, and used. This is a major focus area in the field of AI ethics, which explores the ethical risks posed by AI technologies, something which has not received sufficient attention in the literature on AI in the South African higher education context.

There are reasons to be sceptical about the capabilities of the LLMs and other generative AI systems (Weidinger *et al*., 2022). Even in instances where there is a realistic chance that adopting LLM-based technologies will deliver on their promised benefits, there are important ethical concerns that need to be addressed in order to ensure that these benefits are not undermined by harmful side-effects, and that such benefits are shared equitably among all stakeholders impacted by the technology.

The paper provides a brief overview of the prominent ethical risks associated with the adoption and use of AI-based systems in general, and LLM-based systems in particular, as identified in recent AI-ethics literature. It then discusses five broad ethical principles that can serve as guidelines when assessing the ethical impact of the development, deployment, and use of AI technologies. The paper also considers how these principles might be applied in the context of incorporating AI-based technologies into the higher education sector in South Africa.

Unpacking artificial intelligence

There is no precise, generally agreed upon definition of "artificial intelligence". Roughly, when speaking about AI-systems, people tend to have in mind something along the lines of "computers/machines that display behaviour that can broadly be characterised as intelligent, with "intelligent" here referring to human-like capabilities.[7] In their seminal textbook, Russell and Norvig (2022) describe AI systems as "agents that do the right thing", where the "right thing" is defined as the objective that we give them. Hence, a standard view of AI systems is that they are machines that can act so as to meet "their" human-given objectives. Thus, AI-systems are computational systems that are able to act in an agent-like way so as to meet a given objective. In addition, "AI" is an umbrella term, used to refer to a set of related technologies that make use of computational systems to perform various tasks, including natural language processing, image recognition, speech recognition, data mining for prediction, and online content recommendation.[8]

In the higher education context, Miao, Holmes, Ronghuai, and Hui (2021) divide possible applications for AI into four categories, namely education management and delivery; learning and assessment; empowering teachers and enhancing teaching; and lifelong learning. Examples of specific applications in this context include the automation of admissions and timetables, using "learning analytics" to identify students at risk of failure, intelligent tutoring systems, automated language teaching, automated writing evaluation, and automated discussion-forum monitoring, among others. When it comes to the student- and lecturer-facing applications listed here in particular, it should be emphasised that there is still a lack of robust evidence for their efficacy (Miao, Holmes, Ronghuai, and Hui, 2021). Since much of the

[7] Seeing that there is little agreement on what constitutes "human-like" abilities and on whether machines currently truly exhibit such abilities, it is more accurate to refer to "automated" rather than "intelligent" systems.
[8] A distinction is often made between "narrow" or weak AI and "strong" or artificial general intelligence (Searle 1980; Mitchell, 2019). Narrow AI refers to AI that can perform its specific task very well (even surpassing humans' abilities in that task), but which cannot be deployed to perform a different task. All currently existing AI-systems are narrow systems. Artificial general intelligence (AGI) refers to AI-systems than can perform a wide variety or all tasks that humans can perform, on a par with human abilities. We do not yet have AGI.

current discussion around AI in higher education revolves around the implications of adopting LLMs specifically, it is essential to have a broad understanding of what these systems entail.

LLMs (sometime referred to as “generative AI”) are computational models that have been trained on massive amounts of data to model the statistical properties of the language in their training data (Devlin, 2018; Howard and Rude, 2018; Peters, *et al*., 2018; Radford, 2018; Scao, 2022; Weidinger, *et al*., 2022; OECD, 2023b). These models can then be used to make probabilistic predictions relating to sequences of tokens (words, subsets of words, and other bits of text or pixels), which allows for text (or other outputs, such as images) to be generated that mirrors the training data. Such outputs can be further refined through, *inter alia*, fine-tuning and reinforcement learning from human feedback. Currently, depending on their architecture and training, LLMs are able to create text, images, video, and audio. Examples include OpenAI’s ChatGPT, Google’s Gemini (formerly Bard), Microsoft’s Copilot (formerly Bing chat), and Midjourney. “AI” is now often used as shorthand to refer to generative AI systems. In this paper the term “AI” is used to refer to artificial intelligence systems in general, and specify when referring to LLMs in particular.

Ethics of artificial intelligence

The ethics of artificial intelligence, or AI ethics, involves assessing AI systems to determine if the pose ethical risks. AI ethics is increasingly becoming important because AI systems pose novel problems in as far as they sever objectives and the agency to carry out those objectives; and AI-systems have unprecedented reach, meaning that any harms that arise from their use can occur at scale (Floridi, 2023). In as far as AI systems sever agency and objectives, now there are machines that can make decisions that would previously have been made by humans. Examples include decisions on who to admit to higher education, what information to present to a researcher, or what grade to allocate a student. At least three potential problems arise here: firstly, the systems may, for various reasons, both technical and non-technical, be making these decisions on the wrong basis, leading to bias, unfairness, or other harms (Friedman and Nissenbaum, 1996; Angwin, Larson, Mattu, and Kirchner, 2016; Barocas and Selbst, 2016; Caliskan, Bryson, and Narayanan, 2017; Buolamwini and Gebru, 2018; Dastin, 2018; Noble, 2018; Barocas, Hardt, and Narayanan, 2019; Sheng, Natarajan, and Peng, 2019; Koenecke *et al*., 2020). Secondly, some systems are potentially “black boxes” in that it is often difficult to determine the basis on which any given decision was made. A concern here is the potential lack of recourse if wrong decisions are made (Bryson, 2019; Floridi and Cowls, 2019; Diakopoulos, 2020). At a more fundamental level lies the question of transparency—how can it be determined whether a decision is flawed if basis on which it was made is not known. Thirdly, even if the AI systems meet their objectives flawlessly the very efficacy with which they do so, coupled with their extensive reach, could have unforeseen detrimental consequences. A detailed discussion on these concerns follow.

There are other ethical concerns that arise with the adoption and use of AI systems which have less to do with the technology itself, but more with social factors relevant to adoption and use of the technology. Examples include the technological divide in society and the potential loss of autonomy, as well as the infringement on privacy enterprises that provide these technologies (Floridi and Cowls, 2019).

On a global level, there are three pitfalls that we should note when assessing the potential ethical impacts of the adoption of new technology. The first is technological determinism. This is the view that the widespread adoption of a particular technological innovation is inevitable. This assumption creates a sense of inexorableness and urgency that potentially scuppers any due consideration of the relevant ethical concerns (Bearman, Ryan, and Ajjawi, 2022). Often, a determinist narrative is enthusiastically pursued by the companies that develop and market such technologies, for obvious reasons. The second, and related, pitfall is lending too much credence to overhyped claims regarding the potential benefits of any given technology. No technology is a silver bullet, capable, in and of itself, of solving all manner of social ills. Notwithstanding the benefits associated with a technology, there will always be some side effects of the technology in question. The third is conflating the values at stake. For example, a technology that promises "greater efficiency" may or may not be beneficial on the whole, depending on what it is that is being done more efficiently, how this is being achieved, and who gains from it. "Greater efficiency" does not necessarily translate into benefit for all or even any affected stakeholders.

The overall argument is that ethical risks need to be carefully considered and weighed up against the potential benefits of adopting such systems in various contexts. The argument is not that AI technologies should never be adopted in higher education. Instead, the argument is that the benefits of doing so depend on careful consideration of the capabilities and potential pitfalls of these technologies and of their suitability for particular use cases. Moreover, if the foreseen benefits of a given application can plausibly be said to outweigh the foreseen risks, steps should be taken to mitigate the risks at every point during the development, deployment, and use of the application. Care must also be taken to ensure that the benefits and harms that stem from such adoption are equitably shared by all those affected.

Some ethical concerns

Bias

One of the foremost ethical concerns relating to the wide-scale use of AI-based technologies is that of bias, which is the reflection of cultural stereotypes and/or the taking of actions that unfairly prejudice groups.[9] Hence, "biased" AI systems "systematically and unfairly discriminate against certain individuals or groups of individuals in favour of others" (Friedman and Nissenbaum, 1996). "Unfair" discrimination refers to denying an opportunity or assigning an undesirable outcome to an individual or group on

[9] This understanding of "bias" should be distinguished from *statistical bias*, a conception of "bias" more prevalent in the context of statistics and machine learning. To avoid confusion, Barocas et al. (2019) characterise bias in AI-systems as "demographic disparities in algorithmic systems that are objectionable for societal reasons". I will retain the concept of "bias" here, but note that it refers to such demographic disparities, not statistical error.

grounds that are unreasonable or inappropriate (ibid). Note that this tendency also needs to be systematic for a system to be considered biased.[10]

In their pioneering work on bias in computer-based systems, Friedman and Nissenbaum (1996) identify three broad sources of bias which are pre-existing bias (individual and societal biases that become manifest in the system); technical bias (biases that arises from technical constraints or technical considerations), and emergent bias (bias that arises when a system is deployed in a context with real users). One of the most important insights to draw from their analysis is that bias needs not be the result of intentional design. With the more recent emergence of “big data”[11] and AI systems trained on this data through machine learning[12], this point becomes even more pertinent. Bias may creep in in the measurement phase, where a given state of the world is captured in a dataset; in the machine-learning phase, where the data is turned into a model that summarises patterns in the training data and makes generalisations; in the action *stage* where a given action is taken on the basis of the model’s prediction; and even in the *feedback* that occurs from implementing this system (Barocas, *et al*., 2019).

As Barocas *et al*. (2019) point out, from the outset, subjective decisions and technical difficulties beset the measurement phase, not only because the world is complex, but also because machine learning practitioners often need to define new or relevant categories when defining their “target variable” - that which they are trying to predict. In the machine-learning phase, when a predictive model is trained on data that reflects the messiness of the world, any disparities, distortions, and biases in the training data will be reflected—and possibly amplified—in the model. An example would be the gender stereotypes embedded in the text scraped from the web (Schiebinger, 2014; Caliskan, *et al*., 2017; Noble, 2018;), which can then surface in the output of LLMs (Bender, *et al*., 2021; UNESCO, IRCAI, 2024). Other examples include models that score resumes for programming jobs that discriminate against female applicants due to past biased hiring decisions (Dastin, 2018) and facial recognition software that has much worse accuracy for women of colour than for white men (Buolamwini, *et al*., 2018). Problematically, even when one is aware of such biases in the data and so tries to withhold a demographic category such as gender from the training data, a number of other attributes (“proxies”) in the data may still correlate with the withheld category, which the model could pick up on (Barocas and Selbst, 2016; Barocas *et al*., 2019). In addition, people of demographic minority groups may be underrepresented in the data, or machine learning may work less well for minority groups, if members of majority and minority groups systematically differ in terms of the prediction task (Kearns and Roth, 2019). In the example on university admissions, candidates from historically disadvantaged schools may have systematically lower averages for matric compared to candidates from more advantaged schools, despite being equally capable of academic success at university. If the admissions model is

[10] Some margin of error in automated systems, as with any other system, is all but inevitable.
[11] “Big data” refers to the massive amounts of digital data available as a result of the internet, social media, and the use of billons of smartphones.
[12] “Machine learning” is one of the means of developing AI-systems, where computation is used to discover useful regularities in data (Bryson, 2019). Systems can then be built that exploit these regularities through categorising them, making predictions, or directly selecting actions. Examples include automated fraud detection, insurance pricing, and recommending online content for a user. Since the “learning” and resultant outputs (“decisions”) take place autonomously from direct human intervention, the resultant systems are said to the artificially intelligent.

trained on data where candidates from historically disadvantaged schools are underrepresented, the model will likely be calibrated towards the higher averages of historically advantaged groups and will thus have higher errors rates (false negatives) for the already disadvantaged groups.

In the context of applications built on LLMs, such as ChatGPT, many of these problems of bias also surface (UNESCO, IRCAI, 2024). Any biases in the training data can be reflected in the model output, including sexualised depictions of women, gender biases in output and the generation of other pernicious stereotypes (Sheng *et al*., 2019 Hutchinson*, et al*., 2020; Abid *et al*., 2021; Nozza *et al*., 2021). In addition, underrepresentation in the data of some and overrepresentation of others ultimately leads to the homogenisation of outputs, where dominant views in the training data are reflected and less common, or less *datafied* worldviews and opinions, are lost (Weidinger *et al*., 2022). Many LLMs reflect an Anglophone worldview (Bommasan*, et al*., 2021) as these models are largely trained on English text (Scao, 2022). This can also lead to disparities in performance, such as less facility in underrepresented languages and dialects (Koenecke *et al*. 2020). In addition, models can *amplify* training data biases, thereby exacerbating what is already a pernicious problem (Zhao *et al*., 2017; Wang *et al*., 2019).

In summary, far from being value-neutral computation-based systems, AI-systems can reflect and amplify biases in the societies in which they exist in complex ways. Secondly, biases can also arise as a result of the technicalities involved in building and training AI systems. It is thus unlikely that bias can be completely avoided or corrected for at model level. Hence, while developers have an obligation to attempt to mitigate biases in the systems they develop, users should be made aware of the potential for biased output. If researchers, lecturers, and students were to use LLMs, for example, they would need to know that the output likely represents only the most common or dominant view or group represented in the training data and that some of it may be problematic.

Fairness

The problem of bias in AI systems can give rise to the problem of fairness in the use of the AI systems. One aspect of fairness has to do with whether individuals or groups are affected in same ways, and in a manner is both predictable and justified. Fairness is violated when a classifier model gives different scores to otherwise-identical members of different demographic groups (Obermeyer, 2019). This assumption seems justified in an instance where a hiring algorithm systematically rejects the applications of women that are otherwise similar to those of male applicants (Dastin, 2018). “Fairness” also carries the connotation of “equity”. In this regard, fairness means that equally deserving candidates from different groups should be treated equally. However, as is evident from university admissions algorithm example, treating otherwise identical candidates from advantaged versus disadvantaged schools equally will potentially bar deserving candidates from tertiary education, which also seems unfair (Kearns and Roth, 2019). In both cases, the problem is that an AI-system is potentially reinforcing and even amplifying underlying societal inequities. Moreover, there is no obvious technical remedy, given the problem of proxies. When “fairness” is understood as involving “equity”, it might be required that different groups be treated differently in a given context, such as in the university admissions

example above. Yet, these definitions can be mutually exclusive, depending on our conceptions of "merit" and "equitable treatment", and optimising a model for one of these may lead to unfair solutions on the other understanding of fairness (Kearns and Roth, 2019). Inevitably, optimising for fairness, on whichever conception, also tends to negatively impact accuracy (Kearns and Roth, 2019).
The above problems notwithstanding, proponents of machine learning argue that data-driven decision-making has the potential to be more transparent than human decision-making in that it forces the user to clearly articulate the decision-making objectives and thus *enable then to be explicit about the trade-offs between accuracy and fairness (Barocas, et al*., 2019; Fry, 2019; Kearns and Roth, 2019).

A further problem relating to fairness is that perfect prediction is impossible, which means that mistakes are inevitable (Kearns and Roth, 2019). It is thus important to appropriately temper trust in the outcomes of automated processes. In lieu of doing away with automated decision making, we may need to take steps to ensure that such mistakes are equitably shared, meaning that a particular group should not bear the brunt of erroneous outcomes. Infamously, the COMPAS recidivism-prediction model in the United States of America was shown to have a higher false positive error rate for black defendants and higher false negative error rate for white defendants (Angwin, et al., 2016).

When it comes to using LLMs in an educational context, an important fairness consideration is the potential disparity in performance for different social groups. Of particular concern is the possibility that groups who are already marginalised will be subject to harmful stereotyping and other social biases and/or exclusion (Buolamwini *et al*., 2018; Bender *et al*., 2021; Weidinger *et al*., 2022; UNESCO IRCAI, 2024). This means that some students will potentially carry a heavier burden with regard to possible harm resulting from the use of the LLM, while reaping less benefit than students who are not form the affected marginalised groups. Such inequalities may also be exacerbated by the fact that students may not have equal access to the technology, either due to a lack of hardware or the computational skills required to utilise the technology effectively. Moreover, some LLMs are currently behind a paywall, meaning that students with greater economic means will be able to benefit more from the technology than others.[13] As discussed above, LLMs also risk "locking in" values and views dominant in their training data as well as in their downstream tweaking and applications, potentially leading to the homogenisation of perspectives in those using them (Bommasani *et al*., 2021). If larger LLMs are fine-tuned with specific data for particular uses in the context of teaching and research, care must be taken that the data used is as representative of diverse views and values as possible (at least in as far as is appropriate within the confines of the particular subject matter).

In order to address some of these bias and fairness concerns, higher education authorities and institutions should take steps to minimise the discrimination and exclusion that may come from adopting these systems. UNESCO (2023) recommends identifying and assisting those who do not have access to the necessary hardware, internet connectivity, data, or digital skills needed to use the systems; making provision for students with disabilities or special needs who may face unique challenges in

[13] For example, the ChatGPT version based on GPT-3.5 is free while the better-performing GPT-4 version is behind a paywall (Open AI, 2023a)

accessing these systems; developing criteria and systems for validating AI systems for biases and discrimination; building validation mechanisms to test whether systems are trained on data representative of diversity (including in terms of gender, disability, social and economic status, ethnic and cultural background, and geographic location); and requiring from developers that systems are trained in multiple languages and that support multilingual use; requiring that developers put measures in place to mitigate against promoting dominant cultural norms

Misinformation and other information harms

LLMs are prone to generating false, misleading, and concocted information or poor-quality information (Bender *et al*., 2021; Bommasani *et al*., 2021; Metzler *et al*., 2021; Weidinger *et al*., 2022; Floridi, 2023; Mialon *et al*., 2023; UNESCO, 2023). According to one school of thought, this technology, in itself, will never overcome the problem of generating false information, since LLMs are unable to access or identify information in terms of which they can ground truth (Bender and Koller, 2020; Weidinger *et al*., 2022; Lenat and Marcus, 2023). As Weidinger *et al*. (2022) assert LLMs "are trained to predict the likelihood of utterances. Yet, whether or not a sentence is likely does not reliably indicate whether the sentence is also correct." There is broad agreement that current LLM-based systems are prone to making factual errors and correcting for this with current methods is time-consuming, costly, difficult to scale, and by no means foolproof (Metzler *et al*., 2021; Mialon *et al*., 2023; UNESCO, 2023).

In particularly high-risk domains, such as medicine or law, there is the danger of material harm that may arise from users' acting on misinformation. But outputting misinformation on the basis of statistical prominence can also lead to the further marginalisation of minority opinion (Weidinger *et al*., 2022). An added risk in the context of factual inaccuracy is "algorithm bias" where the users of AI systems are prone to overestimating the capabilities of the system and hence to not be critical of the output of these systems. In the context of education, information resources that deliver factual inaccuracies are highly problematic. If used, it is imperative that students, lecturers, researchers and other stakeholders are made aware of the limitations of these systems and of the potential for algorithmic bias in order to remain appropriately critical towards their output. This shortcoming should also be given due consideration before LLM-based technologies are deployed in the educational setting. In most teaching and research contexts, there seems little value to be had from technologies that are not always factually accurate. It is also important that users develop and retain the skills to effectively search for information outside of these systems in order to assess the veracity of their outputs.

The accessibility and fluency of LLM-based systems makes it extremely easy for students to generate relatively correct text and pass it off as their own. At the same time, it is next to impossible to accurately determine whether text has been generated by an AI (UNESCO, 2023). Universities will have to carefully rethink their assessment practices and take measures to accurately determine whether students are indeed the authors of written work. The most effective methods will probably entail some form of oral assessment to compliment any work not done in a controlled setting. However, South African educators are already time- constrained, and current large classes will make this option less feasible, especially at undergraduate level. Measures will need to be put in place to support lecturers in this regard. It is clear that assessment cannot simply continue as in the past without running the risk of such assessments not reflecting students' knowledge, understanding, or abilities—be it in writing, research or cognition—in any meaningful way.

A related risk is the generation of content that is subject to copyright or consists in intellectual property, given that many large language models contain such material in their training data. Users will need to exercise care and ensure that content generated does not constitute plagiarism or copyright and intellectual property violations. This can prove difficult, however, given that current systems are unable to accurately identify the sources that form the basis of the content that they generate (UNESCO, 2023).

Transparency/accountability

Although automation can help ensure some measure of consistency in decision-making, mistakes are all but inevitable[14] because there are natural limits to prediction, and as Barocas *et al*. (2019) note, a "typical" model deployed in practice may have accuracy of between 0.7 and 0.8. This is better than a random guess, but still leaves room for a substantial number of false negative and false positives. In addition, machine-learning-derived algorithms may develop a decision-making scheme which during training may *seem* to be equivalent that of the human scheme it is being trained on, but the system may, in fact, be reaching its decisions differently and may produce different error patterns. Such learned decision schemes may end up relying on criteria that are objectionable; yet, due to the nature of machine learning, this would be transparent to an observer (Burrell, 2016; Barocas *et al*., 2019).[15] Finally, the system may simply be buggy. Examples of automated decisions that are patently erroneous or unfair are not hard to find (O'Neil, 2016; Obermeyer, 2019; De la Garza, 2020).

The foregoing means that some measure of human oversight is necessary to ensure accuracy of information. Hence, especially when it comes to uses of prediction models that have consequential impacts on people's lives, to establish legitimacy, developers and deployers should be able to justify a given decision scheme in that they are able to explain how the chosen targets relate to goals, they need to validate the accuracy of the deployed system, and allow methods for recourse in case of mistakes (Barocas *et al*., 2019). Yet, the extent to which this is possible and practicable remains in dispute (Coeckelbergh, 2020; Diakopoulos, 2020; Müller, 2023).

In the context of higher education, the issues of transparency and accountability become especially relevant where AI systems are used to assess student work and where researchers use AI technologies to conduct research. In both instances, it is imperative that there remains some mind of human oversight to determine on what basis assessment of student work was done and to ensure that such assessment is fair. In addition, there needs to be avenues for students to query specific assessments

[14] See Barocas et al. (2019) for a detailed discussion.
[15] An example here is the risk-scoring model used by the city of Rotterdam to rank people according to their risk for committing welfare fraud which produced biased outputs. It was found that the system seemed to be taking into account categories such as being a parent, a woman, young, not fluent in Dutch, or struggling to find work (Burgess, et al., 2023).

and the possibility of recourse, if it becomes apparent that a system has erred. In terms of research, it is imperative that researchers remain accountable for their research. This means that researchers need to ultimately take responsibility for the veracity and integrity of their research. They also need to disclose and document their use of LLMs, something which many journals and publishers such as Cambridge University Press already require.

Privacy

Infringement on the privacy of individuals is a major ethical concern with AI-based systems both because of the scale at which such systems are able to access, collect, and process private data, and because of the increasing ease with which even ostensibly anonymised or non-personally identifiable data can be "de-anonymised" (Narayanan and Shmatikov, 2008; Kearns and Roth, 2019; Zuboff, 2019; Véliz, 2020). Consequently, granular surveillance of individuals and blanket surveillance of whole populations by various entities other than legally mandated law enforcement is now possible, including surveillance by other state agencies, businesses, and even individuals (Thompson and Warzel, 2019). In addition, the massive amounts of digital data generated when people use the internet and internet-connected devices and services, is traded between various entities (Müller, 2023). In fact, the business model of most of the internet-based services can be described as consisting of "surveillance capitalism" which involves harvesting in order to granularly target content, be it advertising, or user content such as videos and social media feeds, designed to keep people engaged on a particular platform in order to collect more of their data and target more advertising to them (Vold and Whittlestone, 2019; Zuboff, 2019). This data can also be sold on to anyone, including insurance companies, potential employers, and governments (Véliz, 2020). As the European Commission's High-Level Expert Group on AI put it (2019), AI "enables the ever more efficient identification of individual persons by both public and private entities."

Of particular ethical concern are facial recognition systems and other involuntary methods of identification using biometric data. Such technologies cannot be introduced into the educational setting without consent and, in some instances, the meaningful option to opt out without being prejudiced. Clarity is also needed as to who retains access to the data collected, the purposes it is used for, and the security of the data. Finally, the option sometimes needs to exist for such data to be deleted upon request (sometimes referred to as the "right to be forgotten"), always keeping in mind that simply removing personal identifying information from a given data set does not guarantee its anonymity (Kearns and Roth, 2019).

LLM-based technologies also pose privacy hazards in that they can potentially leak private information present in their training data or entered into them via prompts (Weidinger *et al*., 2022). Such models may also be used to infer sensitive information about individuals. Potentially, malicious actors could leverage LLM-integrated applications to attempt to surreptitiously extract users' data, among other security risks (Greshake, 2023). These risks may be exacerbated if the LLM-technology used in an educational context is provided by an external entity, such as a large corporation, and where users are required to create an account in order to use the products (for example, ChatGPT). This leaves users open to being profiled for purposes such as content curation and online advertising, potentially leaving them vulnerable to manipulation and exploitation (UNESCO, 2023). Hence, students should not be compelled to make use of a commercial AI chatbot such as ChatGPT to complete a research assignment, for example.[16] Care also needs to be taken to ensure that all users of such technologies are made aware of privacy risks that may result. In addition, students, lecturers, and researchers need to have the option of opting out of using such systems if they have legitimate privacy concerns. Institutions should also vet potential AI service providers on their use and handling of any user data that they collect.

Autonomy

The collection of personal data is so widespread because it ultimately gives companies and governments the power to “forecast and influence” human behaviour (Véliz, 2020). This is one area where autonomy becomes a central ethical concern. Autonomy refers to the right to make informed decisions regarding individuals’ own lives, which is a fundamental human right (United Nations, 1948). The concern with digital technologies is that that they, along with insights gained from the analysis of personal data, can give companies, governments, and others unprecedented power to influence the behaviour of human beings and manipulate them. This does not only take place in the form of targeted advertising, but also through targeting online content at people to influences their choices (Susser, 2019). As already mentioned, this can take the form of videos and other online content that is likely to keep people engaged for longer, but which can also result in attempts to manipulate voting and other political behaviour of people (Cadwalladr *et al*., 2018; Pham *et al*., 2022).

An aspect of autonomy that has become even more pertinent with the advent of ubiquitous platforms and, more recently, LLMs is “informational autonomy”. At issue here is the fact that the information that is consumed online is mediated by recommendation algorithms, which help determine and/or rank which users see what content and thus carry the risks of amplification and distortion (Alfano *et al*., 2020; Narayanan, 2023). Recommendation algorithms for online content on platforms such as YouTube “predict” which posts a user are most likely to engage with, given their behavioural data. The result is that the online content that people consume becomes highly personalised to ensure optimum engagement. Hence, people have less control over what they encounter online than they might suppose. This applies to social media, streaming platforms, and even search (to a greater or lesser extent) and includes educational content. The concern is that various platforms and a handful of commercial companies that own them have extensive control over what people are able to access online. This poses a threat to autonomy in that users cannot make informed decisions if they have no knowledge of the content that they are not served with, and they have very little information about and control over the curation processes that mediate their access to the online world (de Villiers-Botha, 2022; Kiri Gunn, 2022; Nys and Engelen, 2022). This dynamic is likely to continue with LLM-based systems deployed by commercial entities and that may be deployed in educational settings

[16] An example might be an assignment where students are required to generate output on a topic with a commercial AI application and then critique it.

(Bommasani *et al*., 2021). Without some insight into how such LLM-based systems fit into the above dynamic and are influenced by commercial considerations, we cannot adequately assess the impact that such systems have on our ability to access relevant information and reach our educational goals.

The risk to autonomy that comes from LLM-based systems is not limited to such systems being deployed by companies whose commercial interests might not coincide with our own interests. There is potential risk from the systems themselves. For example, research suggests that using LLM-based writing assistants can affect users' views (Jakesch *et al*., 2023). A further threat is the homogenisation that such models bring about. Because of their reach, the homogenous worldviews encoded in the dominant LLMs can potentially become default views, robbing users of the possibility of exploring alternative viewpoints and hence meaningfully reflecting on their own and developing their analytic skills.

A final potential risk to autonomy is the possibility that students may fail to acquire important skills if they become too reliant on LLM-based and/or other AI systems. A recent Organisation for Economic Cooperation and Development Report (OECD) (2023a) argues that developments in AI technology means that AI may soon outperform many human beings in many literacy and numeracy skills. This does not necessarily mean that such systems will replace humans in jobs where such skills are required, as jobs generally require complex skillsets. Nevertheless, the OECD forecasts that this will almost certainly affect employment, especially for workers with lower proficiency in literacy and numeracy than machines, which make up a "considerable share of the [OECD member countries'] workforce" (OECD, 2023a, p. 100). Workers best placed to deal with automation will be those with "solid skills" in three key areas: literacy, numeracy, and problem-solving skills in technology-rich environments (digital skills). The education system should be warry of producing students who use such technologies as crutches and who do not develop the literacy, numeracy, and problem-solving skills that may allow them to compete with such technologies. Accordingly, educational institutions should not use AI technologies in ways that will deprive students of opportunities to develop their cognitive abilities and social skills. In essence, the use of technology should clearly contribute to learning and research in a way that would not be possible in the absence of the technology.

Legitimacy

Closely related to the phenomenon of technological determinism is a form of "ethics washing" which involves developers of AI making assertions that possible safety and security ethical risks, particularly are addressed. This can serve to detract from a question of whether a given AI system should be released in the first place. It is not always obvious that implementing AI systems rather than not is ethically justifiable. An example is surveillance technology (High-Level Expert Group on AI, 2019) which raises the question ethical legitimacy which is whether anyone should have such unprecedented, unbridled ability to track the movement of people.

An example in the context of AI-powered applications for the higher education sector might be the adoption of LLM-based tutors. While the prospect of the "personalised teaching" is attractive for many reasons, careful consideration needs to be given to whether the technology available is capable of delivering on this promise (UNESCO, 2023). As (Bommasani *et al*., 2021) point out, at a minimum, such a model would need to have an "understanding" of the subject matter at hand; need some capacity for determining what misconceptions a student might have about subject matter in order to address these effectively; and would need to have some kind of appreciation of pedagogy. These are highly complex capabilities that current automation technologies are unable to replicate.[17]

Scholars highlight the risks that need to be mitigated before adopting automation in teaching, most notably risks to "professional authority, institutional accountability, and public policymaking in education" (Zeid, 2020). Zeid points out that using technology that allows for "outsourcing instruction, assessment, and credentialing functions to [edtech] companies, leads to outsourcing more fundamental decisions to them as well." Lecturers do not only draw on their own skills, expertise, and experience to make decisions regarding instruction in classroom settings, but they also create curricula, course content, lesson plans, textbooks, syllabi, assessments, and education standards. All of this needs to be in line with specific institutional policies and decision-making structures and with broader national policies and guidelines. International, subject-specific norms, standards, and best practice guidelines also inform all of the above functions. In the incorporation of educational technologies from external service providers, the designers of the technologies take on part or all of these functions. In addition, the way in which the technologies are designed and trained will have a significant impact on how and what students are taught and how they are evaluated. There is a tremendous amount of individual and institutional expertise that goes into all of these functions, and not all of them are obviously automatable, especially in a way that leaves room for professional judgement. Moreover, education takes place in highly specific contexts, each with unique circumstances and localised needs and values. The South African higher educational context, for example, is highly democratic, with educators and students having a say in institutional policy and practices. This is not easily reconcilable with the unilateral decisions that determine how educational technology is designed and trained in corporate settings. In outsourcing teaching to technology, some pedagogical and policy decisions are also outsourced to the creators and vendors of the technology, without the necessary oversight and transparency. As Zeid (2020) states, "[w]ithout tools for greater transparency, decisions embedded in code shut students, parents, and educators out of this loop. Instead of more readily available classroom teachers or on-site administrators, corporate entities handle these important decisions." Hence, before educational technology is adopted, careful consideration should be given not only to its capability to perform particular functions, but also to how the responsibilities and duties that accompany those functions will be met.

For reasons above, UNESCO (2023) has adopted a position that AI technology is not yet at the point where it can be usefully deployed in a generalised manner in an educational setting. It cautions as follows

[17] Arguably, computer scientists underestimate the expertise required to master all of these capacities (both task specialisation and domain specialisation).

> before significant progress is possible, it is essential that efforts are put into refining foundation models [LLMs and other forms of generative AI] not only through adding subject knowledge and de-biasing, but also through adding knowledge about relevant learning methods, and how this can be reflected in the design of algorithms and models. The challenge is to determine the extent to which EdGPT models can go beyond subject knowledge to also target student-centred pedagogy and positive teacher-student interactions. The further challenge is to determine the extent to which learner and teacher data may ethically be collected and used in order to inform an EdGPT. Finally, there is also a need for robust research to ensure that EdGPT does not undermine student human rights nor disempowers teachers.

Five ethical principles for AI

A plethora of AI-related principles and guidelines have recently emerged, which is much needed but can be overwhelming. Strikingly, AI ethicists (Floridi *et al*. 2018; Floridi, 2019) have pointed out the convergence of many of these guidelines with one another and with the well-established ethical principles proposed by Beauchamp (1979) in the fields of bioethics and the ethics of healthcare namely beneficence, non-maleficence, justice, and autonomy. Accordingly five ethical principles pertinent to AI have been identified. These are discussed in the ensuing subsections.

Beneficence

A central consideration in the development, adoption, and use of AI technologies is beneficence. The technology should bring some kind of benefit. Before developing, adopting, and using given technologies stakeholders should ask, “What *good* will it bring about?”, or “What benefits follow from this technology?” The “good” that accrues could apply to individuals, society at large, or even the environment. In the context of higher education, incorporating technology such as large language models or other generative-AI systems for teaching or research, for example, should ultimately benefit students, lecturers, and researchers, such as by helping them develop important skills or conduct effective and relevant research.

Non-maleficence

Bringing about an apparent good is not yet *sufficient* reason to create and adopt given technology. It is important to also consider possible harm result to human being when the technology is developed and adopted. For example, it is critical to consider whether reliance on large language models potentially undermines important cognitive skills or whether students from marginalised communities will be even more marginalised through underrepresentation or lack of access. Ultimately, possible harm needs to be outweighed by the foreseen benefits. Steps also need to be taken to mitigate foreseen harms at every point in the lifecycle of technologies. Included here would be things like technical robustness and safety, including ensuring that the system is secure, resilient to attack, accurate, representative, reliable, accessible, auditable, assures privacy, and avoids unfair bias.

Autonomy and other rights

When assessing the foreseen benefits and harms of a given technology, it is not enough to argue that the benefits somehow outweigh the harms on aggregate. It is essential to determine whether some foreseeable harm cannot be "outweighed" by the benefits. For example, basic human rights, such as the right to life, liberty, and security of person, are taken to be inviolable—they may not be infringed on the basis that this will bring about a "greater good" for others or society at large. An example in the AI-context is human autonomy. It is generally thought that it is a fundamental human right to be free to make important decisions about one's own life—to act as an agent (High-Level Expert Group on AI, 2019). This right is often also seen as encompassing a variety of other human rights, including the right to dignity, freedom of expression, the right to a private life and privacy, and freedom of movement (United Nations, 1948). Such rights are, of course, also enshrined in the South African constitution (Constitution of South Africa, 1996).

With AI, there are artificial agents to which human beings would potentially cede some of their decision-making powers to. As autonomous agents, human beings may elect to cede some of these powers; however, such a choice should be free and informed. For example, students would need to be made aware if they were being tutored by an AI-tutor. In such a scenario, their data would need the requisite level of protection and they may even need to be given the option of opting out of using the system, if they have legitimate privacy or other concerns.

An influential formulation according to which it can be tested whether or not the right of others' to autonomy comes from Kant (1785) who stated that human beings should always treat people as ends in themselves and never as means to an end. In essence, this means making sure that people are allowed to pursue their own, freely chosen ends and not simply using them as objects for the benefit of others. Clearly, deceiving, manipulating, and exploiting people or covertly exposing them to risks falls foul of this principle.

Justice and fairness

Once it has been determined that a particular technology is beneficial and does not violate fundamental human rights, it needs to be determined whether the benefits and its possible harm would be equitably distributed among the interested and affected parties. Care must be taken not to perpetuate and amplify past injustices, and to ensure that some social groups are not disproportionately adversely affected by AI-systems.[18] Moreover, AI systems should be as inclusive as possible, both in terms of representation and being accessible to different groups and populations. Asymmetries of information about such systems that could lead to asymmetries of power should be taken into consideration (High-Level Expert Group on AI, 2019). Accountability should also be maintained so that redress against decisions made by AI systems is possible.

[18] In our large language model example, an assignment that requires using commercial AI-models such as ChatGPT, for example, may disadvantage students who are unable to pay for the premium version of the product or who do not have the hardware or skills to use the system. Moreover, since most commercial LLMs are trained in Anglophone data (see discussion above), their output would disproportionately reflect worldviews that do not necessarily apply to our context.

Explicability and transparency

To promote and facilitate adherence to the principles discussed above, measures need to be taken to mitigate against the black box nature of automated systems. Those who develop and deploy AI-based systems need to document their design and development, as well as provide clear explanations of their functionalities, including highlighting their proper uses and limitations (High-Level Expert Group on AI, 2019; National Institute of Standards and Technology, 2023)

Owing to their architecture and training, LLMs pose a challenge to the criterion of explicability. Currently, it is a challenge to understand what some of these models can do, why they output certain behaviours, and how they do so (Bommasani *et al*., 2021). Even in task-specific neural network models, it is difficult to know what a model will do with a particular input. LMMs are capable of performing a diversity of tasks, as they tend to be trained on vast amounts of data (OpenAI, 2023b), and yet details of the training data are often not provided. This is at odds with the requirement that the provenance of training data should be maintained and the AI system's decisions should be attributable to subsets of its training data (National Institute of Standards and Technology, 2023) and strengthens the arguments that LLM-based technologies should not be deployed in high-risk settings, where it is necessary that outcomes or decisions be verifiable.[19]

In the educational setting in general, it is vital that possible use cases be tested, and deployment should occur on an evidence-based basis (UNESO, 2023). Applications should be validated by interdisciplinary experts before being adopted. After adoption, the characteristics, limitations, and potential shortcomings of the AI system need to be clearly communicated to users. Users should also always be made aware when they are interacting with an AI system. To prevent algorithmic bias and anthropomorphising, it should be clearly communicated that the system has no capacities of feeling or understanding.

Conclusion and recommendations

There is much hype around AI technologies currently. Whereas they hold good prospects of bringing about many benefits to human beings, it is important to keep in mind that these tools, as with other tools, can be both beneficial and harmful. Because of the complex nature of the technology, the many ways it could potentially be deployed and used, and its potential reach, care needs to be taken that possible harms are identified and mitigated. Using an ethical framework like the one above to do such an assessment is a start. Ultimately, expertise from many fields will be needed to ensure that any AI system that is adopted is deployed in a way that benefits all who will be affected by it. At a policy level, it is recommended that the following suggestions are taken into account in the development of policies relating to the incorporation of AI into higher education:

[19] This obviously applies to a research setting where factual errors are unacceptable. If researchers were to make use of LLMs in their research in some capacity, it is imperative that they verify all generated output against trusted and legitimate sources.

- Adopt a human-centred, ethics-first approach to incorporating AI into management, teaching, learning, and assessment. The primary purpose of applying AI should be to enhance the capacities of students, lecturers, support staff, and other stakeholders in the field.

- Develop a master plan for using AI for higher education, drawing on a broad range of interdisciplinary expertise, including educators, researchers, computer scientists, engineers, and ethicists, to ensure that the adoption of such systems serves educational priorities and meets genuine educational needs in an ethical manner.

- Adopt an evidence-based approach. Make use of pilot testing, monitoring, and evaluation, and building up an evidence base to test for efficacy and to identify and mitigate any ethical harms before wide-scale adoption.

- Prioritise and enhance AI literacy for all stakeholders in the higher education sector, focusing not only on how to use AI-based technologies, but also on how they work and ethical concerns and responsibilities around their use. Foster the development of AI and general digital skills in all stakeholders in the higher education space.

- Foster local AI talent and local AI innovations for higher education to ensure the development of AI-systems appropriate to the local educational context.

References

Abid, A., Farooqi, M. and Zou, J. 2021. Persistent anti-Muslim bias in Large Language Models. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society,* 298-306.

Alfano, M.; Fard, A. E.; Carter, J. A.; Clutton, P. and Klein, C. 2020. Technologically scaffolded atypical cognition: The case of YouTube's recommender system. *Synthese* (1-2): 1-24.

Angwin, J., Larson, J., Mattu, S. and Kirchner, L. 2016. Machine Bias. *ProPublica*, 23 May 2016.

Barocas, S. and Selbst, A. D. 2016. Big data's disparate impact. *California Law Review*, 103(4): 671-732.

Barocas, S., Hardt, M. and Narayanan, A. 2019. *Fairness and Machine Learning: Limitations and Opportunities*. https://fairmlbook.org

Bearman, M., Ryan, J. and Ajjawi, R. 2022. Discourses of artificial intelligence in higher education: a critical literature review. *Higher Education,* 86(2): 369-385

Beauchamp, T. L. 1979. *Principles of Biomedical Ethics*. New York: Oxford University Press

Bender, E. M. and Koller, A. 2020. Climbing towards NLU: On Meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, July 5-10, 2020: 5185–5198

Bender, E. M., Gebru, T., McMillan-Major, A. and Shmitchell, S. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 3 March 2021, 610-623

Bizcommunity 2023. AI and higher education: What must tertiary institutions consider? *bizcommunity*, 22 https://www.bizcommunity.com/Article/196/499/236028.html.

Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E. and Brynjolfsson, E. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

Bradshaw, S., Neudert, L.M. and Howard, P. 2019. Government Responses to Malicious Use of Social Media, NATO StratCom Centre of Excellence, Riga, Working Paper.

Bryson, J. J. 2019. The artificial intelligence of the ethics of artificial intelligence: an introductory overview for law and regulation. In Dubber, M. Pasquale, F. and S. Das, S. (eds). *The Oxford Handbook of Ethics of Artificial Intelligence*. Oxford: Oxford University Press, pp3-25.

Buolamwini, J. and Gebru, T. 2018. Gender shades: intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability, and Transparency. Proceedings of Machine Learning Research*, 81: 1–15.

Burgess, M., Schot, E. and Geiger, G. 2023. This algorithm could ruin your life. *Wired*, 6 March, 2023.

Burrell, J. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data and Society*, 3(1).

Cadwalladr, C. and Graham-Harrison, E. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *Guardian*, 17 March 2018.

Caliskan, A., Bryson, J. J. and Narayanan, A. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334): 183-186.

Cambridge University Press, 2023. Authorship and contributorship.

https://www.cambridge.org/core/services/authors/publishing-ethics/research-publishing-ethics-guidelines-for-journals/authorship-and-contributorship#ai-contributions-to-research-content.

Christie, J. 2020. The Post Office horizon IT scandal and the presumption of the dependability of computer evidence. *Digital Evidence and Electronic Signature Law Review*, 17: 49-70.

Coeckelbergh, M. 2020. *AI Ethics*. Cambridge, Massachusetts: MIT Press

Dastin, J. 2018. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 11 October 2018.

De la Garza, A. 2020. States' automated systems are trapping citizens in bureaucratic nightmares with their lives on the line. *Time*, 28 May 2020.

De Villiers-Botha, T. 2022. Re-assessing Google as epistemic tool in the age of personalisation. In *Proceedings of SACAIR2022 Conference, the 3rd Southern African Conference for Artificial Intelligence Research*, Online: SACAIR2022 Organising Committee, 323-337.

Devlin, J., Chang, M., Lee, K. and Toutanova, K. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186

Diakopoulos, N. 2020. Transparency. In Dubber, M., Pasquale, F. and Das. S. (eds). *The Oxford Handbook of Ethics of AI*. Oxford: Oxford University Press, pp 197-213

European Commission. 2023. Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) 2021/0106 (COD)

Evans, N. 2021. Nurturing the African higher education teaching and learning ecology using data, information and knowledge from our physical, digital and biological learning spaces. In Ocholla, O., Evans, N. and J. Britz, J. (eds). *Information Knowledge and Technology for Development in Africa*. Cape Town: AOSIS, pp85-102.

Fayyad, U. 2001. The digital physics of data mining. *Communications of the ACM*, 44(3): 62–65.

Floridi, L. 2023. AI as agency without intelligence: on ChatGPT, Large Language Models, and other Generative Models. *Philosophy and Technology*, 36(1): 1-7.

Floridi, L. and Cowls, J. 2019. A united framework of five principles for AI in society. *Harvard Data Science Review*, 1(1).

Floridi, L., Cowls, J., Beltrametti, M., *et al*. 2018. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4): 689-707.

Friedman, B. and Nissenbaum, H. 1996. Bias in Computer Systems. *ACM Transactions on Information Systems*, 14(3): 330–347.

Fry, H. 2019. *Hello World: Being Human in the Age of Algorithms*. New York: W. W. Norton and Company.

Govender, R. 2023. The impact of artificial intelligence and the future of ChatGPT for mathematics teaching and learning in schools and higher education. *Pythagoras, 41*(1): a787.

Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T. and Fritz, M. 2023. More than you have asked for: A comprehensive analysis of novel prompt injection threats to application-integrated Large Language Models. *arXiv preprint arXiv:*2302.12173.

High-Level Expert Group on AI. 2019. *Ethics Guidelines for Trustworthy AI*. Brussels: European Commission.

Howard, J. and Ruder, S. 2018. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:*1801.06146.

Hutchinson, B., Prabhakaran, V., Denton, E., Webster, K., Zhong, Y. and Denuyl, S. 2020. Social biases in NLP Models as barriers for persons with disabilities. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5491-5501.

IEEE. 2019. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition., Online: IEEE.

https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html

Jakesch, M.; Bhat, A., Buschek,D., Zalmanson, L., and Naaman, M. 2023. Co-writing with opinionated Language Models affects users' views. *Association for Computing Machinery*, 1–15.

Kant, I. 1785. *Groundwork for the Metaphysics of Morals*. New York: Oxford University Press.

Kearns, M. and Roth, A. 2019. *The Ethical Algorithm The Science of Socially Aware Algorithm Design*. Oxford: Oxford University Press.

Kiri Gunn, H. 2022. Is there a duty to disclose epistemic risk?. In F. Jongepier, F. and Klenk, M (eds). *The Philosophy of Online Manipulation*. London and New York: Routledge, pp 275-291.

Koenecke, A., Nam, A.J., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J.R., Jurafsky, D., and Goel, S. 2020. Racial disparities in automated speech recognition. Proceedings of the National Academy of Sciences of the United States of America, 117*:* 7684–7689.

Lenat, D. and Marcus, G. 2023. Getting from generative AI to trustworthy AI: What LLMs might learn from Cyc. *arXiv preprint arXiv:*2308.04445.

Marwala, T. 2023. Resistance is futile – South Africa must urgently adapt to the new age of artificial intelligence. *Daily Maverick*, 7 December 2023.

Metzler, D., Tay, Y., Bahri, D. and Najork, M. 2021. Rethinking search: Making domain experts out of Dilettantes. *ACM SIGIR Forum*, 55(1): 1–27.

Mialon, G., Dessì, R., Lomeli, M., Nalmpantis, C., Pasunuru, R., Raileanu, R., Rozière, B., Schick, T., Dwivedi-Yu, J., Celikyilmaz, A. and Grave, E. 2023. Augmented language models: a survey. *arXiv preprint arXiv:*2302.07842.

Mitchell, M. 2019. *Artificial Intelligence: A Guide for Thinking Humans*. Penguin UK.

Müller, V. C. 2023. Ethics of artificial intelligence and robotics. *The Stanford Encyclopaedia of Philosophy* (Fall 2023 Edition). Zalta, E. N. and Nodelman, U (eds.)

https://plato.stanford.edu/archives/fall2023/entries/ethics-ai/

Narayanan, A. and Shmatikov, V. 2008. Robust de-anonymization of large sparse datasets. *IEEE Symposium on Security and Privacy*, 111-125

Narayanan, A. 2023. *Understanding Social Media Recommendation Algorithms*. New York: Knight First Amendment Institute.

National Institute of Standards and Technology. 2023. *AI Risk Management Framework*. Online: U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

Noble, S. U. 2018. *Algorithms of Oppression*. New York: New York University Press.

Nozza, D., Bianchi, F. and Hovy, D. 2021. HONEST: Measuring hurtful sentence completion in Language Models. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, 2398-2406

Nwosu, L. I. 2023. Fourth Industrial Revolution tools to enhance the growth and development of teaching and learning in higher education institutions: A systematic literature review in South Africa. *Research in Social Sciences and Technology, 8*(1): 51-62.

Nys, T. and Engelen, B. 2022. Commercial online choice architecture: when roads are paved with bad intentions. In Jongepier, F and Klenk, M. (eds). *The Philosophy of Online Manipulation*. London: Routledge, pp135-155.

Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464): 447-453.

OECD. 2019. *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449.

OECD. 2023a. *Is Education Losing the Race with Technology?: AI's Progress in Maths and Reading.* Paris: Educational Research and Innovation, OECD Publishing.

OECD. 2023b. AI language models: Technological, socio-economic and policy considerations. In *OECD Digital Economy Papers*, 352. Paris: OECD Publishing.

Office of Science and Technology Policy. 2022. *Blueprint for an AI Bill of Rights*. The White House https://www.whitehouse.gov/ostp/ai-bill-of-rights/ .

OpenAI. 2023a. ChatGPT. https://chat.openai.com/.

OpenAI. 2023b. GPT-4 Technical Report. https://doi.org/10.48550/arXiv.2303.08774

Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K. and Zettlemoyer, L. 2018. Deep Contextualized Word Representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies Volume 1.* New Orleans, Louisiana: Association for Computational Linguistics, 2227–2237.

Pham, A., Rubel, A. and Castro, C. 2022. Social media, emergent manipulation, and political legitimacy. In Jongepier, F. and Klenk, M (eds). *The Philosophy of Online Manipulation*. New York and London: Routledge, pp 353-369.

Radford, A. and Narasimhan, K. 2018. Improving language understanding by generative pre-training. *OpenAI*.https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf./

Russell, S. J. and Norvig, P. 2022. *Artificial Intelligence A Modern Approach*. Fourth ed. Harlow: Pearson Education Limited.

Scao, T.L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A.S., Yvon, F., Gallé, M. and Tow, J., 2022. Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv*:2211.05100.

Schiebinger, L. 2014. Scientific research must take gender into account. *Nature*, 507: 9.

Schwab, K. 2016. *The fourth industrial revolution.* Coligny: World Economic Forum.

Searle, J. R. 1980. *Minds, brains, and programs. Behavioural and Brain Sciences,* Vol. 3, No. 3, pp. 417–424.

Sedola, S., Pescino, A. J., and Greene, T. 2021. *Blueprint: Artificial Intelligence for Africa.* Kigali: Smart Africa, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) and GFA Consulting.

Sheng, E., Chang, K.W., Natarajan, P. and Peng, N. 2019. The woman worked as a babysitter: On biases in language generation. *arXiv preprint arXiv*:1909.01326.

Singer, P., 1986. Introduction. In Singer, P. (ed). *Applied Ethics*. Oxford: Oxford University Press, pp1-8.

Susser, D., Roessler, B. and Nissenbaum, H. 2019. Technology, autonomy, and manipulation. *Internet Policy Review*, 8(2).

Thompson, S. A. and Warzel, C. 2019. Twelve million phones, one dataset, zero privacy. *New York Times*, 19 December 2019.

UNESCO. 2022. *Recommendation on the Ethics of Artificial Intelligence*. Paris: United Nations Educational, Scientific and Cultural Organisation.

UNESCO. 2023. *Guidance for Generative AI in Education and Research*. Paris: United Nations Educational, Scientific and Cultural Organization

UNESCO, IRCAI. 2024. *Challenging Systematic Prejudices: an Investigation into Gender Bias in Large Language Models*. Paris: United Nations Educational, Scientific and Cultural Organisation

United Nations.1948. Universal Declaration of Human Rights 217 A (III). Paris.

Véliz, C. 2020. *Privacy Is Power: Why and How You Should Take Back Control of Your Data*. London: Penguin Random House.

Vold, K. and Whittlestone, J. 2019. Privacy, autonomy, and personalised targeting: rethinking how personal data is used. In Véliz, C. (ed). *Data, Privacy and the Individual*, Madrid: Centre for the Governance of Change.

Waghid, Y., Waghid, Z. and Waghid, F. 2019. The Fourth Industrial Revolution reconsidered: non advancing cosmopolitan education. *South African Journal of Higher Education, 36*(3): 1-9.

Wang, T., Zhao, J., Yatskar, M., Chang, K.W. and Ordonez, V. 2019. Balanced datasets are not enough: Estimating and mitigating gender bias in deep image representations. In *Proceedings of the IEEE/CVF international conference on computer vision*, 5310-5319.

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P., Mellor, J.F., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., Biles, C., Brown, S.M., Kenton, Z., Hawkins, W.T., Stepleton, T., Birhane, A., Hendricks, L.A., Rimell, L., Isaac, W.S., Haas, J., Legassick, S., Irving, G. and Gabriel, I. 2022. Taxonomy of risks posed by Language Models. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. New York, NY: Association for Computing Machinery, 214–229.

World Economic Forum. 2016. *The Future of Jobs: Employment, skills and workforce strategy for the fourth industrial revolution.* https://www.weforum.org/publications/the-future-of-jobs/: World Economic Forum.

World Economic Forum. 2023. *Future of Jobs Report.* Coligny: World Economic Forum.

World of Work. 2023. Artificial intelligence & Higher Ed in SA: What Universities Must Consider Now. *World of Work*, 13 February 2023. https://iieworldofwork.iie.ac.za/artifical-intelligence-higher-ed-in-sa-what-universities-must-consider-now.

Xing, B., Marwala, L. and Marwala, T. 2018. Adopt fast, adapt quick: adaptive approaches in the South African context. In Gleason, N. W. (ed) *Higher Education in the Era of the Fourth Industrial Revolution.* Singapore: Palgrave Macmillan.

Zeid, E., 2020. Robot teaching, pedagogy, and policy. In Dubber, M., Pasquale, F. and Das. S. (eds.), *Oxford Handbook of Ethics of AI.* Oxford: Oxford University Press, pp788-803.

Zhao, J., Wang, T., Yatskar, M., Ordonez, V. and Chang, K.W. 2017. Men also like shopping: Reducing gender bias amplification using corpus-level constraints, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark: Association for Computational Linguistics, 2979-2989.

Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.

Making a Case for a Human-Centred Approach to the Adoption and Use of Artificial Intelligence in Higher Education in South Africa

W. Qampi and A. Saidi
Council on Higher Education
Qampi.W@che.ac.za

Abstract

Artificial intelligence represents one of the latest sets of tools developed by human beings to advance the interest of human beings as individuals and as communities and societies. The value that any tool brings depends on how it is used. Ordinarily tools have rules, instructions or protocols for their good use. When those rules, instructions or protocols are ignored, the tools may cause harm to human beings. The same would apply to the use of artificial intelligence tools. Unfortunately, it appears that in many parts of the world, including South Africa, no rules or protocols governing the use of artificial intelligence tools have been developed either at national level or at the level of higher education institutions. The result is that the whimsical adoption and use of artificial intelligence tools carries the risk of inflicting different forms of harm on human beings including cause the loss of employment or source of livelihood; discrimination, bias and exclusion; loss of human interaction and interrelationships; inequality and stereotypes; marginalisation and subjugation, to mention a few. Rules or protocols for the use of tools protect human beings from any possible harm that can emanate from the whimsical use of the tools. This is the basis of a human-centred approach to the design, development and use of tools. This paper makes and presents a case for a human-centred approach to the adoption and use of artificial intelligence in higher education in South Africa. It contends that a human-centred approach would make human beings derive optimum benefits from adopting and utilising artificial intelligence tools, while preventing and/or mitigating possible harm that can result from unregulated use of artificial intelligence tools.

Keywords: Artificial intelligence, ethics, higher education, human-centeredness, human values, *Ubuntu*

Introduction

One of the technological developments that has been thrust to prominence by the Fourth Industrial Revolution is artificial intelligence (AI) whose use is increasingly becoming pervasive in all spheres of life, including in higher education (Lubinga *et al*., 2023). AI is defined as computer-driven "systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals" (European Commission, 2018, p. 1). There are several types of AI, which include Limited Memory AI, Narrow AI, Artificial Generative AI, and Super AI (Joshi, 2019). Taken together, the various types of AI encompass a broad spectrum of applications across many industry sectors including higher education (Compron and Burke, 2023). It is believed that the adoption and use AI in higher education have the potential to enhance teaching and learning, mainly through AI-driven tutoring systems, personalised learning, effective assessment, and effective grading. Similarly it has the potential to increase the efficiency of administrative and student support services, as well as research support services (Rudolph *et al*., 2024).

However, alongside the potential benefits of enhancing teaching and learning, and increasing efficiency of support functions, there are significant ethical and practical concerns about the use of AI in higher education. These concerns include bias, non-compliance with data privacy policies and/or legislation, diminishing of human agency, compromising academic integrity, and exacerbation of existing inequalities and epistemic hegemony, to name a few (Rudolph *et al*., 2024). As a result of these concerns, even the most ardent proponents of the adoption of AI in higher education, including Marwala (2024), urge that caution should be exercised in the process of exploring the possibility of leveraging AI in higher education. This paper explores how the caution urged by proponents such as Marwala (2024) can be exercised in exploring the possible adoption and utilisation of AI tools in higher education in South Africa. It contends that the said caution can be effectively exercised effectively by adopting human-centred approach.

Overview of the use of AI in higher education

The use of AI in higher education has the potential to be transformative in numerous ways (Opesemowo and Adekomaya, 2024) than any other technological development that took place previously (Bates *et al*., 2022). It has the potential to shift the focus from transferring knowledge to processing knowledge, and from disciplinary learning to interdisciplinary learning (Chiu, 2024), among others. Although AI tools were already being utilised for administrative support service provision before the Covid-19 pandemic, since the pandemic, the growth of AI tools has accelerated. Generative AI tools such as ChatPDF, ChatGPT, Copilot, OtterAI, Quilbot, Gradescope and Grammarly are some of the more popular and useful digital applications for higher education.

The pervasiveness of AI in higher education is due to developers having taken much interest in AI for higher education because the sector is highly influenced by the development of information and communication technologies. AI is anchored in the developments of information and communication technologies (Compron and Burke, 2023). Globally, higher education institutions are incorporating AI to enhance learning experiences (Maisiri, 2024); improve administrative efficiency (Sharma *et al*, 2022); and address various educational challenges such as the language barrier to knowledge (Funda and Piderit, 2024), and the enhancement of quality education (Ungerer 24).

The use of AI in higher education in South Africa

The higher education sector in South Africa leads on the African continent regarding the adoption and use of AI (Ade-Ibijola and Okonkwo, 2023). The Covid-pandemic saw many higher education institutions in South Africa reworking their traditional curricula to be provisioned online. Since then, the institutions have not looked back, and are increasingly adopting and using AI in most of the activities, starting from processing of applications for admission, to assessment in programmes (Woldegiorgis, 2024). However, there is an unevenness in the adoption and use of AI among higher education institutions. The globally ranked universities in South Africa use AI more than other higher education institutions, clearly confirming the long existing differences among the institutions in relation access to resources (Lubinga *et al*. 2023).

Although students, academics and researchers in higher education institutions are increasingly adopting and using AI tools, this is happening without regulations at both institutional and national levels. Only a few institutions have begun the process of developing and implementing guidelines and/or policy frameworks to govern the use of AI in teaching and learning, research, and engaged scholarship (Boda, 2024). However, it is critical to have the guidelines and/or policy frameworks in place to avoid risks associated with the use of AI in higher education. Such risks include the exacerbation of existing inequalities, and mismanagement of data, loss of employment owing to the automation of jobs, and the creation of jobs that clearly favour high level skills (Alexander, 2022), notwithstanding the fact that the majority of the population do not possess such high level skills. The guidelines and/or policy frameworks should not only explain to AI users the opportunities that stem from incorporating AI into teaching, learning and research, but also highlight the basics of ethical implications of AI in higher education.

Human centredness

As demonstrated by Mashao and Aigbavboa (2024), AI is simply one of the set of tools that the human being has developed to serve the interests and needs of human society. It arguably represents the latest of sets of tools that the human being has developed on the evolutionary trajectory. Therefore, AI as a set of tools in service of humanity, should be designed, adopted and utilised in ways that harness the value for serving the interests and needs of human beings, while limiting any negative effects on human society and on the environment on which human society depends for its sustenance. The concept of human centredness in the design, adoption and utilisation of technology means that the technology in question should enhance and not replace the natural ability and capability of people. With respect to AI, it means that it should augment the innate abilities and capabilities of people who work in various sectors without necessarily replacing them and causing them to lose their sources of livelihoods (Raebun, 2023). A human-centred approach to the adoption and utilisation of AI focuses on leveraging AI tools to bolster human decision-making, enrich human interactions, and increase efficiency. It is about placing the interest of the human being, as the end user, at the centre of designing AI tools, making them user-friendly, and enhancing the outcomes of their use (Surireddy, 2023). This approach results in empowering employees with the tools they need to take charge of their jobs and improve productivity (Chan, 2023). In higher education, human centredness with respect to AI means adopting and using AI to make teaching, learning, research and engaged scholarship less onerous; to make students learn and acquire the necessary knowledge and skills without enormous effort; and to make researchers successfully undertake complex studies without requiring massive financial and human capital. Human centred adoption and use of AI in higher education does not alter or compromise the universal fundamental value and practices of higher education. It only makes these to be pursued in a more resource effective and time effective manner, thereby making the return on investment greater.

A human-centred approach to technology, including AI, starts with undertaking studies to understand the dynamics of the ever-changing world and local environments, and how the human being adapts to such changes. Another important dimension is to engage extensively with internal and external stakeholders, and use the outcomes of such engagements as the basis for decisions to adopt and use particular technology or AI tools. Once the technology is acquired, there should be effective communication to the stakeholders to provide sufficient information that will allow the end-users to make effective use of the technology. Such information may also be packaged into guidelines and or policy frameworks (Chan, 2023). This means, for example, that the adoption and use of AI in higher education should be guided and be regulated by higher education leaders at both institutional as well as system levels (Lubinga *et al*., 2023). A human-centred approach also required continuous monitoring and evaluation so that if there are unforeseen imponderables, they can be detected and attended to in time before the cause irreparable harm.

Centrality of context

Although the human-centred approach to the adoption and use of AI tools has universal goals and outcomes as discussed in the preceding section, it is important to note the practical aspects may vary from one context to another. This is because the culture of a society determines what is acceptable and not acceptable in a particular society. Therefore the culture of a society is a crucial factor to be considered in the adoption and use of technical breakthroughs including AI (Funda and Piderit, 2024). There is no one 'prototype' of human-centred approach that fits and works effectively in all societies in a 'one size fits all' fashion. The implication of this is that frameworks developed by international bodies such as United Nations Scientific, Educational and Cultural Organisation (UNESCO) should be used only as reference material in the process of developing context-specific institutional and/or national guidelines and policy framework to govern human-centred adoption and use of AI in higher education. They should not be copied *as the* 'blue print' to be transplanted in different institutions and/or different countries or societies without alteration, because culture varies from one society to another, and form one institution to another.

It is also critical to understand that the guidelines and policy frameworks of multilateral bodies such as UNESCO were developed based on data that is predominantly from the West (Goffi, 2023) because the bulk AI tools available were designed, developed and created in North America and Europe (Kazim and Koshiyama, 2021). The rules, instructions and protocols for the proper use of the AI tools are based on the concerns from the West (Goffi, 2023). Although these universal guidelines can potentially provide a skeleton framework to guide regions on ethical concerns of AI and how they are approached, they are unable to fully recognise the importance of diverse voices in tackling the risks and benefits of AI (Mokoena, 2024). The aforementioned examples of the universal guidelines on proper use of AI developed by the UNESCO and the World Health Organisation (WHO) are examples of guidelines and/or policy framework that are heavily influenced by the Global North. Africans are quite sceptical when it comes to the adoption and utilisation of new technologies such as AI (Ade-Ibijola and Okonkwo, 2023).

Although the frameworks developed by these multilateral bodies are heavily influenced by issues in the West, they emphasise the principles that have universal applicability. These are respect of human rights, respect for society and the environment, robustness and safety, transparency, contestability, responsibility or accountability, justice and fairness, and privacy. These principles underpin the human-centre approach because they are essential for addressing societal challenges. However, they need to be expounded in a manner that relates to any particular social and cultural context (Kwoa *et al*., 2023). These principles are as important in Africa as they are in the West. It is therefore important when developing guidelines and policy frameworks to govern the adoption and use of AI, higher education institution and countries in Africa and elsewhere in the Global South, build on these universal principles and values.

Flowing from the above, a human-centred approach to the adoption and use of AI tools in higher education in South Africa should include, among others, the development and implementation of guidelines and/or frameworks that are underpinned by South African human values. The constitution of South Africa is founded on values such as human dignity, the realisation of equality and the advancement of human rights and freedoms, non-racialism and non-sexism (Republic of South Africa, 1996). Another important South African human value is *Ubuntu* which has roots in Southern African cultures. The word *Ubuntu* is derived from the Nguni languages, and it is defined as an African humanist philosophy that is “characterised by interconnectedness of all things and beings; the spiritual nature of people; their collective/individual identity and the collective/inclusive nature of family structure; oneness of mind, body, and spirit; and the value of interpersonal relationships” (Mungai, 2015 as cited in Zvomuya, 2020). In South Africa, *Ubuntu* is explained through the common Nguni aphorism: *“Umuntu ngumuntu ngabantu*” which is loosely translated as “a person is a person through others” or “I am because you are”). *Ubuntu* values include compassion, solidarity, harmony, consensus, hospitality, sympathy (Chigangaidze *et al*., 2022); respect, dignity, equity and interdependence (Ngubane and Makua, 2022); the promotion of teamwork and collaboration, cohesion and belonging (Mupedziswa *et al*., 2019. Other values of *Ubuntu* are humaneness, communalism, harmony, and responsiveness.

Ubuntu is the foundation for the basic values that manifest themselves in the ways African people think and behave toward each other, and towards everyone else they encounter (Mangaliso, 2001). Although the *Ubuntu* philosophy and the aforementioned aphorism are in Nguni, it is important to note that the philosophy exists in other regions outside of South Africa, and is known by different names such as *Bomoto* in Congo; *Gimuntu* in Angola; *Botho* in Botswana; *Umunthu* in Malawi; *Vumuntu* in Mozambique; *Umunhum, Vhutu; Humunnhu/Ubuthosi* in Zimbabwe; *Bumuntu* in Tanzania; and *Umuntu* in Uganda, just to name a few (Mupedziswa *et al*., 2019).

Impacts of not foregrounding human-centeredness in utilisation of AI tools

This sections discusses the utilisation of AI tools in higher education without foregrounding human-centredness. The impacts of this are discussed to make a case for a human-centred approach to the adoption and use of AI tools in higher education. The discussion focuses on AI tools used for tutoring, performance of administrative functions, expanding access to higher education, language learning, assessment and feedback from assessment, and research.

AI for tutoring services

One of the areas in higher education which has found favour with the use of AI is the provision of tutoring services. AI tools such as Intelligent Tutoring Systems (ITS) are used in this respect to mimic tutoring environments such as the traditional, face-to-face tutoring (Alam and Mohanty, 2023). 'Dr Maths' is an example of an ITS developed by the Council for Scientific and Industrial Research (CSIR) in South Africa. It provides South African students who are studying mathematics with tutoring support in the form of personalised real-time assistance from humans, supported by automated language clarification and topic identification (Lubinga *et al*., 2023). Tutoring systems such as 'Dr Maths' do not only support students in providing personalised learning experiences, but also contribute significantly to reducing the workload of lecturers. Students tend to feel more comfortable interacting with robots for learning purposes, and these virtual assistants or tutors, are able to think, act, interact and supply customised content and personal care (Ungerer, 2019). These AI tutors may provide high-quality, one-on-one tutoring which is always available to all students. In essence, they transform classroom learning (Tarisayi, 2024, p. 31). They also provide personalised feedback and support to students, as well as simulate traditional tutoring learning experiences.

Unfortunately, human-centredness is not foreground in the design and utilisation to the AI tutoring tools. For example, human interaction is the essential basis for teaching and learning in higher education. However, the use of ITS and related AI tools eliminates human interaction completely. Similarly, the *Ubuntu* value of interdependence does not have a place in AI-facilitated tutoring environment (Ngubane and Makua, 2022). All this has the potential to negatively impact the relationship between students and educators or lecturers, and to reduce personal interaction and mentorship, which constitute one of the critical features of education (Ausat *et al*., 2023; Sallam *et al*., 2023; Qadir, 2023, cited in Al-Mughairi and Bhaskar, 2024). The isolation denies students opportunities for personal connections, teamwork and collaborations. Of course isolation enhances the value of autonomy, which, to all intents and purposes, is Western-European value that conflicts with the value of *Ubuntu,* which advocates for a form of wholeness that comes through one's relationship and connectedness with other people in the society (Akpa-Inyang and Chima, 2021). The absence of personal interaction between the educator or tutor and students militates against participatory and interactive learning and yet a fundamental aspect of any good form of pedagogy is to optimise the participation of students in their learning processes. Isolation makes active learning impossible, and yet it is active learning that advances engagement with learning material as personal interactions provides students with the chance to discuss issues, ask questions, debate and share their thoughts and experiences (Ngubane and Makua, 2022).

AI for provisioning administrative support services

Higher education institutions have adopted AI tools for use in provisioning administrative support services. Such AI tools include chatbots which perform routine tasks such as organising meetings, communication, managing schedules and documenting student enrolment and providing support for staff members (Adeshola and Adepoju, 2023, as cited in Rudolph *et al*., 2024). As previously alluded to, AI can provide student support through virtual tutoring, and lecturers benefit from having virtual tutors who are available at any time to attend to their needs. Educators or lecturers can utilise AI tools as their teaching assistants to perform administrative tasks. Doing so leaves them with more time to focus on fostering critical and problem-solving skills, and even to focus on mentoring roles for students (Maisiri, 2024).

The increased efficiencies that result from the use of AI tools in provisioning administrative support services in higher education institutions may result in eliminating the need to fill administrative support positions within the institutions. In worst-case scenarios, it may result in cutting down on administrative support positions and thereby leading to retrenchments. In either case, human beings are negatively affected by adopting the AI tools in provisioning administrative support services, without prioritising human interests and needs. Those who are unable to find employment, and those who are retrenched because higher education institutions are increasingly adopting AI tools in the provisioning of administrative support services have their human dignity impaired. According to the constitution, one of South Africa's founding values is human dignity (Republic of South Africa, 1996). Similarly, the *Ubuntu* philosophy also endorses dignity in society (Ngubane and Makua, 2022).

Furthermore, by making people with administrative skills and competencies unable to find employment, and by precipitating retrenchment of administrative staff, the adoption of AI in provisioning administrative support services, would perpetuate social inequality (Funda and Piderit, 2024). In general, the adoption and use of AI tools can exacerbate inequalities in the labour market, unless it is undertaken in a human-centred fashion. In South Africa, for instance, the bulk of the country's workforce is unskilled (Marwala, 2021) and yet the adoption and utilisation of AI tool threaten to make unskilled and/or semiskilled jobs redundant (Mokoena, 2024) which would then worsen the already existing unemployment crisis in the country, resulting in stripping people of their human dignity.

A human-centred approach to the adoption of AI tools would require that the capabilities of the various AI tools available is assessed, and a decision is made to adopt and use only those tools that should augment the innate abilities and capabilities of people who work in in the administrative support functions in higher education institutions, without necessarily replacing them and causing them to lose their sources of livelihoods (Raebun, 2023).

AI for expanding access to higher education

AI technologies are being utilised by higher education institutions to expand access to higher education, particularly to the marginalised or disadvantaged communities (Rudolph *et al*. 2024). Literature suggests that AI tools present novels ways of providing high-quality education globally, especially to those who ordinarily would not have access to it (Ungerer, 2019). Countries with large proportions of people with access to the Internet, as well as access to devices such as smart mobile phones and/or laptops, can expand access to higher education significantly by using AI tools (Maisiri, 2024). In such countries, access to higher education is increasingly being extended to the majority of their peoples.

However, the digital divide remains a challenge. Not all people or classes of people in any country have equal access to the Internet, or access to devices such as smart mobile phones and/or laptops. The digital divide separates those that have access to the Internet and devices such as smart phones and laptops, from those who do not. In general, the digital divide replicates the urban-rural divide, with majority of urban population having access to digital resources, and the majority of the rural population not having readily access to digital resources. Therefore, despite the potential of expanding access to higher education, the digital divide means that the access to higher education cannot be extended to all and sundry. Digital divide is a manifestation of inequality in society. Therefore, unless the digital divide is eliminated, the adoption and use of AI tools can only serve to entrench inequality which is the root cause of many human rights violations.

In South Africa, the digital divide is compounded by challenges of loadshedding which makes electricity not available for extended periods in a day (Mutiso, 2024). Loadshedding has now morphed into load reduction, but the effect is the same. Electricity is to AI like what a heartbeat is to the human body. Without a constant and stable electricity supply, the benefits of adopting and using AI tools will never be fully realised. Unfortunately, even with back-up power, the rolling blackouts still affect access to online (digital) education because the mobile phone towers are also powered by electricity. When their battery power is depleted during extend periods of loadshedding and/or load reduction, they cease to function, and thus impairing connectivity to the Internet (Szutowicz, 2023). Network operators such as Vodacom have to prioritise spending billions of rands in ensuring that they continue to provide connectivity when there is no power instead of investing that money into rolling out infrastructure and new technologies in rural areas (Szutowicz, 2023). Once against, this points out to the fact that unless and until the basic needs of human being, such as uninterrupted access to electricity, are attended to first, the full potential of AI cannot be realised and be leveraged on.

As discussed above, the full potential of using AI to expand access to higher education is compromised by the digital divide and other factors that cause or accentuate inequality. The adoption of AI in higher education needs to be undertaken in a way that considers the human value of equality. The adoption and utilisation AI tools in higher education without strategies that consider achieving equality will continue to widen the knowledge gap among the ‘haves’ and ‘have-nots’, as well as between the Global North and Global South to avoid accentuating existing inequalities (Rudolph *et al.*, 2024).

AI for learning languages

AI tools facilitate the learning of languages by providing translation, writing tutorials, checking spellings and grammar. They assist in academic writing through auto spell and grammar checking, and auto compilation of references, among others. Interactive software applications, online lessons and platforms, automated translators and multimedia resources augment traditional approaches, offering dynamic and engaging learning experiences. These technological advancements have improved access to language learning resources and support, allowing students to practice the language being learnt anytime, anywhere, and, maybe even more importantly, in a fun way (Zucchet, 2023).

The challenge for South Africa and other countries in the Global South is that the languages that can be learnt using AI tools are mostly English, French, Germany, Spanish, Russian and Mandarin. These also happen to be the languages of teaching and learning, and scholarship across the world. It is undeniable that these are not only foreign languages to South Africa and to other countries the Global South, they are also languages of imperialism and colonial hegemony (Makeleni *et al*., 2023). It therefore appears that AI tools are entrenching the languages of imperial and colonial hegemony at the expense of local indigenous languages. As many scholars have argued, language is not simply a medium of communication, it is rather a medium of tradition and culture, and one of factors that define the humanness of human beings. When the language of a particular community of people is subjugated, the culture, the traditions and the humanness of those people are also subjugated and marginalised. It is important that this issues is considered carefully before wholesale adoption of AI tools to promote and facilitate the learning of languages in higher education.

Experts warn that the Large Language Models (LLMs) used in AI are, in the main, not based on indigenous African languages has the effect of the excluding millions of Africans from benefitting from those AI tool, considering that only a small proportion of the population of Africa can communicate in English, French and other imperial languages. When the masses are excluded in this way, the inequality becomes stark (Adebayo *et al*., 2024). Since the LLMs are mostly based on English and the other colonial languages, the use AI in higher education in South Africa, for example, has the potential to undo the little gains of language transformation made since 1994. It makes the use of African languages in teaching and learning, and in scholarship, less attractive. Few postgraduate students would still seek to write their theses in African languages when they could do so less onerously in English because of the assistance of AI tools. Similarly, few academics and researchers would seek to prepare manuscripts for publication in African languages when they could prepare them effortlessly in English with the assistance of AI tools. Petersen (2023) argues that multilingualism assists in removing barriers in teaching and learning and facilitates better communication, understanding and comprehension of knowledge. This inclusion of African languages will also encourage social cohesion, diversity in academic spaces and create a positive attitude and pride in their native languages (Petersen, 2023). AI in higher education will not be able to achieve all of this if it cannot include the indigenous languages. Students who are not proficient in the dominant language and cannot articulate themselves in the dominant language will struggle to participate in discussion forums and tutorials, thus potentially causing AI tools to generate an evaluation of the students' participation and knowledge application inaccurately. The ability to speak and understand English or any of the European languages is not an indicator of effectiveness of learning in most disciplines.

A human-centred approach to the use of AI tools in learning languages would involve developing AI tools for Africa indigenous languages has slowly. Unfortunately, at this stage, there have not been significant developments in this regard (Makeleni *et al*., 2023).

AI for assessments and provision of feedback to students

AI tools are also used in higher education to provide increased efficiency and effectiveness in activities such as review and evaluation of learning, formal assessments, grading, as well as providing feedback (Chen *et al*., 2020). AI tools such as chatbots and virtual assistants can augment the role of educators or lecturers in providing feedback to students, review and evaluation using huge quantity of datasets, and automating routine assessment tasks such as grading (Vashishth, 2024, cited in Rudolph *et al*., 2024). AI can also prepare exercises and assessments within the curriculum that are customised to students' particular weaknesses, and such exercises and assessments are used to gauge how well students have grasped the knowledge that they have engaged with (Maisiri, 2024).

The use of AI tools in assessments presents a unique challenge as automating the decision making in assessments can reinforce historic stereotypes, discrimination and exclusion (Mokoena, 2024). This reinforcement of historic stereotypes, discrimination and exclusion through advanced mechanisms is a form of digitised knowledge hegemony. It has the potential to negatively inform the content, assessment and grading methods that these AI tools provide especially towards African contexts. Adopting these AI tools, which are inherently biased against people of particular races, language groups, and from particular countries, without moderating the biasness, is indeed repugnant, and a disservice to the higher education institutions that implement these AI tools.

The AI tools are programmed by feeding data to their algorithms and training the algorithms to generate results efficiently and accurately. The data that these algorithms are fed is determined by its developers, and if there are no inputs from developers from other regions (like the Global South) to provide diverse perspectives from their worldviews, the challenges of bias, discrimination and exclusion, result. Bias from AI technologies poses significant ethical challenges in higher education, as these systems have been trained to inherit and amplify the prejudices present in the data that they were trained with. The prejudices embedded in these technologies include racism and sexism (Rudolph *et al*., 2024). A human-centred approach would require that data from South Africa and other countries in the Global South should be included in the sets of data that these algorithms are fed and trained on.

AI for research

In higher education AI tools have the potential of providing automated assistance for research activities. The automation of research activities did not start with AI. Software packages such as Statistica and Atlas.ti have assisted researchers to analyse data efficiently to inform research findings. Interactive AI applications such as ChatGPT can assist researchers and even research students with generating research questions, designing methodologies, generating summaries of literature, and providing overall research support (Al-Mughairi and Bhaskar, 2024). These applications can also assist in drafting research reports and manuscripts for publication (Rudolph *et al*., 2024). Applications such as *Trinka,* which checks that research paper for submission to journal conform to style guides of the journals; and *Scholarcy,* which provides summaries of literature and provides important points from the literature, are frequently used in research. They provide advanced and automated research assistance in an efficient and effective way (Rudolph *et al*., 2024).

The main concerns regarding the use of AI tools in research relate to authenticity and accuracy. Crawford *et al*., (2023c) assert that utilising AI in research has limitations which could lead to subpar quality, inaccuracies, biases, and a lack of originality. Insufficient or inaccurate data fed into AI during its development leads to biased decision-making by the AI tools, which is a risk to research especially from the African context (Damfeh *et al*., 2022). AI may not be able to interpret and accurately summarise literature that speaks to topics outside of Western countries. The potential bias and discrimination from these algorithms can misinform research; which compromises the integrity of research.

Towards a human-centred approach to the adoption and use of AI

At the national level, a major part of a human-centred approach to the adoption and use of AI is having in place guidelines and/or policy frameworks that articulate the rules, instructions and protocols that govern the use of the AI tools. Unfortunately, in South Africa, there has been little guidance from the government on how AI should be utilised; and higher education institutions have been slow to propose solid strategies and policies on the use of AI in education (Funda and Piderit, 2024). Therefore, unlike other regions such as China's, the European Union, Australia and the United States of America's, the use of AI in South Africa remains largely a *laissez faire* affair (Boda *et al*., 2024).

The human-centred approach requires that the guidelines and/or policy frameworks that are developed and implemented to govern the use of AI should be developed informed by results of feasibility and/or impact studies, and the outcomes of extensive stakeholder engagements. The feasibility and/or impact studies should focus on identifying possible negative effects on human beings, and how such possible effects could be prevented and/or mitigated. Similarly, the extensive stakeholder engagements should focus on how human values can be considered in making the choice of the AI tools to adopt and use, and how to make use of them. AI tools that have the potential of eliminating any need of human intervention in the performance of tasks to be avoided or adapted to minimise the potential impact. As discussed earlier, the use AI should not lead to many people losing their sources of livelihoods. It should rather contribute to making human beings more effective and efficient in the performance of their job activities, in making decisions, and in overcoming challenges.

The feasibility studies and stakeholder engagements should pay particular attention to the issue of digital divide and other forms of inequality that are manifested or accentuated through the adoption and use of AI tools. Technological innovations that generate or exacerbate inequalities should not be adopted and used unless accompanied with interventions that seek to reduce the inequalities and/or its ramifications. AI technologies should not be available only to the elite in society.

Within higher education institutions, a human-centred approach also requires that students, academics and other staff members are assisted with acquiring the necessary knowledge of the AI tools available, and acquire the understanding of how they can be leveraged to facilitate learning, teaching, and support research. It is hard to imagine an institution being adequately prepared for AI tools if there is almost no training at all on digitalisation and its use in teaching and learning, research, and engaged scholarship (Woldegiorgis, 2024). Training educators on ethical use of AI could assist in addressing many of the concerns about AI (Tarisayi, 2024). Therefore, it is essential to have ongoing training and support for educators and students on the use and ethics of AI. Institutions have a duty to focus on capacity development, ensuring that educators and students acquire adequate skills for the responsible use of AI (Ade-Ibijola and Okonkwo, 2023).

Conclusion

This paper has argued for a human-centred approach to the adoption and utilisation of AI tools in higher education. Its point of departure is that AI is a set of tools aimed at making human being more effective and efficient. As such, AI should seek to advance the interests of human beings, and assist human beings in their quest to realise their aspirations. Ideally, therefore, the adoption and use of AI should not be harmful to human beings. Any harm that a human being suffers as a result of adopting and utilising AI is a red flag that calls for the use of the AI to be reassessed.

A human-centred approach to the adoption and use of technology means placing human interests and needs at the centre. It means putting in place mechanism for optimising the benefits that human beings can derive from AI while minimising or eliminating any possible harm. The harm that a human being can suffer from the adoption and use of AI is often not physical. It comes in the forms of losing employment or sources of livelihood; suffering discrimination, bias and exclusion; losing human interaction and interrelationships; suffering inequality and stereotypes; and suffering neo-colonialism in its many different dimensions and variants, to mention a few.

The human-centred approach requires that the guidelines and/or policy frameworks to govern the use of AI should be developed, informed by results of feasibility and/or impact studies, and the outcomes of extensive stakeholder engagements. It also requires continuous monitoring and evaluation, so that unforeseen impacts are detected in the nick of time, and attended to, accordingly.

References

Adebayo, B., Bhalla, N. and Harrisberg, K. 2024. From Swahili to Zulu, African Techies develop AI language tools. CNBC Africa, 18 June 2024. https://www.cnbcafrica.com/2024/from-swahili-to-zulu-african-techies-develop-ai-language-tool.

Ade-Ibijola, A. and Okonkwo, C. 2023. Artificial Intelligence in Africa: Emerging challenges. *Responsible AI in Africa, Social and Cultural Studies of Robots and AI*. https://doi.org/10.1007/978-3-031-08215-3_5. Date accessed: 12 June 2024.

Akpa-Inyang, F. and Chima, S. C. 2021. South African traditional values and beliefs regarding informed consent and limitations of the principle of respect for autonomy in African communities: a cross-cultural qualitative study. *BMC Medical Ethics.* https://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-021-00678-4.

Alam, A. and Mohanty, A. 2023. Foundation for the future of higher education or 'misplaced Optimism'? Being human in the age of artificial intelligence. *Innovations in Intelligent Computing and Communication*. https://link.springer.com/chapter/10.1007/978-3-031-23233-6_2.

Alexander, R. 2022. Key Opportunities and Challenges for 4IR in South Africa. *DSI/NRF South African Research Chair in Industrial Development.* https://www.uj.ac.za/wp-content/uploads/2021/10/sarchi-wp-2021-08d-alexander-october-2022.pdf.

Al-Mughairi, H. and Bhaskar, P. 2024. Exploring the factors affecting the adoption of AI techniques in higher education: Insights from teachers' perspectives on ChatGPT. *Journal of Research in Innovative Teaching and Learning*. https://www.emerald.com/insight/content/doi/10.1108/JRIT-09-2023-0129/full/html.

Bates, T., Cobo, C., Mariño, O. and Wheeler, S. 2020. Can artificial intelligence transform higher education? *International Journal of Educational Technology in Higher Education*. https://doi.org/10.1186/s41239-020-00218-x.

Boda, R., Gunning, E. and Ntuli, L. 2024. The EU AI Act Passes: Should South Africa follow suit and regulate Artificial Intelligence? *ENSight,* 19 March 2024. https://www.ensafrica.com/news/detail/8261/the-eu-ai-act-passes-should-south-africa-foll#:~:text=While%20AI%20remains%20largely%20unregulated,unlawful%20processing%20of%20personal%20information.

Chan, K. A. 2023. Digital transformation: 4 tips to keep it human-centred. *The Enterprise Project*. https://enterprisersproject.com/article/2023/1/digital-transformation-4-tips-keep-it-human-centered.

Chen, L., Chen, P. and Lin, Z. 2020. Artificial intelligence in education: A review. *IEEE Access* 8, 75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510.

Chigangaidze, R. K., Matanga, A, A. and Katsuro, T. R. 2022. *Ubuntu* philosophy as a humanistic–existential framework for the fight against the Covid-19 pandemic. *Journal of Humanistic* Psychology 62(3): 319-333.

Compron, H. and Burke, D. 2023. Artificial intelligence in higher education: the state of the field *International Journal of Educational Technology in Higher Education* 20(22). https://link.springer.com/article/10.1186/s41239-023-00392-8.

Damfeh, E. A., Wayori, B. A., Appiahene, P., Mensah, J. and Awarayi, N. S. 2022. Human-centred artificial intelligence: a review. *International Journal of Advancements in Technology* 13(8) ttps://www.longdom.org/open-access/humancentered-artificial-intelligence-a-review-94388.htm

Goffi, E. R. 2023. Teaching ethics applied to AI from a cultural standpoint: What African "AI Ethics" for Africa? in Corrigan, C. C., Asakipaam, S. A., Kponyo, J. J. and Luetge. C. (eds) *AI Ethics in Higher Education: Insights from Africa and Beyond*. Springer Briefs in Ethics. Pp13 – 26. Available at: https://doi.org/10.1007/978-3-031-23035-6_2

European Commission. 2018. *A Definition of AI: Main Capabilities and Scientific Disciplines*. European Commission. https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf.

Funda, V. and Piderit, R. 2024. A review of the application of artificial intelligence in South African higher education. Proceedings of the 2024 Conference on Information Communications Technology and Society. https://ieeexplore.ieee.org/abstract/document/10507113.

Josh, N. 2019. 7 types of artificial intelligence. *Connective World.* https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence/.

Kazim, E. and Koshiyama, A. S. 2021. A high-level overview of AI ethics. *Patterns.* https://doi.org/10.1016/j.patter.2021.100314.

Kwoa, P. G., Kanubala, D. D. and Sonna, B. 2023. AI ethics education for future African leaders. *AI Ethics in Higher Education: Insights from Africa and Beyond*, Springer Briefs in Ethics. https://doi.org/10.1007/978-3-031-23035-6_2.

Lubinga, S., Maramura, T. C.and Masiya, T. 2023. Adoption of Fourth Industrial Revolution: Challenges in South African higher education institutions. *Journal of Culture and Values in Education* 6(2). https://repository.up.ac.za/handle/2263/94027?show=full.

Maisiri, J. 2024. AI as a beacon of hope for education reform. *Mail & Guardian,* 9 April 2024. https://mg.co.za/thought-leader/opinion/2024-04-09-ai-as-a-beacon-of-hope-for-education-reform/.

Makeleni, S., Mutongoza, B. H. and Linake, M. A. 2023. Language education and artificial intelligence: an exploration of challenges confronting academics in Global South universities. *Journal of Culture and Values in Education* 6(2):158-171.

Mashao, D and Aigbavboa, C. 2024. Artificial Intelligence as the Latest Tool of *Homo habilis*…. Will it be the Last Tool? *Kagisano* 15. (this issue). Pretoria: CHE

Marwala, T. 2021. The Fourth Industrial Revolution in Higher Education. In Brink. C. (ed). *The Responsive University and the Crisis in South Africa*. Leiden: Brill. Pp 300 – 311.

Marwala, T. 2024. Generative AI and education. *Medium,* 9 April 2024. https://medium.com/@tshilidzimarwala/generative-ai-and-education-1e203acdf83f.

Mokoena, K. K. 2024. A holistic *ubuntu* artificial intelligence ethics approach in South Africa. *Verbum et Ecclesia* 45(1):a3100. https://verbumetecclesia.org.za/index.php/ve/article/view/3100/7364.

Mupedziswa R., Rankopo M. and Mwansa, L. 2019. *Ubuntu* as a Pan-African philosophical framework for social work in Africa. In Twikirize J. M. and Spitzer H. (eds.) *Social Work Practice in Africa: Indigenous and Innovative Approaches*. Kampala: Fountain Publishers. pp. 21-38

Mutiso, R. M. 2024. AI in Africa: Basics over buzz. *Science* 383(669). https://www.science.org/doi/10.1126/science.ado8276.

Ngubane, N. I. and Makua, M. 2022. Intersection of *Ubuntu* pedagogy and social justice: Transforming South African higher education. *Transformation in Higher Education* 6(0): a113. doi.org/10.4102/the.v6i0.113.

Opesemowo, O. A. G. and Adekomaya, V. 2024. Harnessing artificial intelligence for advancing sustainable development goals in South Africa's higher education system: A qualitative study. *International Journal of Learning, Teaching and Educational Research* 23(3): 67 – 86.

Petersen, F. 2023. Why South African universities have a duty to develop multilingualism. https://www.ufs.ac.za/templates/news-archive/campus-news/2023/september/why-south-african-universities-have-a-duty-to-develop-multilingualism.

Popenici, S. A. D. and Kerr, S. 2017. Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(22). https://telrp.springeropen.com/articles/10.1186/s41039-017-0062-8.

Raebun, A. 2023. What is a human-centric AI? A critical, empathetic approach to technology. *The Work Innovation Lab*. https://asana.com/resources/what-is-human-centric-ai

Republic of South Africa. 1996. *The Constitution of the Republic of South Africa*. https://www.justice.gov.za/constitution/SAConstitution-web-eng.pdf.

Rudolph, J., Ismail, M. F. B. M. and Popenici, S. 2024. Higher education's generative artificial intelligence paradox: The meaning of chatbot mania. *Journal of University Teaching and Learning Practice* 21(6).

Surireddy, R. 2023. Human-centred technology: developing useful and user-friendly solutions to increase access to public benefits. *Steady IQ*. https://www.steadyiq.com/resources/human-centered-technology .

Szutowicz, T. 2023. The terrible toll of load shedding on SA's Mobile Networks.,*Tech Central,* https://techcentral.co.za/the-terrible-toll-of-load-shedding-on-sas-mobile-networks/219424/.

Tarisayi, K. 2024. Preparing for AI's transformational potential: rethinking teacher education in South Africa. *International Education Trend Issues.* 2(1).

UNESCO. 2023. *Guidance for Generative AI in Education and Research*. UNESDOC Digital Library. https://unesdoc.unesco.org/ark:/48223/pf0000386693.

Ungerer, L. M. 2019. AI in higher education: Considering the ecosystem in an emerging country context in emerging education futures. In Moravec, J. W. (ed.) *Education Futures*: *Experiences and Visions from the Field*. Minneapolis: Educational Futures.

Woldegiorgis, E. T. 2024. The challenges of online learning in African higher education: A critical reflection on the digital divide associated with the Covid-19 pandemic. Woldegiorgis, E. T. and P. Jonck. (eds) *Higher Education in the Face of a Global Pandemic.* Leiden: Brill.

Zvomuya, W. 2019. *Ubuntuism* as an international turning point for social work profession: new lenses from the African pot of knowledge. *African Journal of Social Work* 10(1).

Zucchet, E. 2023. How artificial intelligence is revolutionising language learning. https://www.berlitz.com/blog/artificial-intelligence-ai-language-learning.

The Kagisano Series: 2001 to 2024

Number 15 (2024):	Artificial Intelligence and Higher Education in South Africa
Number 14 (2022):	Inferences of Resilience and Agility of Higher Education from the Diverse Experiences with Emergency Remote Teaching and Learning
Number 13 (2022):	The Fourth Industrial Revolution and Higher Education in South Africa
Number 12 (2020):	Student Governance and Engagement for Academic Success
Number 11 (2017):	Constituting Higher Education
Number 10 (2016):	Student Funding
Number 09 (2013):	The Aims of Higher Education
Number 08 (2013):	Academic Freedom
Number 07 (2010):	Universities of Technology: Deepening the Debate
Number 06 (2010):	Community Engagement in South African Higher Education
Number 05 (2006):	Universities of Technology
Number 04 (2006):	Ten Years of Higher Education under Democracy
Number 03 (2003):	The General Agreement on Trade in Services and South African Higher Education: What Should South Africa do?
Number 02 (2003):	Principles for Governance of South African Higher Education: Reflections on Cooperative Governance in South African Higher Education
Number 01 (2001):	Re-inserting the 'Public Good' into Higher Education Transformation

Council on Higher Education

1 Quintin Brand Street

Persequor Technopark

Pretoria

South Africa

Tel: +27 12 349 3887/3840

www.che.ac.za

Research, Monitoring and Advice Directorate