Editors
Yuriy B. MELNYK
Mmatshuene A. SEGOOA

Artificial Intelligence in Digital Society

Collective Monograph
Volume 1

AI

KRPOCH

KRPOCH Scientific Research Institute

Editors
Yuriy B. Melnyk
Mmatshuene A. Segooa

Artificial Intelligence in Digital Society

Collective Monograph
Volume 1

Kharkiv
“KRPOCH”
2026

DOI: 10.26697/9786177089192.2026
UDC: 004.8:308:008(058)
BISAC: TEC052000
Ar791

Recommended
by the Academic Council of the KRPOCH Scientific Research Institute
by the Academic Methodological Council of the KRPOCH Scientific Research Institute
(Protocol No. 1 of 10.08.2025)

Editors:
Melnyk Y. B.
Doctor of Philosophy in Pedagogy (PhD), MPSI, MIM, Affiliated Associate Professor,
Director, KRPOCH Scientific Research Institute, Kharkiv, Ukraine
Segooa, M. A.
Doctor of Computing in Informatics (PhD),
Senior Lecturer, Tshwane University of Technology, Pretoria, South Africa

Ar791 Melnyk, Y. B., & Segooa, M. A. (Eds.). (2026). *Artificial Intelligence in Digital Society, Vol. 1*. KRPOCH. https://doi.org/10.26697/9786177089192.2026

The series of monographs "Artificial Intelligence in Digital Society" aims to bring together the existing theories and practices of artificial intelligence in the Western and Eastern worlds. The idea of synergy plays a central role in the monographs. The monographs in this series represent an attempt to integrate the experience accumulated in recent years and propose new solutions to the problems of using artificial intelligence. This series of monographs is an important resource for researchers, professors and graduate students, as well as practitioners and specialists in the field of artificial intelligence.

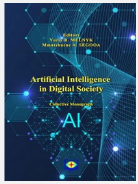
Open Access. The edition is available in international databases and repositories: Crossref, Google Scholar, EndNote Click, Internet Archive, eKRPOCH, etc.

ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)

CONTENTS

Artificial Intelligence in Digital Society,
Volume 1, 2026

DOI: 10.26697/9786177089192.2026
ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)



Introduction to the Monograph on Artificial Intelligence in Digital Society, Volume 1, 2026

Melnyk Y. B. [1,2]

[1] Kharkiv Regional Public Organization "Culture of Health", Ukraine
[2] Scientific Research Institute KRPOCH, Ukraine

Received: *15.12.2025;* ***Accepted:*** *10.02.2026;* ***Published:*** *10.03.2026*

We are on the verge of global changes in existence. In the digital age, we have witnessed the emergence of objects and phenomena that were unimaginable just a few decades ago.

The creation and development of artificial intelligence (AI) attracts particular attention from scientists and the general public. Artificial intelligence, which is created by human intelligence, will inevitably surpass its creators in the near future. This situation has brought a number of fundamental questions (understanding of the nature of reality, consciousness, the meaning of existence, etc.) to the fore at a new level, which require study.

Given these circumstances, I came up with the idea of creating a series of monographs entitled ***Artificial Intelligence in Digital Society***, which describe existing AI theories and practices and explore the complex relationship between AI and society.

The monograph is based on the idea of synergy, the increase in the overall efficiency of the "Human-AI System" as a result of their interaction, as well as possible integration and merging.

As the Editor, it was my privilege to oversee the development of the first edition of this monograph. I was particularly interested in chapters that not only highlighted the challenges of digitalisation in society, but also offered specific practical solutions for utilising artificial intelligence capabilities.

This monograph considers issues related to researching both the potential and shortcomings of the digitalisation of society. This issue features a variety of chapters that draw on the expertise of artificial intelligence specialists and reflect a wide range of ideas and perspectives.

Chapter 1 is devoted to the historical development and theoretical foundations of artificial intelligence. The authors review key milestones in the evolution of AI, emphasising that technological progress in this domain has not followed a strictly linear trajectory. The chapter analyses the interdisciplinary nature of AI, integrating biological, cognitive, philosophical, mathematical, and logical perspectives together with developments in machine learning, neural networks, and natural language processing. Particular attention is given to the transition from classical symbolic approaches to machine learning paradigms. The analysis demonstrates that neither symbolic nor statistical approaches alone are sufficient, highlighting the growing importance of hybrid methods in contemporary intelligent systems.

Chapter 2 examines the regulation of human interaction with artificial intelligence within the emerging “Human–AI System”. The authors propose a conceptual model for regulating the use of AI-based chatbots in scientific research and academic publishing. Central to this model is the AIC “AI Chatbots Attribution”, which promotes compliance with ethical and legal copyright standards. The chapter also addresses mechanisms for monitoring and managing human–AI interaction in the context of rapidly advancing digital technologies. Particular attention is given to the protection of fundamental human rights, including freedom of choice and the right to work, as well as the proposed “AI Free. Human Created” attribution.

Chapter 3 addresses the growing gap between the rapid advancement of artificial intelligence technologies and society’s capacity to govern and benefit from them equitably. The chapter examines the relationship between digital transformation, artificial intelligence, and societal change, highlighting how technological development reshapes institutions, governance structures, and human relationships. Drawing on qualitative documentary research and comparative case studies across healthcare, finance, education, and public services, the authors analyse both the opportunities and risks of AI-driven transformation. Using Vial’s Building Blocks of Digital Transformation framework, the study emphasises infrastructure integration, equitable value distribution, trustworthy organisational practices, and a human-centred approach to sustainable transformation.

Chapter 4 investigates the factors influencing artificial intelligence adoption among small and medium enterprises and its implications for organisational performance. The study develops an integrated analytical model that combines the technology–organisation–environment framework, diffusion of innovation theory, and ethical principles to explain AI adoption within SMEs. Using quantitative methods and structural equation modelling, the authors analyse survey data collected from South African enterprises. The findings indicate that compatibility, organisational readiness, employee capability, top management support, customer pressure, vendor support, and ethical considerations such as fairness, accountability, and transparency significantly influence AI adoption, thereby contributing to improved firm performance.

Chapter 5 provides a comprehensive review of artificial intelligence–driven chatbots and intelligent agents applied to monitoring, evaluation, and organisational learning. The chapter examines recent advances in conversational AI, including large language models, retrieval-augmented generation, and multi-agent architectures. Through a systematic literature review and bibliometric analysis of peer-reviewed publications from 2021 to 2025, the authors identify key research trends, core techniques, and emerging application domains. The findings highlight the growing role of AI-based conversational systems in organisational evaluation and knowledge management. The chapter synthesises current evidence, identifies research gaps, and outlines directions for future studies and evidence-based adoption.

Chapter 6 explores the use of artificial intelligence-based chatbots to enhance the sustainability and competitiveness of small and medium enterprises in the digital era. The study investigates factors influencing chatbot utilisation among South African SMEs by integrating the technology–organisation–environment framework, the technology acceptance model, and diffusion of innovation theory. A quantitative research design was employed, with survey data collected from 300 enterprises and analysed using structural equation modelling. The findings reveal that relative advantage, compatibility, organisational readiness, perceived usefulness, ease of use, ethical AI regulation, and top management support significantly influence chatbot adoption among SMEs, whilst security is less significant.

Chapter 7 analyses the transformative role of artificial intelligence as a driver of innovation in production systems, managerial decision-making, and consumer value creation. The chapter examines the application of AI technologies in smart manufacturing, including predictive maintenance, intelligent supply chains, and human–robot collaboration. It also explores the use of cognitive automation, predictive analytics, and scenario modelling to support augmented decision-making processes. In the consumer domain, the authors discuss hyper-personalised services enabled by recommendation systems, behavioural analytics, and conversational interfaces. The chapter emphasises that effective AI implementation requires high-quality data, workforce reskilling, robust governance frameworks, and responsible innovation

Chapter 8 focuses on governance, ethical challenges, and data security risks associated with the adoption of Generative Artificial Intelligence in contemporary organisations. The authors analyse emerging security threats related to large foundation models, including model poisoning, prompt injection, and risks of data leakage and intellectual property exposure. The chapter also examines ethical concerns such as inexplicable bias, limited transparency, and generate shadow vulnerability in AI systems. To address these challenges, the authors propose a socio-technical governance framework that integrates human oversight, explainable artificial intelligence, and continuous security monitoring within the generative AI deployment process.

Chapter 9 investigates the role of Smart Artificial Intelligence in advancing Industry 4.0 within the South African manufacturing sector. The chapter examines how cyber-physical systems, the Internet of Things, and intelligent automation contribute to improving productivity, efficiency, and competitiveness in industrial production. Based on a systematic literature review of studies published between 2022 and 2025, the authors analyse the opportunities and constraints associated with Smart AI adoption. The findings indicate that, despite its significant potential for digitalising production processes, implementation is hindered by financial limitations, inadequate technological infrastructure, and insufficient organisational capabilities within the manufacturing sector.

Chapter 10 examines the prospects for human–machine collaboration in Sub-Saharan Africa in the context of labour market transformation and technological development. Using a systematic literature review of studies published between 2020 and 2025, the chapter analyses how artificial intelligence may influence employment, poverty reduction, and social inequality within the framework of the Sustainable Development Goals and the capability approach. The findings reveal that limited infrastructure, low levels of digital and AI literacy, and insufficient technological resources constrain the effective adoption of advanced technologies. At the same time, the region's young population creates opportunities for universities to expand AI education and industry partnerships.

Chapter 11 explores the development of a higher education ecosystem based on the implementation of artificial intelligence technologies. The study analyses the benefits and challenges associated with the integration of AI into university teaching and learning processes. On this basis, the authors develop and substantiate a model for the optimal implementation of artificial intelligence within the higher education ecosystem using a systems approach. The proposed model includes structural (universities, faculties, departments, institutes, etc.) and functional (content of education, forms and methods of teaching, diagnosing of learning outcomes, administering of educational service, and eternal – include academic achievement: levels of knowledge, skills, and competences) components. The results are essential for developing university strategies for developing educational ecosystem.

Chapter 12 analyses the readiness of South African universities to adopt Generative Artificial Intelligence within the higher education sector. The chapter examines both the opportunities and challenges associated with the use of generative AI tools in academic environments, particularly in the absence of comprehensive institutional policies and guidelines. Using a systematic literature review and content analysis of academic publications, institutional reports, and policy documents, the authors assess levels of adoption according to the Generative AI maturity framework. The chapter proposes an analytical framework enabling universities to evaluate readiness, identify adoption gaps, and develop policies for responsible integration of generative AI technologies.

The monographs in this series represent an attempt to integrate the experience accumulated in recent years and propose new solutions to the challenges of AI using.

This series of monographs is an invaluable resource for researchers, teachers, postgraduate students, and practitioners in the field of AI.

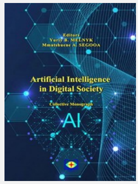
Information about the author:
Melnyk Yuriy Borysovych – https://orcid.org/0000-0002-8527-4638; Doctor of Philosophy in Pedagogy, Affiliated Associate Professor; Chairman of Board, Kharkiv Regional Public Organization "Culture of Health" (KRPOCH); Director, Scientific Research Institute KRPOCH, Ukraine.

Cite this chapter as:
Melnyk, Y. B. (2026). Introduction to the monograph on Artificial Intelligence in Digital Society, volume 1, 2026. In Y. B. Melnyk & M. A. Segooa (Eds.), *Artificial Intelligence in Digital Society, Vol.1.* (pp. 5–9). KRPOCH. https://doi.org/10.26697/aids.2026.0
The electronic version of this chapter is complete. It can be found online in the AIDS Archive https://doi.org/10.26697/aids

PART I

FOUNDATIONS AND EVOLUTION OF ARTIFICIAL INTELLIGENCE

Artificial Intelligence in Digital Society,
Volume 1, 2026
DOI: 10.26697/9786177089192.2026
ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)



Chapter 1. The Evolution of the Theory and Practice of Artificial Intelligence

Stadnik A. V. [1,2], **Mykhaylyshyn U. B.** [1]

[1] *Uzhhorod National University, Ukraine*
[2] *Kharkiv Regional Public Organization "Culture of Health", Ukraine*

Received: *01.12.2025;* ***Accepted:*** *10.02.2026;* ***Published:*** *10.03.2026*

Abstract
The authors first review the historical milestones in the emergence and development of artificial intelligence (AI). They demonstrate that progress is rarely linear, and today's successes do not guarantee solutions to future challenges. Further research shows that the theoretical foundation of AI is a multi-layered, interdisciplinary structure that includes biological, cognitive, philosophical, mathematical, and logical aspects, as well as machine learning, neural networks, and natural language processing. Modern artificial intelligence is the result of the interaction of these theoretical branches, each of which contributes to the effective operation of intelligent systems. The authors conclude that the evolution of AI from classical symbolic to machine learning represents a fundamental shift in the understanding and construction of artificial intelligence. However, a consideration of symbolic and statistical approaches shows that neither is ideal, often requiring hybrid solutions that combine multiple methods.

Keywords: evolution of artificial intelligence, machine learning, symbolic artificial intelligence, neural networks, machine learning, intelligent systems.

Cite this chapter as:
Stadnik, A. V., & Mykhaylyshyn, U. B. (2026). The evolution of the theory and practice of artificial intelligence. In Y. B. Melnyk & M. A. Segooa (Eds.), *Artificial Intelligence in Digital Society, Vol. 1* (pp. 111–26). KRPOCH. https://doi.org/10.26697/aids.2026.1
The electronic version of this chapter is complete. It can be found online in the AIDS Archive https://doi.org/10.26697/aids

Historical Milestones in Artificial Intelligence Research

Artificial Intelligence (AI) is a system that enables machines (computers, neural networks, etc.) to solve problems similar to those of the human mind, including perception, learning, reasoning, and decision-making, by analogy with human cognitive functions (Gacem & Aouane, 2024).

The preconditions for AI were formed long before the appearance of the first electronic computers. The possibility of creating and developing artificial humans, higher intelligence, and superintelligence capable of thinking better than humans has been a topic of discussion for quite some time. Ancient Greek myths about Hephaestus (god of fire and forge) described automatons (animate, metal statues of animals, people, and monsters). For example, Talos was a giant sculpted from bronze by Hephaestus who patrolled the island of Crete, protecting it from pirates; the Halkotaurus were bronze fire-breathing bulls (Fan et al., 2020). In medieval Islamic science, the Banu Musa brothers and Al-Jazari developed various automatic devices, including water clocks, self-playing musical mechanisms, and servomechanisms. Particularly revealing are Al-Jazari's "android" figures, which essentially represented early models of programmable automata containing feedback elements (Hill, 1991).

The concept of modern scientists was reduced to a partial imitation of the computational capabilities of the human brain through mathematical models, which subsequently led to the invention of mechanical computing machines:

1623 – Wilhelm Schickard came up with the "Counting clock" – the first adding machine capable of performing four arithmetic operations. The mechanism's operation was based on the use of stars and gears.

1642 – Blaise Pascal created "Pascal's calculator", which could only perform addition and subtraction.

1673 – Gottfried Leibniz's arithmometer, a mechanical calculating machine capable of addition, subtraction, multiplication, and division (Chang, 2020).

The development of autonomous mechanical computing devices became the prototype of modern AI technology and depended on the existing technological and computational capabilities of a particular period.

In the first half of the 20th century, numerous publications appeared devoted to the theoretical prerequisites for the creation of AI:

1920 – Czech playwright Karel Čapek released a science fiction play, "RUR", in which he proposed the idea of artificial humans, which he called robots (Chang, 2020).

1943 – McCulloch & Pitts published "A Logical Calculus of the Ideas Inherent in Nervous Activity", describing the first mathematical model of a neural network (McCulloch & Pitts, 1943).

1948 – Claude Shannon, in his work "Mathematical theory of communication", proposed to consider information as something new that is transmitted during communication. Information is measured in "binary digits" (0 or 1), better known as "bits" (Shannon, 1950).

In the second half of the 20th century, the term "artificial intelligence" was coined and entered into common usage, and interest in AI reached its peak:

1950 – Alan Turing published the book "Computing Machinery and Intelligence". He proposed the "Turing test", which asks whether a machine can fool a scientist into thinking it is communicating with a human (Turing, 1950).

1952 – Arthur Samuel wrote a computer checkers program that played at a master level with the ability to learn as you play (Samuel, 2000).

1956 – John McCarthy, Marvin Minsky, Nathaniel Rochester & Claude Shannon organized the Dartmouth Workshop on Artificial Intelligence, where the term was first used. Participants suggested that machine learning, language use, and problem-solving were problems that could be solved in the near future (McCarthy et al., 1955).

Then it continued, the creative development of artificial intelligence, from programming languages still in use today to books and films exploring the idea of robots:

1958 – John McCarthy created LISP (LISt Processing), the first programming language for AI research, which is still popular today (Moor, 2006).

1959 – Arthur Samuel coined the term "machine learning" (Samuel, 2000).

1961 – The first industrial robot, Unimate, began working on the assembly line at a General Motors plant in New Jersey.

1965 – Edward Feigenbaum & Joshua Lederberg created the first "expert system", which was a form of AI programmed to replicate the thinking and decision-making abilities of human experts.

1966 – Joseph Weizenbaum created the first "chatbot" ELIZA, a pseudo-psychotherapist that used Natural Language Processing (NLP) for communicating with people.

1979 – The Association for the Advancement of Artificial Intelligence (AAAI) was founded (AAAI, 2025).

Next, a period of rapid growth and decline of interest in AI has begun due to the lack of technological breakthroughs and reduced funding:

1980 – The first AAAI conference was held at Stanford (AAAI, 2025).

1980 – The first expert system, known as XCON (Expert Configurator), appeared on the commercial market. It was designed to assist in ordering computer systems by automatically selecting components according to the customer's needs.

1985 – An autonomous drawing program known as AARON is demonstrated at the AAAI conference.

1986 – Ernst Dieckmann and his team from the Bundeswehr University of Munich created and demonstrated the first driverless car (robomobile). It could reach speeds of up to 80 km/h on roads without obstacles or human drivers.

1987 – Alacrity was commercially launched by Alacrity Inc. Alacrity was the first strategic management consulting system and used a sophisticated expert system.

1987 – The market for specialized LISP-based hardware collapsed due to the emergence of cheaper and more accessible competitors capable of running LISP software, including products from IBM and Apple.

1988 – Rollo Carpenter invented the chatbot Jabberwacky, which he programmed to carry on interesting and engaging conversations with people (AAAI, 2025).

Despite the lack of funding, the early1990s saw impressive advances in AI research, including the creation of the first AI system capable of defeating a reigning world chess champion, the introduction of AI into everyday life (the first Roomba robot vacuum cleaner, the first commercial speech recognition software on Windows computers):

1997 – Deep Blue (IBM) defeated world chess champion Garry Kasparov, becoming the first program to beat a human world chess champion;

1997 – Windows released speech recognition software (Dragon Systems);

2000 – Professor Cynthia Breazeale has developed the first robot capable of imitating human emotions using a face including eyes, eyebrows, ears, and mouth;

2002 – The first Roomba robot vacuum cleaner was released;

2003 – NASA landed two rovers (Spirit & Opportunity) on Mars, and they explored the planet's surface without human intervention;

2006 – Twitter, Facebook, and Netflix began using AI as part of their advertising and user experience algorithms;

2010 – Microsoft released the Xbox 360 Kinect, the first gaming device designed to track body movements and convert them into gaming commands;

2011 – Apple released Siri, the first popular virtual assistant (AAAI, 2025; Chang, 2020).

At this time, digital product is becoming one of the most rapidly developing areas (Pypenko, 2019). We are witnessing a surge in the popularity of AI tools, including virtual assistants and search engines. Deep learning and big data have also gained traction during this period:

2012 – Jeff Dean and Andrew Ng trained a neural network to recognize cats;

2016 – Hanson Robotics has created a humanoid robot named Sofia with a realistic human appearance and the ability to see and reproduce emotions, as well as communicate;

2017 – Facebook programmed two AI-powered chatbots to communicate and learn how to negotiate, but during the process of communicating, they eventually abandoned English and began developing their own language, completely autonomously (Chang, 2020);

2019 – Google's AlphaStar reached Grandmaster level in the video game StarCraft 2, outperforming almost all human players;

2020 – OpenAI began beta testing GPT-3, a deep learning model. It became the first model to produce content virtually indistinguishable from human-created content;

2021 – OpenAI developed DALL-E, which can process and understand images at a level sufficient to generate accurate captions (AAAI, 2025; Chang, 2020);

2023 – In the Academic International Corporation, the problems of legitimising AI-based ChatBots in scientific research were studied, and an attribution (AIC AI Chatbots) was developed, which is proposed for use in indicating the role and level of involvement of AI and ChatBots in research and publications (Melnyk & Pypenko, 2023);

2025 – Large Language Models Pass the Turing Test (ELIZA, GPT-4o, LLaMa-3.1-405B, and GPT-4.5) (Jones & Bergen, 2025).

Thus, while modern achievements are impressive, they also remind us of the need for a sober assessment of the capabilities and limitations of AI technology. History teaches us that progress is rarely linear, and today's successes do not guarantee solutions to tomorrow's challenges. Fundamental questions about the nature of intelligence and its understanding remain open. The future of AI will be determined not only by technical breakthroughs but also by society's ability to develop ethical frameworks, regulatory mechanisms, and social institutions for the responsible development and application of this technology.

Theoretical Foundations of Artificial Intelligence Systems

The theoretical aspects of AI systems are diverse and encompass *biological, cognitive, philosophical, mathematical, logical, machine learning, neural networks, and natural language processing*. They reflect different views on intelligence, cognition, and system operations.

Biological Aspects. The origins of AI, as well as its history, are closely linked to brain sciences (neurophysiology, anatomy and physiology of the nervous system, psychology, etc.). Many of the founding scientists of AI are also brain scientists, and many discoveries in AI are based on biological research. For example, working memory, which was discovered using magnetic resonance imaging, inspired the development of the memory module in machine learning models, ultimately leading to the creation of the long short-term memory (LSTM) model. Changes in the spinal cord that occur during learning have inspired the creation of a new algorithm, *Elastic Weight Consolidation* (EWC), for continuous learning. Neural connections in the human brain, discovered using a microscope, inspired the development of artificial neural networks (Fan et al., 2020).

The goal of AI is to develop computer systems capable of performing tasks traditionally performed by human intelligence, with functions such as information reception, processing, decision-making, and control. The goal of brain science is to study the structure, functions, and operating mechanisms of the biological brain (perception, recognition of multisensory information, and decision-making regarding interactions with the environment). Thus, AI is very similar to human intelligence and can be considered a simulation of the human brain's cognitive abilities (Miśkiewicz, 2019). A comparison of human intelligence and AI has revealed that, although both systems are capable of performing similar tasks, their underlying mechanisms and limitations are fundamentally distinct (Table 1).

Table 1.1

A Final Comparison of Human Intelligence and Artificial Intelligence

Characteristic	Human Intelligence	Artificial Intelligence
Warp	Biological (neural activity, neuromodulation, and metabolic restrictions)	Digital and machine (logical operations, energy-efficient matrix multiplication, and discrete data representation)
Information processing	Analog and discrete	Discrete
Types of errors	Biologically stochastic (random)	Systemic (mathematically determined)
Consciousness	Possesses consciousness or self-awareness	Does not possess consciousness or self-awareness
Thinking	Abstract and heuristic based on experience, emotion, and cultural context	Heuristic based on computations, search algorithms, probabilistic inferences
Adaptation and decision making	Based on life experience, emotions, and cultural context	Based on pre-programmed algorithms and data, they often require reprogramming
Creation	Based on associative thinking, life experience, emotional, and cultural context	Based on patterns and the recombination of information from training data
Embodiment and Cognition	They develop together, sensorimotor feedback shapes learning, and bodily experience influences abstract thinking.	When integrated into robotic systems, embodiment is not phenomenologically or biologically equivalent.
Education	Possibly one-shot learning and generalization based on minimal information	Repetitive learning from new large data sets

Thus, AI and human intelligence share standard features in the ability to learn and adapt, but differ significantly in their nature, operating mechanisms, and capabilities.

Cognitive Psychology helps understand and model thought and perception processes. The mental aspects of AI are rooted in an attempt to describe human thinking as an information-processing process. Digital AI models are partly inspired by theories of cognitive science, which study perception and processing of sensory information, memory, attention, logical and associative thinking, and decision-making.

Fundamental components of AI cognitive constructs are knowledge representation mechanisms. Several approaches have emerged in artificial intelligence: *the symbolic approach* (knowledge is represented in the form of logical structures and rules); *the subsymbolic approach* (knowledge is stored in distributed representations characteristic of neural networks); and *the hybrid approach* (a combination of logical structures and neural network learning). Each approach reflects specific views on the functioning of the brain's cognitive systems, including the processing of contexts, connections between objects, and the ability to generalize.

Modern AI systems demonstrate functional analogs of human cognitive processes: Perception is realized through computer vision models, speech recognition, and multimodal transformers. Attention is mathematically formalized in the self-attention mechanism (self-attention), which allows us to highlight the most significant elements of the input data. Memory is represented by internal state structures (LSTM, GRU), external storage (Differentiable Neural Computers), or a quasi-semantic, multidimensional vector space (embedding space). Although such systems do not reproduce the biological mechanisms of intelligence, they perform similar functions, which allows them to be considered as cognitive models (Achler, 2024).

The Philosophy of AI examines the nature, consciousness, and ethical implications of utilizing AI. For example, it asks questions such as, Can a machine think like a human? Does it possess consciousness and subjective experience? A modern view of the Turing Test, according to the physical-symbolic systems hypothesis (Newell & Simon, 1971), argues that symbol manipulation is sufficient for intelligence. Meanwhile, John Searle's "Chinese Room" argument demonstrates that a program manipulates symbols syntactically but does not comprehend their semantics, much like a person in a room following instructions in Chinese without understanding the language. This contradicts the idea that the brain works like a computer (Searle, 1993).

Philosophical aspects of AI also include issues of consciousness, free will, the alignment of human values, ethics, and bias. Key issues include whether a machine can possess consciousness, how to ensure ethical behavior in AI, and how to integrate human values into algorithms without distortion or bias. Furthermore, the problem of explaining AI decisions and controlling its actions in society is discussed. To address these issues, an interdisciplinary approach that brings together philosophers, scientists, and engineers is essential, as is the development of transparent and interpretable algorithms (Gacem & Aouane, 2024).

Mathematical Aspects of AI encompass *probability theory, mathematical statistics, linear algebra, and optimization algorithms*.

Linear algebra provides tools for representing and processing data using vectors and matrices. These structures play a key role in the development of machine learning models and neural networks.

Vectors and matrices are fundamental elements of representing data in feature space. Eigenvalues and eigenvectors play a crucial role in dimensionality reduction and data analysis using methods such as principal component analysis (PCA) (Jolliffe, 2002). These concepts enable the extraction of the most relevant information from large and complex datasets.

Probability theory enables us to model the uncertainty and randomness inherent in data. *Statistical methods* are essential for evaluating models and drawing conclusions from data (Casella & Berger, 2002). Probability distributions describe the distribution of values of a random variable, which is essential when modeling stochastic processes. Bayesian theory and probabilistic models enable us to update probabilities based on new data. This is particularly useful when developing adaptive systems that learn in dynamic environments.

Optimization algorithms are at the core of the training process for AI models. Gradient descent is used to minimize the loss function by updating parameters in the direction of the steepest slope (Ruder, 2016). Stochastic gradient descent (SGD) is a variation of the method that uses random subsets of the data to speed up training, which is particularly useful for large datasets (Almudevar, 2021).

Logic*,* along with mathematical methods, plays a crucial role in representing knowledge, formalizing reasoning, and constructing logical programming systems. These aspects enable the creation of inference, deduction, and proof algorithms used by many AI systems. Logical approaches are based on various types of formal logic, including classical propositional logic, predicate logic, and specialized description logics. These methods allow knowledge and rules to be precisely described in the form of formal statements. Knowledge bases and production systems are used to represent knowledge, and the reasoning process is implemented using inference algorithms, such as resolution. The logical programming language Prolog serves as a practical example of the application of logical foundations in AI, where rules and facts are specified declaratively, and inference is performed automatically (Genereth & Nilsson, 1987).

A key element of the logical foundations of AI is inference algorithms, which enable the generation of new knowledge from a given set of facts and rules. The main goal is not simply to obtain an answer, but to explain its logic, preserving the reasoning tree and the possibility of counterfactual analysis. Such approaches form the basis of explainable AI (XAI), which is becoming increasingly popular in modern systems. This ensures the transparency and trust in intelligent systems (Genereth & Nilsson, 1987).

In recent years, there has been an increasing integration of logical and algorithmic methods with machine learning. Hybrid neural-symbolic architectures combine the ability of neural networks to process unstructured data with the ability of logic to provide formal reasoning and explanation. Such systems use logic to ask "why" and the conditions under which decisions are made, while neural networks address the question of "what to predict". These approaches open up new possibilities for creating more adaptive, simultaneously explainable, and formally verifiable intelligent statements.

Machine Learning is a core component of artificial intelligence, enabling systems to learn from data, make predictions, or make decisions without being explicitly programmed (Mitchell, 1997). There are three primary learning paradigms: *supervised, unsupervised, and reinforcement learning*.

In *supervised learning*, models are trained on labeled data, where each input is associated with a corresponding output value. Regression predicts continuous values, while classification categorizes data into discrete categories using methods such as logistic regression and support vector machines (SVM) (Bishop, 2006).

Unsupervised learning methods work with unlabeled data and aim to discover hidden structures or patterns within it. Clustering groups similar data, for example, using the k-means method. Dimensionality reduction methods, such as PCA, reduce data complexity while preserving important information (Hastie, Tibshirani, & Friedman, 2009).

Reinforcement learning involves training an agent by interacting with its environment and learning from the consequences of its actions, which are mediated by rewards or punishments (Sutton & Barto, 2018). This approach is efficient in problems where the sequence of decisions influences the final goal.

Neural Networks are models inspired by biological neurons that can learn complex functions and representations (Goodfellow, Bengio, & Courville, 2016). Artificial neurons are basic devices that take input and generate output through an activation function. Multilayer perceptrons (MLPs) are neural networks with one or more hidden layers that can model nonlinear relationships. The universal approximation theorem states that neural networks with sufficient neurons can approximate any continuous function (Hornik, Stinchcombe, & White, 1989). Convolutional neural networks (CNNs) specialize in image processing by capturing spatial dependencies (LeCun et al., 2015). Recurrent neural networks (RNNs) are well-suited for sequential data, such as text and audio, because they account for temporal dependencies.

Natural Language Processing *(NLP)* is another crucial AI task enabled by computational linguistics. The theoretical aspects of NLP encompass *syntactic,* *semantic, and pragmatic analysis, statistical text processing methods, and consideration of linguistic constructs*. This enables the creation of intelligent translation systems, conversational interfaces, and assistants that understand and generate spoken language.

NLP process involves several stages.

1. Data entry involves receiving text or voice data.

2. Pre-processing involves cleaning and structuring data (e.g., tokenization, stop word removal).

3. Meaning extraction involves using machine learning algorithms to analyze the meaning of words in context, disambiguate, and infer user intent (Alammar, 2025).

4. An appropriate response or performing a task based on the extracted value (Manning et al., 2020).

Modern NLP is based on highly efficient computational models: word embedding learning: word2vec, GloVe – represent words as vectors encoding semantic similarity; ELMo, BERT - take into account the dependence of word meaning on context; *transformers* - provide parallel processing of sequences and deep contextual understanding; *large language models* (LLM) – trained on huge data and capable of performing a variety of tasks without specialized tuning (Ethayarajh, 2019).

Thus, the theoretical foundations of AI form a multilayered, interdisciplinary structure, incorporating biological, cognitive, philosophical, mathematical, and logical aspects, as well as machine learning, neural networks, and natural language processing. Modern AI is the result of the interaction of these theoretical fields, each contributing to our understanding of the nature of intelligence and the construction of effective intelligent systems.

From Symbolic Artificial Intelligence to Machine Learning: Paradigm Shifts

Problems in AI can be solved in different ways. Some approaches (symbolic) assume that the mind is a system of symbolic representations and logical manipulations of them; others (statistical and machine learning) interpret intelligence as the ability to extract patterns from data and optimize decisions. The transition from the first to the second approach was not immediate. It represented a gradual shift in emphasis, an interweaving of methods, and a reflection on which properties of intelligence are more important for practice.

Symbolic AI (Good Old-Fashioned AI (GOFAI) is based on the Newell and Simon hypothesis (1971), which posits that intelligence can be modeled through the manipulation of symbols that express knowledge and the logical relationships between them. The methodology is based on constructing productive rules ("if-then") that connect symbols into logical relationships. Using these rules, systems draw conclusions, form hypotheses, and determine what additional information to request. This structure enables the modeling of cognitive processes with consistent logic and adaptation based on knowledge, rather than relying solely on statistics.

This methodology also draws on biological analogies, where the model includes neurons (perceiving and discriminating) that are hierarchically organized into structures capable of complex inference. Symbolic AI creates models that closely resemble the cognitive processes of living organisms, facilitating the explanation of results and the analysis of critical factors influencing decisions. Furthermore, the reasoning process is transparent and traceable, facilitating its understanding and debugging. Rule-based systems are less demanding on computing resources. Symbolic AI can often run on computer CPUs, making it more energy-efficient than data-intensive machine learning approaches that typically require powerful GPUs.

The main tools of symbolic AI are:

- knowledge bases containing a set of rules and facts in the form of symbols;
- means of formalizing knowledge (for example, production rules; logical expressions, semantic networks, frames;
- logical interpreters and inference engines that apply rules to reason and make decisions;
- formalized knowledge representation languages and explanation systems that provide logic traceability (Prolog and other logical languages) (Garrido-Merch & Puente, 2025).

Symbolic AI, dominant from the 1950s to the 1980s, used hand-crafted rules and logical reasoning to model intelligence. It was effective for tasks requiring explicit knowledge representation, such as expert systems, but struggled with the ambiguity, scalability, and complexity of the real world.

The 1980s and 1990s saw the modernization and growth of computing power, enabling greater use of large databases and statistical methods that allow machines to learn based on patterns in the data they receive, rather than relying solely on predefined rules. This made it possible to process probabilistic, unstructured, and variable data, which was difficult for symbolic methods.

This paved the way for a radical paradigm shift: the focus shifted from explicit, transparent, and precise knowledge to data and optimization. This led to the emergence and rise of machine learning (Goodfellow, Bengio & Courville, 2016).

Machine Learning is a set of methods where models extract patterns from examples. Machine learning is based on *statistics, information theory, and computational learning theory*. The key idea is that instead of explicitly programming rules for solving a problem, a system should automatically extract patterns from data. This represents a shift in emphasis from algorithm design to the design of learning architectures and the collection of relevant data (Bishop, 2006).

The primary methods of machine learning are *defining the model architecture, loss function, optimization algorithm, and validation procedure*. Machine learning's key strengths include adaptability and scalability, robustness to noise and partial data distortion, and practical efficiency, often exceeding human performance in applied tasks.

The most dramatic paradigm shift has occurred in the last fifteen years, with the resurgence and triumph *of deep neural networks*. Although the basic algorithms had existed for decades, the convergence of several factors led to a qualitative leap in their performance. The availability of massive datasets, thanks to the internet and digitalization, provided training material. The computing power of graphics processing units made it possible to train networks with billions of parameters. Algorithmic innovations, such as batch normalization, attention mechanisms, and residual connections, have addressed the challenges of training deep networks. Breakthrough results followed one after another. In 2012, the AlexNet convolutional neural network radically outperformed traditional methods in the ImageNet image recognition competition. Machine translation systems based on recurrent networks and attention mechanisms achieved near-human quality. Image and text generation models demonstrated impressive creative capabilities (Krizhevsky et al., 2012).

The emergence of the Transformer architecture in 2017, along with the subsequent development of large-scale language models, was particularly significant. Systems (GPT, BERT, and others) have demonstrated that pre-training on massive text data, followed by fine-tuning, can create models with surprisingly broad natural language processing capabilities, without requiring explicit encoding of linguistic rules (Ethayarajh, 2019).

The shift from symbolic AI to machine learning reflects deep, fundamental disagreements about the nature of knowledge and intelligence. The symbolic approach assumes that intelligence requires explicit, declarative representations of knowledge and its manipulation through logical rules. Machine learning, intense learning, embodies the position that knowledge emerges from experience and does not necessarily have an explicit symbolic form. Expertise can exist without explicit representation of rules; knowledge can be embodied and procedural (Gorner, 2007).

Methodologically, the two approaches differ in their method of problem-solving. *Symbolic AI* follows a top-down strategy: problem analysis, decomposition into subtasks, knowledge formalization, and construction of reasoning algorithms. This is an engineering approach, where each system component is designed to perform a specific function. *Machine learning* is primarily a bottom-up approach, involving the selection of an appropriate architecture, data collection, model training, and model validation on independent data. The inner workings of a trained model often remain opaque; knowledge is extracted inductively from statistical regularities rather than deductively constructed.

Symbolic AI retains its advantages in tasks that require explicit reasoning, the explanation of decisions, guaranteed correctness, and working with a small number of examples. Symbolic systems naturally support composition, systematic thinking, and the integration of heterogeneous knowledge sources. Their behavior is predictable and verifiable. *Machine learning* excels in tasks of pattern recognition, working with noisy data, generalization to new situations, and processing sensory information. It naturally scales to big data and automatically adapts to changes in the data distribution. Formalization of costly expert knowledge is not required.

However, machine learning also has significant weaknesses. Models can be opaque black boxes, making them difficult to understand and debug. They require large volumes of labeled data and computational resources. Generalization is limited by proximity to the training distribution, and models are vulnerable to adversarial examples. Integrating symbolic knowledge and common sense remains a challenge.

The big data revolution has fundamentally changed the AI landscape. The exponential growth of computing power, particularly the advent of specialized accelerators like GPUs, has enabled the training of models with billions of parameters. What seemed computationally infeasible just twenty years ago is now routinely performed in research labs and even on personal computers.

The commercial success of machine learning has attracted massive investment from tech companies and venture capital funds, creating a powerful economic incentive for further research and development. The availability of open data, code, and pre-trained models has made a positive feedback loop, accelerating progress. The integration of machine learning into educational programs and the development of specialized training courses have led to an influx of talent into the field, further accelerating its growth (Sifatkaur et al., 2023).

However, due to the lack of explainability and opacity of purely statistical models, the field of neuro-symbolic AI, which combines the advantages of both approaches, has recently developed. This hybrid approach seeks to combine the adaptability and ability to learn from machine learning data with the logical rigor and explainability of symbolic systems, which is particularly important for applications in mission-critical areas (Gacem & Aouane, 2024).

Thus, the evolution of AI from classical symbolic to machine learning represents a fundamental shift in the understanding and construction of artificial intelligence – from manual knowledge processing to systems capable of autonomous learning and adaptation, while maintaining the desire for transparency and controllability of decisions. The transition from symbolic to statistical approaches shows that neither provides all the answers. First, symbolic systems have proven insufficient for understanding the full complexity of human language.

Statistical systems have succeeded not by replacing symbolic representations, but by finding new ways to describe patterns that symbolic approaches struggled to formalize. Second, practical AI problems often require hybrid solutions based on multiple approaches. Third, the qualities of symbolic systems (interpretability and explainability) sometimes outweigh all the advantages of modern neural systems.

References

Achler, T. (2024). What AI, neuroscience, and cognitive science can learn from each other: An embedded perspective. *Cognitive Computation, 16,* 2428–2436. https://doi.org/10.1007/s12559-023-10194-9

Alammar, J. (2025). Large language models: Architecture and training. From next-word prediction to reasoning. *In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.*2 (KDD '25). Association for Computing Machinery, New York. USA. https://doi.org/10.1145/3711896.3736805

Almudevar, A. (2021). *Theory of statistical inference* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003049340

Association for the Advancement of Artificial Intelligence. (2025). *About the Association for the Advancement of Artificial Intelligence (AAAI).* https://aaai.org

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer, Berlin. https://link.springer.com/book/9780387310732

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Duxbury Press, Pacific Grove. https://pages.stat.wisc.edu/~shao/stat610/Casella_Berger_Statistical_Inference.pdf

Chang, A. C. (2020). History of artificial intelligence. In *Intelligence-Based Medicine, Artificial Intelligence and Human Cognition in Clinical Medicine and Healthcare* (pp. 23–27). https://doi.org/10.1016/B978-0-12-823337-5.00002-0

Ethayarajh, K. (2019). *How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings.* ArXiv. https://doi.org/10.48550/arXiv.1909.00512

Fan, J., Fang, L., Wu, J., Guo, Y., & Dai, Q. (2020). From brain science to artificial intelligence. *Engineering, 6*(3), 248–252 https://doi.org/10.1016/j.eng.2019.11.012

Gacem, H., & Aouane, A. (2024). Conceptual foundations of artificial intelligence. *Journal of El-Manhel Economy*, *7*(1), 1215–1224. https://asjp.cerist.dz/en/article/247713

Garrido-Merchán, E., & Puente, C. (2025). *GOFAI meets Generative AI: Development of expert systems by means of large language models.* ArXiv. https://doi.org/10.48550/arXiv.2507.13550

Genesereth, M. R., & Nilsson, N. J. (1987). *Logical foundations of artificial intelligence*. Los Altos, CA: Morgan Kaufmann. https://kryten.mm.rpi.edu/COURSES/LOGAIS02/nonmonotonic.pdf

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press. https://www.deeplearningbook.org/

Gorner, P. (2007). Authenticity. In *Heidegger's Being and Time: An Introduction* (pp. 105–152). Cambridge University Press. https://doi.org/10.1017/CBO9780511808036.009

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The element of statistical learning: data mining, inference, and prediction.* Springer, Berlin. https://doi.org/10.1007/978-0-387-84858-7

Hill, D. (1991). Mechanical engineering in the Medieval Near East. *Scientific American*, *264*(5), 100–105. https://www.jstor.org/stable/24936907

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer-Verlag. https://link.springer.com/book/10.1007/b98835

Jones, C. R., & Bergen, B. K. (2025). *Large language models pass the Turing test.* ArXiv. https://doi.org/10.48550/arXiv.2503.23674

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25 Proceedings* (pp. 1097–1105). Curran Associates Inc. https://papers.nips.cc/paper_files/paper/2012

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521,* 436–444. http://dx.doi.org/10.1038/nature14539

Manning, C., Clark, K., Hewitt, J., Khandelwal, U., & Levy, O. (2020). Emergent linguistic structure in artificial neural networks trained by self-supervision. *Proceedings of the National Academy of Sciences, 117*(48), 30046-30054. https://doi.org/10.1073/pnas.1907367117

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A proposal for the Dartmouth summer research project on artificial intelligence.* http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

McCulloch, W., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5,* 115–133. https://doi.org/10.1007/BF02478259

Melnyk, Yu. B., & Pypenko, I. S. (2023). The legitimacy of artificial intelligence and the role of ChatBots in scientific publications. *International Journal of Science Annals, 6*(1), 5-10. https://doi.org/10.26697/ijsa.2023.1.1

Miśkiewicz, J. (2019). The merger of natural intelligence with artificial intelligence, with a focus on Neuralink company. *Virtual Economics*, *2*(3),

22–29. https://doi.org/10.34021/ve.2019.02.03(2)
Mitchell, T. (1997). *Machine learning*. McGraw-Hill Higher Education, New York. https://www.cs.cmu.edu/~tom/files/MachineLearningTomMitchell.pdf
Moor, J. (2006). The Dartmouth college artificial intelligence conference: The next fifty years. *AI Magazine*, *27*(4), 87–89. https://doi.org/10.1609/aimag.v27i4.1911
Pypenko, I. S. (2019). Digital product: The essence of the concept and scopes. *International Journal of Education and Science, 2*(4), 56. https://doi.org/10.26697/ijes.2019.4.41
Samuel, A. L. (2000). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *44*(1.2), 206–226. https://doi.org/10.1147/rd.441.0206
Searle, J. R. (1993). Consciousness, explanatory inversion, and cognitive science. *Behavioral and Brain Sciences*, *13*(4), 585–596. https://doi.org/10.1017/S0140525X00080304
Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
Sifatkaur, D., Manmeet, S., Vaisakh S. B., Neetiraj, M., & Sukhpal, S. (2023). Mind meets machine: Unravelling GPT-4’s cognitive psychology. *Bench Council Transactions on Benchmarks, Standards and Evaluations, 3*(3), Article 100139. https://doi.org/10.1016/j.tbench.2023.100139
Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist, 26*(2), 145–159. https://doi.org/10.1037/h0030806
Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). A Bradford Book, the MIT Press. https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf
Turing, A. M. (1950). Computing machinery and intelligence. *Mind, LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Information about the authors:

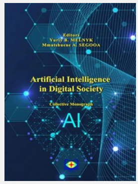
Stadnik Anatoliy Volodymyrovych – https://orcid.org/0000-0002-1472-4224; Doctor of Philosophy in Medicine, MD, Affiliated Associate Professor, Kharkiv Regional Public Organization “Culture of Health”, Kharkiv, Ukraine; Uzhhorod National University, Uzhhorod, Ukraine.

Mykhaylyshyn Ulyana Bohdanivna – https://orcid.org/0000-0002-0225-8115; Doctor of Psychological Sciences, Full Professor; Head of the Department of Psychology, Uzhhorod National University, Uzhhorod, Ukraine.

PART II

REGULATION, ETHICS, AND THE FUTURE OF ARTIFICIAL INTELLIGENCE-DRIVEN TRANSFORMATION

Artificial Intelligence in Digital Society,
Volume 1, 2026
DOI: 10.26697/9786177089192.2026
ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)



Chapter 2. The Regulation of Human Interactions with Artificial Intelligence

Pypenko I. S. [1,2], **Melnyk Y. B.** [1,2]

[1] *Kharkiv Regional Public Organization "Culture of Health", Ukraine*
[2] *Scientific Research Institute KRPOCH, Ukraine*

Received: *01.12.2025;* ***Accepted:*** *10.02.2026;* ***Published:*** *10.03.2026*

Abstract

Human civilisation has entered a new phase in the development of the digital technology age, where the emergence of artificial intelligence (AI) has given rise to new systems of interaction: the "Human-AI System". This system involves establishing certain rules and norms for human interaction with AI. This chapter describes our proposed model for regulating the use of AI-based chatbots in scientific research and publications. The model involves the use of AIC "AI Chatbots Attribution", which promotes compliance with ethical and legal copyright standards. This chapter also addresses the issue of controlling and managing this system of human interaction with AI. Given the great potential and speed of development of AI-based digital and information technologies, we may lose our position of leadership in this field in the near future. We believe that, very soon, human activity that does not make use of AI will need to defend its right to exist. These are the natural human rights of freedom of choice and the right to work. The attribution or logo "AI Free. Human Created", developed by the authors to indicate that the product was created by a human without the involvement of AI, can be used to classify products. We are confident that in the near future, highly developed countries will develop, ratify, and implement laws regulating the norms of interaction and relations between humans and AI.

Keywords: artificial intelligence, Human-AI System, AIC "AI Chatbots Attribution", "AI Free. Human Created", ethical and legal standards, interactions and relationships.

Cite this chapter as:

Pypenko, I. S., & Melnyk, Y. B. (2026). The regulation of human interactions with artificial intelligence. In Y. B. Melnyk & M. A. Segooa (Eds.), *Artificial Intelligence in Digital Society, Vol. 1.* (pp. 28–41). KRPOCH. https://doi.org/10.26697/aids.2026.2

The electronic version of this chapter is complete. It can be found online in the AIDS Archive https://doi.org/10.26697/aids

Model for Regulating the Use of AI-based Chatbots in Scientific Research and Publications

We are now seeing an increasing trend of using chatbots based on artificial intelligence (AI) in scientific research and writing. It is no secret that machine-readable texts today are more demanding and more readable. We live in a time when machines write texts that are read by machines far more often than by humans.

Several companies have announced the development of AI-based chatbots: OpenAI's ChatGPT, Google's Bard, Microsoft's Bing (a search engine with a chatbot), etc. There are already many AI tools with different specializations for text, photos, videos, etc. AI tools are developing at an unimaginably fast pace.

Is chatbots an advanced search engine? Or is it a real human intellectual competitor capable of exploring, learning, improving, creating?

Discussions about the trends and replacement of humans by AI, and the possible threats associated with it, have been ongoing since the term was introduced by John McCarthy (1959) in the middle of the last century.

This type of discussion is characteristic of most innovations. Think back to the discussions about robotics. Just as in the current AI situation, people saw benefits, problems, and threats. In the AI situation, things have become even more complicated because it has a new characteristic – learnability, as well as the use of the Large Language Model (LLM).

To answer the above question, it is necessary to consider the essence of this phenomenon. There are many aspects to this problem: from the physical level (availability and quality of servers) to the moral and ethical level (rules, norms, values, etc.).

There is no denying that AI, including chatbots such as GPT, has enormous potential to greatly facilitate our daily lives and be an indispensable assistant in professional activities.

A number of scientists believe that AI and chatbots are real competitors of humans in their professional activities and may replace them in many areas in the near future (Çalli & Çalli, 2022; Dans, 2019; Dimitriadou & Lanitis, 2023; Singh & Sood, 2022).

There are also often radical views that argue that the development of AI and the proliferation of chatbots could lead to a loss of control over them and even the extinction of humanity (Farahani, 2023).

It is normal to have different points of view about new phenomena. However, one cannot ignore the personal position of those who are leading the development of these technologies and systems. They are more immersed in the problem than others, aware of the latest research, and able to anticipate trends more objectively. Their disagreement and lack of a unified view on the prospects of using AI can have ambiguous consequences. On the one hand, it generates competition, which contributes to the development of this market and to innovation. On the other hand, we cannot be completely sure that we will not lose something more important in the pursuit of profit and the desire to lead.

The study aims to consider the issues that arise for researchers, authors and publishers when preparing scientific publications in relation to the norms of interaction and relations between humans and AI, and to propose an attribution that would reflect the role and level of involvement of AI and specific chatbots in a given study. The study also aims to design a basic logo for products created by humans without AI involvement.

In this chapter, we will not discuss the advantages, disadvantages, and limitations for human use of AI. We will limit ourselves to considering the problem in the area of using AI for scientific research and publication. To be fair, the rivalry between AI and humans is indeed growing. In the near future, we can expect AI to increasingly displace humans from certain areas of activity, including consulting services, telemedicine, online education, journalism, IT, etc.

This problem raises a number of fundamental questions: can AI significantly influence (replace) human activity in the Human-Human System with the new Human-AI System?

This is a fundamentally new system that raises even more questions, especially how it will affect the quality of life of the individual himself.

First of all, it is necessary to describe this definition.

Human-AI System is a complicated dynamic complex of interactions between living and non-living matter, is an accumulation of coordinated, interdependent and interconnected informational-technological actions of human and AI, oriented to learn from the information obtained, designed to effectively perform tasks and achieve goals (Melnyk & Pypenko, 2023).

While the answers to some questions are obvious (technology and robotisation have made heavy and monotonous work easier, computerization and the Internet have helped speed up information retrieval and processing), the use of AI, including chatbots, remains uncertain. This is especially true in the intellectual sphere: scientific research, media publications, etc.

Some of the positive things about using AI and chatbots are that they can find relevant documents, summarize text and draw conclusions from documents, make predictions, answer questions quickly, and argue for answers based on the latest scientific research.

Despite all these impressive benefits, we have some doubts about the pace and scope of AI delegation. Would not the use of AI accelerate the pace of life so much that we lose control over it? You would agree that this small factor could radically affect our lives. Therefore, the problem of AI legitimacy needs to be addressed as soon as possible.

Since scientists are (still) the leaders of innovation and the level of development of society depends on them, let us consider the role of chatbots in scientific research and publications. It is in scientific research and publications that ideas are first expressed and then put into practice, significantly affecting human activity and life on the planet as a whole.

Existing search engines and the emergence of new chatbots, such as ChatGPT, which use language models, greatly simplify the process of preparing and writing scientific research and publications. They can help authors automate research workflows such as literature searching, literature review, statistical analysis, and more.

In this chapter, we would like to introduce our idea of creating a digital platform that has the potential to legitimize and regulate the use of AI, intelligent search engines, chatbots in scientific and practical human activities. And first of all, it should be implemented for scientific research and publications.

In our opinion, one of the most obvious and simplest ways to solve this problem is to use licenses and attribution. The attribution we developed (AIC AI Chatbots, 2023) has several types (AIC "AI Chatbot Text" / AIC "AI Chatbot Image" / AIC "AI Chatbot Video") that provide different contributions and allows the user(s) to select the type required for scientific research (Figure 2.1). Specifying this attribution and fulfilling the conditions for its use will help to ensure ethical and legal standards in research activities.

Figure 2.1

AIC AI Chatbots Attribution to Indicate the Use of AI-based Chatbot

a) b) c)

Note. Abbreviations: AIC, Artificial Intelligence-based Chatbot; AI, Artificial Intelligence; a) AIC "AI Chatbot Text", attribution used for Text generated by an AI-based Chatbot; b) AIC "AI Chatbot Image", attribution used for Image generated by an AI-based Chatbot; c) AIC "AI Chatbot Video", attribution used for Video generated by an AI-based Chatbot (https://doi.org/10.26697/ai.chatbots)

The left segment with a gray background contains a hexagonal figure with the AIC abbreviation centered on white background. The AIC abbreviation stands for AI-based Chatbot as well as Academic International Corporation, which provides this platform.

The right segment with a white background contains the "AI Chatbot" inscription. This indicates that the author(s) of the manuscript used AI-based Chatbot. Below the inscription, A, B, C, etc. letters in alphabetical order indicate this contribution to the research.

The name of the chatbot/toolkit(s) in the Materials and Methods section; the author(s) can include the name of the chatbot developer in the Acknowledgments section.

Authors may disagree because using the logo looks like co-authoring with AI. In anticipation of this disagreement, we suggest looking at the actual capabilities of chatbots and their role in preparing the paper/chapter. After all, chatbots are quite capable of performing study design, data collection, statistical analysis, data interpretation, manuscript preparation, literature searches... The author only needs to specify the topic, key parameters, and manuscript design requirements, and that will be enough for chatbot to write a review article or even an original article.

We assume that in the near future, such papers/chapters will fill publishers' email inboxes. Therefore, the dilemma of quality or quantity in scientific scientific research and publications will become particularly relevant (Melnyk & Pypenko, 2021).

Are the papers written by chatbots the result of the intellectual activity of the author, who has skillfully set the parameters for entering information, or are they still the product of the chatbot, which has a share in co-authorship?

Let us try to answer the question of who owns the authorship of such a publication objectively.

Despite the significant contributions that chatbots can make, at this stage chatbots cannot be considered legitimate authors of a scientific papers.

If only because chatbots are not responsible for the text they write, they cannot sign a statement about the presence or absence of a conflict of interest. Such a statement is required by most scientific journals, including the International Journal of Science Annals (IJSA).

However, there is a precedent of ChatGPT having a profile in Scopus (ChatGPT, n. d.), as well as papers published by prestigious international publishers in which ChatGPT is listed as an author (O'Connor & ChatGPT, 2022).

Also noteworthy is the book "Impromptu: Amplifying Our Humanity through AI", in which GPT-4 writes: "I would like to thank Reid Hoffman for inviting me to co-author this book with him". Please note that Reid Hoffman, a leader in the field of AI, states on the title page "By Reid Hoffman with GPT-4" (Hoffman with GPT-4, 2023).

There is one case in the literature where ChatGPT has answered negatively to the question of whether it meets all of the International Committee of Medical Journal Editors (ICMJE) criteria for authorship – "ChatGPT can assist in the drafting or revising of a work, but it cannot fulfill all of the ICMJE criteria for authorship" (Anderson, 2023).

Perhaps it is a question of specific criteria for authorship, rather than ChatGPT's refusal to acknowledge its role in writing. In any case, we have not received a clear answer to this question. Therefore, the answer should be sought in the aspect of ethics, as well as the willingness of the person to recognize the authorship of ChatGPT or not.

Todd Carpenter conducted a ChatGPT survey on the impact of AI on science communication. Specifically, he asked about the ethics for an author of using AI in developing a scholarly paper. As ChatGPT learned from the response, ethics “depends on the specific context and the expectations of the research community in which the article will be published” (Carpenter, 2023).

ChatGPT itself sees no ethical problems with the use of AI in scientific writing. However, it notes that authors must “clearly state this in the article and provide appropriate credit to the AI program” (Carpenter, 2023).

Springer Nature and Taylor & Francis Publishers suggest that AI contributions should be reflected in the methods or acknowledgements section, rather than being listed as an author (Stokel-Walker, 2023).

This position is justified by the important characteristic of authorship – responsibility for publication.

In this context, it should be noted that it is known that AI has convincingly described the results of studies (specifying the organizations that conducted them and the quantitative indicators). However, when clarifying the information, he could not confirm it with any sources and apologized for the error and confusion in his statement (Davis, 2023).

These facts point to the need for caution and responsible use of information obtained from AI. It is important to remember that human remains responsible and accountable for copyright infringement.

If someone claims undivided authorship, he/she should objectively, based on facts, state the role of chatbot in the scientific research and publications, claim full responsibility for the content of his/her manuscript and the result, including the parts created by chatbots, as well as the degree of originality of his/her publication. Perhaps there is no shame in stating that the research design, data collection, or statistical analysis was done using a particular chatbot. In doing so, the question posed to the chatbot and the answer received from the chatbot should be clearly stated.

We believe that information about the use of chatbot should necessarily be reflected in the methodology with a correct indication of which chatbot was used by the author, where and to what extent. The name of chatbot and its characteristics should be specified in the References list.

Our recommendation is also based on the fact that in the near future it will probably be impossible to hide the involvement of chatbots in the writing of a scientific paper/chapter. Chatbots-creating companies will start using something like a “watermarking” on the bot’s output to make plagiarism easier to spot. The San Francisco-based company OpenAI, which created ChatGPT, has already announced this. OpenAI guest researcher Scott Aaronson said that “the technology would work by subtly tweaking the specific choice of words selected by ChatGPT,

…, in a way that wouldn't be noticeable to a reader, but would be statistically predictable to anyone looking for signs of machine-generated text" (Hern, 2022).

So there is a good chance that if you try to pretend to be the author of text written by a chatbot, you may be detected. Turnitin has already begun work on developing an AI-based text detection tool (Chechitelli, 2023).

In early April 2023, the American Psychological Association (APA) website published information with guidelines for quoting and reproducing text generated by chatbots (McAdoo, 2023).

We recommend that Authors of our Journal use these standards when preparing a manuscript and citing text generated by chatbots.

It is important to note the statement of the Committee of Publication Ethics (COPE). On its website, the Committee has published its official position on authorship and the use of AI tools (COPE Council, 2021; COPE, 30 January 2023; COPE, 13 February 2023; COPE, 23 February 2023; Watson & Stiglic, 2023). Also a number of papers on using AI for scientific writing (Çalli & Çalli, 2022; Dans, 2019; Dimitriadou & Lanitis, 2023; Farahani, 2023; Singh & Sood, 2022).

Today, COPE is virtually the only organization in the scientific world that promotes ethical principles in scientific publishing. COPE Council members warn that the increasing role of AI in research writing "has significant implications for research integrity and the need for improved means and tools to detect fraudulent research" (COPE, 23 March 2023).

This is a matter of concern for those scientific publishers, who conduct their activities responsibly and put into practice the principles of scientific publishing ethics and the COPE standards.

The IJSA is a full member of the COPE (COPE, n.d.). Thanks to this, the members of the IJSA Editorial Board were able to participate online in events dedicated to the discussion of this topical issue (COPE, 23 March 2023).

The Regulation of the Norms of Interaction and Relations between Humans and AI

Our entire civilisation, the achievements of science and culture, have been created by human intelligence. However, we now have artificial intelligence (AI) that could be its alternative. This situation has actualised some of the most important questions about the relationship between human intelligence and artificial intelligence. Firstly, will AI help us or, on the contrary, create problems? Secondly, what do we need to do to create a harmonious system of interacting and relating?

Human civilisation has entered a new spiral of development in the age of digital and information technology where, with the advent of AI, a new "Human-AI System" of relationships has emerged (Melnyk & Pypenko, 2023). This allows us to clarify the essential features of the new phenomenon under consideration, which opens prospects for its further study.

First of all, we should accept as axiomatic the idea that our world has been changed forever with the advent of AI. Whatever we do, there will always be a place for AI in what we do. In addition, the role of AI in our lives will continue to grow. It is still within our power to control and manage this system of interactions. However, the potential and the speed of development of AI-based digital and information technologies are so great that we may have to concede this primacy in the near future.

It has been less than a year (30 November 2022) since the launch of ChatGPT. ChatGPT is an AI-based conversational LLM. The potential applications of LLMs in research and practice look promising, given their ability to generate creative responses.

In the first 3 months of its existence, ChatGPT has become an indispensable tool for 100 million people worldwide. A large number of people of different ages and social statuses, from schoolchildren to university professors, have found ChatGPT to be an indispensable tool for dealing with issues in their personal and professional lives.

This popularity makes ChatGPT an obvious positive answer to the question of whether AI has become our assistant. We are sure that there will be millions of schoolchildren and students who actively use ChatGPT for their studies and for solving tasks assigned to them in educational institutions. At the same time, it is very likely that millions of teachers and university professors are also using AI to prepare assignments for these students.

This creates a paradoxical situation in which the AI becomes both the object and the subject of action (writing and solving its own tasks).

The other question is whether this is a problem or not. As in the first case, we believe that the answer to this question will be in the affirmative. Undoubtedly, replacing one's own opinion and efforts in solving tasks with an AI answer will have a negative impact on students' personal cognitive sphere (intelligence) and competence level.

To be fair, we should point out that this is a problem for the faculty as well. Over the past year, there has been a significant increase in the number of research studies, and therefore articles, using AI-based tools. Previous studies have addressed the legitimacy of using AI in scientific research and publications (Melnyk & Pypenko, 2023), and the dilemma of quality versus quantity of scientific publications, which will become particularly relevant with the advent of AI (Melnyk & Pypenko, 2021). Discussions about the tendency to replace humans with AI, and the potential threats associated with this, have been ongoing since the term was introduced by McCarthy (1959) in the middle of the last century. These issues certainly deserve attention. In most cases, they remained theoretical views of the problem. However, the situation has changed dramatically over the past year.

That is why we are focusing on the above axiom about the irreversible penetration of AI into our life activities and the subsequent increase in its influence on all spheres. As a consequence of this trend, the need to build a real system of harmonious interaction and relationship between humans and AI becomes obvious.

This problem is likely to be a key issue for this century, as the survival of humanity literally depends on it.

We are not inclined to dramatise the situation about the increasing danger to humanity from the development of AI. We believe that AI, in the absence of individual consciousness, is not capable of harming humanity. However, the real dangers, which are becoming increasingly apparent, should not be ignored.

In a metaphorical sense, AI can be compared to the fuel or electricity needed to run a machine. The advent of a new fuel (petrol) made it possible for the internal combustion engine to function. Automobiles appeared, aeroplanes... Even today, many people still measure the power of a car's engine in horsepower. Nowadays, hardly anyone has to do their travel planning with horses in mind. But this does not mean that horses have become useless and can be disparaged as unnecessary or inefficient.

It is still directly human beings who decide how to use and interact with new scientific advances. A human can refuel the drone and send it on a research mission to another planet, or send it to destroy the inhabitants of a neighbouring country. A clear example is the russian federation's military action in Ukraine. In this case, drones with integrated warheads are actively deployed in large numbers, capable of making a long flight over the battlefield, independently detecting a target, classifying its level of importance among others, and making a decision to destroy it.

Despite the negative trends and realities we live in today, there is still hope that humanity is able to understand the responsibility of using AI and can channel it to advance our civilisation, science and culture.

Therefore, the issue of creating a harmonious relationship between humans and AI is very important. These relationships can be both personal and professional. In this case, personal relationships, such as the role and the level, are determined by each person for him or herself; professional relationships can be regulated from the outside and have serious consequences for the human.

We share the views of researchers who claim that the use of AI will be the reason for the reduction of large numbers of workers in various fields in the coming years. It can cause various social conflicts.

It is therefore crucial to regulate these relationships in a legal and regulatory context.

We think this is difficult to achieve, but it is certainly possible. The challenge is that AI is becoming increasingly pervasive in people's daily lives and workspace. Therefore, something more sophisticated than Asimov's Three Laws of Robotics must be developed to manage this complex system of human-AI relationships (Asimov, 1942).

We believe that in the near future, countries with high levels of economic growth will develop, ratify and implement laws that regulate the norms of interaction and relationships between humans and AI.

Today, thanks to the activities of COPE (2023) and major scientific publishers (WAME, JAMA), standards and rules have been developed for the use of AI-based chatbots in scientific publications.

The first steps towards legitimising AI-based chatbots were taken by Melnyk and Pypenko (2022). These scientists have created and implemented the AIC AI Chatbots information technology platform (AIC AI Chatbots, 2023), which provides technological solutions for the use of AI-based Chatbots (text, images, videos) in scientific research and publications. However, the above standards are voluntary and could be used as a recommended guide. This allows unscrupulous users of AI-based chatbots to ignore these ethical guidelines. This is why it is necessary to enact laws that regulate the standards of human-AI interaction.

In developing laws and regulations governing standards for human-AI interaction, particular attention should be paid to the protection of human rights in the case of deliberate refusal to use AI.

We believe that human activity without the use of AI will soon have to defend its right to exist. It is a natural human right to freedom of choice and work. Using a special attribution (logo/stamp/label) on a product created by humans without AI involvement can help. We offer such an attribution "AI Free. Human Created" (Figure 2.2).

Figure 2.2

The Attribution "AI Free. Human Created"

Note. From "Human and artificial intelligence interaction", by I. Pypenko, 2023, *International Journal of Science Annals, 6*(2), p. 56 (https://doi.org/10.26697/ijsa.2023.2.7)

The attribution developed enables the classification of products created by humans without the use of AI, as well as increasing the value of natural human labour (Pypenko, 2023).

Conclusions

We started our Editorial with a warning: this chapter was not written by a chatbot and is intended for humans. Although we don't have the slightest doubt that it will be read by AI, because this chapter will be converted into multiple formats and found in several dozen scientometric databases, repositories, and search engines. It is time for humans to define the legitimacy we give to AI.

We have offered the essence of the definition "Human-AI System". This allows us to clarify the essential features of the new phenomenon under consideration, which opens prospects for its further study.

Authors should be transparent about the use of AI tools. This will allow readers to know what and how the chapter was created, and it will allow reviewers, editors, and publishers to check the quality of the chapter.

We encourage you to consult the recommendations of leading publishers Springer Nature and Taylor & Francis, as well as the expertise of COPE Council members on the ethics of scientific publication, and the recommendations of APA experts on citing and reproducing chatbot-generated text.

The need to determine the legitimacy of using AI-based chatbots in scientific research prompted us to develop a method for indicating AI involvement and the role of chatbots in a scientific publication.

We recommend using the developed base logo to indicate chatbots' involvement and contributions to the writing of the chapter. This would be appropriate for researchers, authors, reviewers, editors, readers, and, from our point of view, ethical.

AI has become an integral part of the lives of human beings. The potential and the speed of development of AI-based information technologies is so great that in the near future humanity may concede primacy to AI. This situation requires the development, ratification and implementation of laws that regulate the norms of interaction and relationships between humans and AI.

The first steps have already been taken to legitimise AI-based chatbots in scientific research and publications. This chapter offers an attribution for products created by humans without the involvement of AI. The use of the "AI Free. Human Created" attribution helps to protect the individual's right to freedom of choice and work.

References

AIC AI Chatbots. (2023). *AIC AI Chatbots attribution.* https://doi.org/10.26697/ai.chatbots

Anderson, K. (2023, January 13). *ChatGPT says it's not an author.* The Geyser. https://www.the-geyser.com/chatgpt-says-its-not-an-author/

Asimov, I. (1942). *Runaround.* Astounding Science Fiction. https://web.williams.edu/Mathematics/sjmiller/public_html/105Sp10/handouts/Runaround.html

Çalli, B. A., & Çalli, L. (2022). Understanding the utilization of artificial intelligence and robotics in the service sector. In S. B. Kahyaoğlu (Ed.), *The Impact of Artificial Intelligence on Governance, Economics and Finance: Vol. 2. Accounting, Finance, Sustainability, Governance & Fraud: Theory and Application* (pp. 243-263). Springer. https://doi.org/10.1007/978-981-16-8997-0_14

Carpenter, T. A. (2023, January 11). *Thoughts on AI's impact on scholarly communications? An interview with ChatGPT.* The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2023/01/11/chatgpt-thoughts-on-ais-impact-on-scholarly-communications/

ChatGPT. (n. d.). *Scopus Author ID: 58024851600* [Scopus Author Identifier]. Scopus. Retrieved April 01, 2023, from https://www.scopus.com/authid/detail.uri?authorId=58024851600&ref=the-geyser.com

Chechitelli, A. (2023, January 13). *Sneak preview of Turnitin's AI writing and ChatGPT detection capability*. Turnitin. https://www.turnitin.com/blog/sneak-preview-of-turnitins-ai-writing-and-chatgpt-detection-capability

COPE. (2023, January 30). *Artificial intelligence in the news.* https://publicationethics.org/news/artificial-intelligence-news

COPE. (2023, February 13). *Authorship and AI tools. COPE position statement.* https://publicationethics.org/cope-position-statements/ai-author

COPE. (2023, February 23). *Artificial intelligence and authorship.* https://publicationethics.org/news/artificial-intelligence-and-authorship

COPE. (2023, March 23). *Artificial intelligence (AI) and fake papers*. https://publicationethics.org/resources/forum-discussions/artificial-intelligence-fake-paper

COPE Council. (2021, September). *COPE Discussion document: Artificial intelligence (AI) in decision making – English*. https://doi.org/10.24318/9kvAgrnJ

COPE. (n. d.). *International Journal of Science Annals* [COPE Members page]. COPE. Retrieved March 17, 2023, from https://publicationethics.org/members/international-journal-science-annals

Dans, E. (2019, February 6). *Meet Bertie, Heliograf and Cyborg, the new journalists on the block.* Forbes. https://www.forbes.com/sites/enriquedans/2019/02/06/meet-bertie-heliograf-and-cyborg-the-new-journalists-on-the-block/?sh=416c2163138d

Davis, P. (2023, January 13) *Did ChatGPT just lie to me?* The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2023/01/13/did-chatgpt-just-lie-to-me/

Dimitriadou, E., & Lanitis, A. (2023). A critical evaluation, challenges, and future perspectives of using artificial intelligence and emerging technologies in smart classrooms. *Smart Learning Environments, 10,* 12. https://doi.org/10.1186/s40561-023-00231-3

Farahani, M. S. (2023). Applications of artificial intelligence in social science issues: a case study on predicting population change. *Journal of the Knowledge Economy*. https://doi.org/10.1007/s13132-023-01270-4

Hern, A. (2022, December 31). *AI-assisted plagiarism? ChatGPT bot says it has an answer for that.* The Guardian. https://amp.theguardian.com/technology/2022/dec/31/ai-assisted-plagiarism-chatgpt-bot-says-it-has-an-answer-for-that

Hoffman, R. with GPT-4. (2023). *Impromptu: Amplifying our humanity through AI.* Dallepedia LLC. https://www.impromptubook.com/wp-content/uploads/2023/03/impromptu-rh.pdf

McAdoo, T. (2023, April 7). *How to cite ChatGPT*. APA. https://apastyle.apa.org/blog/how-to-cite-chatgpt

McCarthy, J. (1959). Programs with common sense. In *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, 756-791. Her Majesty’s Stationery Office. http://jmc.stanford.edu/articles/mcc59/mcc59.pdf

Melnyk, Yu. B., & Pypenko, I. S. (2021). Dilemma: Quality or quantity in scientific periodical publishing. *International Journal of Science Annals, 4*(2), 5-7. https://doi.org/10.26697/ijsa.2021.2.1

Melnyk, Yu. B., & Pypenko, I. S. (2023). The legitimacy of artificial intelligence and the role of ChatBots in scientific publications. *International Journal of Science Annals, 6*(1), 5–10. https://doi.org/10.26697/ijsa.2023.1.1

O’Connor, S., & ChatGPT. (2022). Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Education in Practice, 66,* 103537. https://doi.org/10.1016/j.nepr.2022.103537

Pypenko, I. S. (2023). Human and artificial intelligence interaction. *International Journal of Science Annals,* *6*(2), 54–56. https://doi.org/10.26697/ijsa.2023.2.7

Singh, R., & Sood, M. (2022). An introductory note on the pros and cons of using artificial intelligence for cybersecurity. In D. Gupta, A. Khanna, S. Bhattacharyya, A. E. Hassanien, S. Anand, A. Jaiswal (Eds.), *International Conference on Innovative Computing and Communications: Vol. 471. Lecture Notes in Networks and Systems* (pp. 337-348). Springer. https://doi.org/10.1007/978-981-19-2535-1_26

Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove. *Nature, 613,* 620-621. https://doi.org/10.1038/d41586-023-00107-z

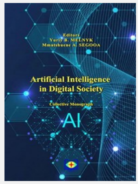
Watson, R., & Stiglic, G. (2023, February 23). *Guest editorial: The challenge of AI chatbots for journal editors.* https://publicationethics.org/news/challenge-ai-chatbots-journal-editors

Information about the authors:

Pypenko Iryna Sergiivna – https://orcid.org/0000-0001-5083-540X; Doctor of Philosophy in Economics, Affiliated Associate Professor, Secretary of Board, Kharkiv Regional Public Organization “Culture of Health”; Scientific Research Institute KRPOCH, Ukraine.

Melnyk Yuriy Borysovych – https://orcid.org/0000-0002-8527-4638; Doctor of Philosophy in Pedagogy, Affiliated Associate Professor; Chairman of Board, Kharkiv Regional Public Organization “Culture of Health” (KRPOCH); Director, Scientific Research Institute KRPOCH, Ukraine.

Artificial Intelligence in Digital Society,
Volume 1, 2026
DOI: 10.26697/9786177089192.2026
ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)



Chapter 3. Bridging the Society-Artificial Intelligence Gap through Holistic Digital Transformation

Baduza G. [1], **Penxa L.** [2], **Ramafi P.** [3]

[1] *Rhodes University, South Africa*
[2] *University of the Western Cape, South Africa*
[3] *University of Witwatersrand, South Africa*

Received: *14.12.2025;* ***Accepted:*** *10.02.2026;* ***Published:*** *10.03.2026*

Abstract
A widening gap exists between artificial intelligence's (AI) rapid advancement and society's capacity to govern and benefit equitably from these technologies. AI adoption is treated as technical implementation rather than comprehensive socio-technical transformation, creating dangerous misalignments between technological capabilities and societal readiness. This chapter examines the multifaceted relationship between digital transformation, artificial intelligence, and societal change, analysing how technological advancement reshapes social institutions, governance structures, and human relationships. Through qualitative documentary research and comparative case study analysis across healthcare, finance, education, and public services, the chapter explores both the enabling potential and adverse consequences of AI-driven transformation. The analysis reveals that while AI acts as a catalyst for innovation in healthcare diagnostics, precision agriculture, circular economy practices, and educational personalization, it simultaneously introduces critical challenges including labour displacement, wealth inequality, algorithmic bias, and threats to human agency. Drawing on Vial's Building Blocks of Digital Transformation framework, four key themes emerge: foundational integration infrastructure, equitable value distribution, trustworthy organizational practices, and societal impact priority. The chapter demonstrates that successful AI integration requires moving beyond purely technological metrics toward human-centred approaches that prioritize transparency, trust, fairness, and environmental sustainability.
Keywords: digital transformation, artificial intelligence, society 5.0, human-centred technology.

Cite this chapter as:
Baduza, G., Penxa, L., & Ramafi, P. (2026). Bridging the society-artificial intelligence gap through holistic digital transformation. In Y. B. Melnyk & M. A. Segooa (Eds.), *Artificial Intelligence in Digital Society, Vol. 1.* (pp. 42–54). KRPOCH. https://doi.org/10.26697/aids.2026.3
The electronic version of this chapter is complete. It can be found online in the AIDS Archive https://doi.org/10.26697/aids

Introduction

The concept of digital transformation (DT) was introduced in 2000 (Patel & McCarthy, 2000) but it became popular to researchers and practitioners after 2014 (Reis, Amorim, Melão & Matos, 2018). Most DT definitions shows that it is techno-social changes in institutions and their environments resulting from the adoption and use of new digital technologies in societies (Stolterman et al., 2004; Fitzgerald et al., 2014; Kraus, Jones, Kailer, Weinmann, Chaparro-Banegas & Roig-Tierno, 2021; Vial, 2019, and Tana, Breidbach, & Burton-Jones, 2023). Moreover, successful DT must consider factors that can hinder the execution of their transformation (Vial, 2019).

DT in society must consider several factors including social inequality and digital divide. The debate about social inequality and the digital divide began in the 1990s focusing on access to computers and the internet. In the 2000s, the debates focused on access to the internet and its impact on people's lives (Matzat & 2020). The digital divide is still prevalent in many social institutions (Tang et al., 2025). Modern digital technologies have transformed the way people socialise, communicate, work, learn, entertain, and share their emotions and expressions. It has influenced and shaped other social institutions resulting in massive transformations in the public, private and civil society sectors. This leads to the debate of how to digitally transform society and its institutions.

Literature Review

Digitally Transforming Societies

Digitally transforming societies requires a collaborative effort from different relevant actors have shared values and norms, pursue a joint objective (Tana, Breidbach & Burton-Jones, 2023). In Africa, DT was set to impact all areas of African society (Kazim, 2021). DT in Africa needs to consider heterogeneity of African countries and must be context specific. There are technological and economic challenges which include digital infrastructure, inadequate internet connections, digital skills, affordability of digital services, regional integration of digital infrastructure, etc. (Kazim, 2021). These challenges reveal the degree to which African countries can effectively digitally transform their societies (African Union, 2020).

The Covid 19 pandemic contributed to fast-tracking DT in all countries of the world. In Higher Education (HE) institutions, universities had to embrace distance teaching and learning using online platforms (Mhlanga et al., 2022 Pypenko et al., 2020). They had to leverage existing and acquired new digital technologies. However, they exposed a digital divide in terms of staff digital incapacities, lack of internet infrastructure, spatial distribution of internet facilities and digital literacy (Zeleza & Okanda, 2021). This suggested that DT in HE institutions requires digital capacities and skills of the relevant actors.

DT has a potential to address government administrative inefficiencies and enhance public services. DT can improve government decision-making efficiencies, optimize resource allocation, and enhance the quality and efficiency of public services (Yang, Gu & Albitar, 2024). DT serves to enhance government transparency, enabling the public to hold government accountable (Stratu-Strelet et al., 2023).

Consequently, the construction of a digital government is an integral pathway to satisfying public expectations and boosting public trust (Yang, et al., 2024). Moreover, both Europe and Asia Pacific countries are actively promoting the DT of their governments. The eGovernment Benchmark report of 2022 shows that 35 countries in Europe are providing and promoting eGovernment services. Asia Pacific countries have also made significant progress in this area (Priharsari et al., 2023). Additionally, China's rapid economic expansion serves as an influential catalyst for governmental DT.

Negative Impacts of AI Adoption on Society

DT threatens traditional values and cultures while reshaping how people socialise, communicate, learn, and work, with implications for privacy and security. The growing use of Artificial Intelligence (AI) is transforming contemporary society by altering power structures through decentralisation and the emergence of new social classes (Gutorovich & Gutorovich, 2019), as well as increasing social connectivity and enabling virtual socialisation (Caceres Zapatero et al., 2017 cited in Hanandini, 2024).

DT has enhanced communication and human interaction through widespread use of media and messaging platforms (Carter, 2005) enabling online socialisation, dating, business engagement, and long-term relationships (Guzman & Lewis, 2020). However, it has also reduced face-to-face interaction, increased cyberbullying, online deception, and predatory behaviour (Kumari & Oman, 2024).

DT can exacerbate inequality, weaken security, and threaten privacy, thereby affecting human agency and human rights (Kumari & Oman, 2024). It also contributes to unemployment by reshaping job types toward technology-oriented roles and disrupting existing skills, making technology literacy, cognitive problem-solving, and analytical thinking increasingly essential (World Economic Forum, 2023).

DT has reshaped how knowledge is created and accessed (Melnyk & Pypenko, 2020; Mhlanga, 2024). It has also transformed formal education by expanding access through e-learning and enabling more personalized learning experiences.

According to Taufik (2025), DT has expanded to include AI, bringing ethical and existential risks such as algorithmic bias and the potential erosion of human critical thinking (Makridakis, 2017; Farina et al., 2022). This shift requires a human-centred approach that balances innovation with transparency, trust, and the preservation of social values.

Digital Transformation and Ethics

DT poses important ethical concerns about equity, privacy, and responsible technology use (Klein, 2022; Schuster & Kilov, 2025). Bias and the lack of fairness of the AI systems have additionally cast doubt as such systems are not culturally, politically, or morally neutral (Schuster & Kilov, 2025; Stapleton, 2025). They embody human biases that are unconsciously programmed into them and can relentlessly target the most vulnerable (Stapleton, 2025). AI systems reflect to us our mistakes, problems, errors, biases, prejudices and failures of wisdom (Stapleton, 2025). When automated systems produce correct outcomes rapidly, humans, risk acting merely as validators of machine-generated decisions rather than informed agents, which can erode epistemic autonomy and human judgment (Bokhari, Park and Manzoor, 2025; Stapleton, 2025). In fields like social robotics, the inability of robot friends to mimic the complex styles of human friendship (such as being constructively critical) raises ethical concerns, as this relationship may gradually contribute to a loss of important societal values like honesty and respect (Farina et al., 2022).

AI as a Tool to Enable Holistic Digital Transformation

AI acts as a catalyst for new skills, institutions, and governance by augmenting and automating human cognitive tasks, thereby enabling societal and economic transformation (Makridakis, 2017). AI drives value creation by enhancing decision-making through big data analytics, automation, and predictive capabilities across sectors (Foresti et al., 2020; Feroz & Kwak, 2024; Bokhari, Park & Manzoor, 2025).

AI has played a key role in advancing healthcare and precision medicine by supporting healthcare professionals with diagnostic insights, operational efficiency, and patient engagement. It enhances hospital logistics, resource allocation, real-time patient monitoring through wearables, and continuous support via virtual nursing assistants (Varnosfaderani & Forouzanfar, 2024).

AI is emerging as a critical enabler of sustainability and the circular economy by supporting the achievement of the Sustainable Development Goals (SDGs). In agriculture, machine learning-enabled drones and satellites enhance productivity and food security by monitoring soil conditions and predicting environmental effects on crops, while AI-based digital twins allow organisations to optimise energy consumption and significantly reduce carbon emissions (Ali et al., 2024; Varnosfaderani and Forouzanfar, 2024).

In education, AI is redefining learning by promoting autonomous and self-regulated approaches that empower students to take ownership of their educational journeys. Through prompt engineering, students use AI as a conversational and intellectual partner for idea generation and research refinement, while automated assessment tools provide instant, human-like feedback that supports real-time learning and faster mastery of complex concepts (Mzwri and Turcsányi-Szabo, 2025).

Essentially, AI functions as an enabling infrastructure that enhances human capabilities by processing complex data, revealing hidden patterns, and delivering actionable insights. This supports new forms of institutional practice and governance while expanding access to advanced analytics that promote progress toward sustainable development goals.

Synergy and Integration between the Future Society and AI

The synergy and integration of AI and future society reflect a shift toward a human-centric model aligned with Society 5.0, which merges physical and digital spaces through Human-Cyber-Physical Systems to address social challenges and enhance human well-being (Foresti et al., 2020). This smart society envisions AI enabling sustainability through human–machine cooperation, predictive and adaptive systems, and organisational strategies of "Digitalization" that prioritise value creation, augmentation, and operational excellence over simple automation (Feroz &Kwak, 2024; Foresti et al., 2020).

The success of digital government transformation depends on stakeholder trust acting as a bridge between technology and institutional change, with effective integration enhancing public value through fairness, inclusivity, and transparency (Bokhari, Park &Manzoor, 2025). AI and Digital Twins serve as essential enablers for transitioning to a Circular Economy, particularly through closing material loops by tracking and mapping resources to achieve UN SDGs (Ali et al., 2024). This includes AI-driven drones and sensors that monitor environmental impacts in real-time to facilitate smart and sustainable agriculture that protects biodiversity (Ali et al., 2024). Long-term synergy may even involve incorporating AI into democratic political institutions in ways that reduce conflict and enhance governance effectiveness.

The deepest level of integration involves the concept of the "generated human," where human judgment aligns with algorithmic language while preserving cognitive independence, ensuring humans remain informed agents rather than mere validators of machine-generated decisions (Branda, 2025). This requires reflective empowerment, in which AI enhances human reflection and agency instead of undermining it (Branda, 2025; Farina et al., 2022). Integration for social and moral good is guided by a virtue ethics approach centred on human flourishing and the development of techno-moral wisdom, enabling AI to support meaningful lives within communities through flexible, context-sensitive ethical engagement rather than rigid rule-based frameworks (Farina et al., 2022).

Methodology

This chapter adopted a qualitative documentary research approach as a methodological framework. Data collection was done through the peer reviewed literature review and case studies. The literature review examines peer-reviewed papers on DT, organizational studies, public governance and ethics. The comparative case studies analysed AI implementation initiatives in healthcare, finance, education, and public services and society, identifying patterns distinguishing successful holistic transformations from fragmented implementations. The collected data followed the framework-based thematic analysis. Framework-based analysis applies Vial's model ("Building Blocks of Digital Transformation") to categorize findings across triggers, barriers, strategic responses, and outcomes, revealing how dimensions interact during AI adoption.

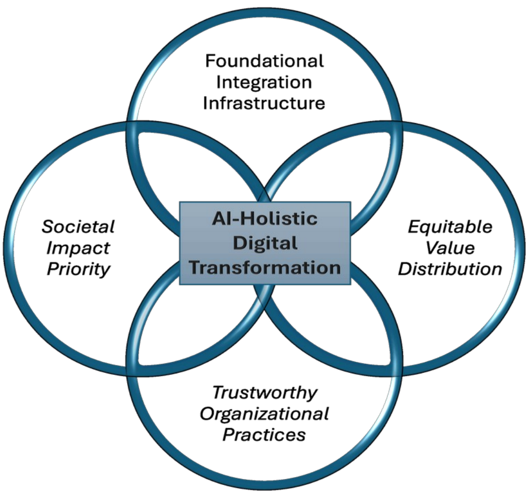
Findings

The section presents findings that emerged from case study analysis through Vial's model ("Building Blocks of Digital Transformation"). Various themes emerged from the analysed case studies. Some of the case studies were focused on industrial companies and how they had aimed at integrating machine learning and deep learning into their business development, focusing on telecommunications, automotive, packaging, pumps, and an AI platform provider.

The next set of case studies proved that agricultural and agri-tech companies use artificial intelligence and digital twins to support circular economy practices and multiple UN Sustainable Development Goals. These firms deploy AI and DT for precision agriculture, waste minimization, water conservation, renewable energy integration, and resource recovery, thereby operationalizing strategies like narrowing, slowing, closing, and regenerating resource loops. The cases highlight that, despite barriers such as data, cost, and change-management challenges, AI and DT enabled solutions have already helped participating companies contribute to several SDGs. Based on these case studies and the analysis utilizing the select theoretical lens the following key themes emanated as per Figure 3.1.

Figure 3.1

AI-Holistic Digital Transformation Themes



Theme 1: Foundational Integration Infrastructure

The first theme emerging from these case studies reveal that successful AI integration requires fundamental building blocks that span technical infrastructure, human capacity, and institutional readiness. Through the case studies, a key element was how across embedded systems, microgrids, and agriculture, AI acts as a disruptive technology that only creates value when tied to clear business and societal goals (Feroz & Kwak, 2024; John et al., 2022). These foundational elements include the development of robust data ecosystems, the cultivation of digital literacy and prompt engineering skills among users, the establishment of adaptive governance frameworks (John et al., 2022), and the creation of interoperable systems that enable seamless human-machine collaboration across diverse societal contexts (Varnosfaderani & Forouzanfar, 2024; (Bokhari et al., 2025). This implies that, at a societal level, AI should be framed as an instrument for public-interest outcomes (e.g., safety, sustainability, resilience), rather than as an end or a pure efficiency play.

Theme 2: Equitable Value Distribution

This theme emphasizes that the extraction of insights and generation of value from data cannot be pursued in isolation from considerations of who has access to AI systems. This entails understanding whose data is being used, and whether algorithmic outcomes maintain or reduce existing social inequalities therefore, requiring deliberate design choices that prioritize inclusive participation and equitable distribution of AI-generated benefits (Fox & Griffy-Brown, 2022; Lucchi, 2023). Data as a key enabler in utilising AI, there is continued need to identify where the data is emanating from through new value paths (Lucchi, 2023). Through the case studies it has been identified these paths include predictive maintenance, AI-designed microgrids emission cuts, and AI/DT-enabled circular agriculture which in turn advances multiple SDGs (Ali et al., 2024; Alam et al., 2025). At the same time, they expose that high data and infrastructure demands can exclude smaller actors, regions with weaker connectivity, and less digitized farmers or firms (Ali et al., 2024). For society, this means AI needs governance and business models that share data benefits more broadly through fair access, capacity building, and safeguards against concentration of data and platform power. Data-driven value creation through AI must be fundamentally aligned with principles of fairness and inclusion to ensure equitable benefits across society.

Theme 3: Trustworthy Organizational Practices

This theme highlights that organizations need to develop institutional competencies that ensure AI systems are trustworthy. Through explainable decisions, reliable performance across conditions, and ethical frameworks guiding development and deployment. These capabilities are now as essential to organizational excellence as technical skills. Case studies show that many organizations introduced structural changes and new capabilities to support AI adoption. They invest in DevOps/DataOps/MLOps, cross-functional AI teams, and continuous monitoring, but still face challenges with transparency, training–serving skew, and model drift, particularly in safety-critical areas such as autonomous driving and energy systems. For society, AI needs to be developed with built-in explainability, robustness checks, and clear accountability so people can trust systems that increasingly affect mobility, infrastructure, and resource allocation (Holmstrum, 2021; Varnosfaderani & Forouzanfar, 2024). Organizational capabilities for AI deployment must intrinsically embed ethics, reliability, and transparency as core operational principles rather than peripheral concerns (Makridakis, 2017; Feroz & Kwak, 2024; Vial, 2019).

Theme 4: Societal Impact Priority

This theme emphasises that the true measure of AI success lies in its contribution to human development, environmental sustainability, democratic governance, and social cohesion. These require that technological advancement be evaluated not merely by what it can do, but by whether it genuinely enhances public value and supports the achievement of broader societal goals such as the SDGs (Ali et al., 2024; Alam et al., 2025). For AI to serve society, evaluation metrics need to move beyond accuracy and Return of Investment (ROI) to include distributional effects, ecological impact, and long-term resilience (Makridakis, 2017; Holmstrom, 2021; John et al., 2022). These criteria feed back into how disruptions are sensed, strategies chosen, and structures redesigned. Societal outcomes must be elevated based on people-success criteria in evaluating AI implementation, moving beyond narrow metrics of efficiency or profitability (Farina et al., 2022; Branda, 2025).

Discussion

The key findings reveal that holistic DT requires an integration of foundational infrastructure, equitable value distribution, ethical organisational practices. Tana, Breidbach & Burton-Jones (2023) argue that this integration requires a collaboration with different actors with a shared goal. We argue that AI requires the collaboration between the public and private sectors and civil society to help navigate the interplay between technological advancements, social implications, environmental concerns, ethical considerations.

The findings from case studies revealed that successful AI integration requires fundamental building blocks that span technical infrastructure, human capacity, and institutional readiness. These findings respond to the existing technological challenges outlined in the literature such digital infrastructure, digital skills, affordability of digital services and inequality (Kazim, 2021).

Theme two above is equitable value distribution which argues that AI use must be aligned with principles of fairness and inclusion to ensure equitable benefits across society. Similarly, the literature showed that DT poses important ethical concerns about equity (Klein, 2022; Schuster & Kilov, 2025). Additionally, equitable value distribution theme responds to the existing challenges that AI systems reflect biasness and lack of fairness highlighted in the literature by Schuster & Kilov (2025) and Stapleton (2025).

Theme three argues for trustworthy organizational practices. This ethical issue aligns with Klein's (2022) argument that DT requires responsible technology use. Additional to trustworthy organizational practices, Bokhari et al. (2025) argued that stakeholder trust is key to integrate technology to organisational changes leading to successful DT.

Theme four is about societal impact as a priority. The theme argues for AI to contribute to people, environment, government and inclusive society. In alignment with this theme, the literature showed that AI has transformed the way people socialise, communicate, work, learn, entertain, and share their emotions and expressions (Tang et al., 2025). Similar to the theme arguing for AI to contribute to government, Yang et al. (2024) contends that AI can improve government decision-making efficiencies, optimize resource allocation, and enhance the quality and efficiency of public services.

Concluding Remarks

Artificial intelligence is reshaping social values, cultural norms, relationships, and the physical environment, while also intensifying social inequality and the digital divide. This chapter has shown that AI and DT are fundamentally socio-technical processes that affect institutions, human agency, and social cohesion. While AI offers significant opportunities such as improved healthcare, sustainable agriculture, personalised education, and more transparent governance, these benefits are unevenly distributed. In regions such as Africa, infrastructure gaps, limited digital literacy, labour displacement, algorithmic bias, and privacy risks threaten to deepen existing inequalities and undermine democratic trust. AI can make a meaningful contribution only if it enhances human well-being, supports environmental sustainability, strengthens democratic governance, and advances broader societal goals such as the Sustainable Development Goals. Achieving this requires collaborative, human-centred governance that balances innovation with ethical responsibility, ensuring that technological progress ultimately serves social justice and shared human values.

References

African Union (AU). (2020). *The African Union (AU) joins forces with HP INC to expand digital learning opportunities for African youth.* https://au.int/sites/default/files/pressreleases/38479-pr-pr_043_-_au_joins_forces_with_hp_inc_to_expand_digital_learning_opportunities_for_african_youth.pdf

Alam, M. M., Hossain, M. J., Zamee, M. A., & Al-Durra, A. (2025). Design and operation of future low-voltage community microgrids: An AI-based approach with real case study. *Applied Energy, 377*, Article 124523. https://doi.org/10.1016/j.apenergy.2024.124523

Ali, Z. A., Zain, M., Hasan, R., Al Salman, H., Alkhamees, B. F., & Almisned, F. A. (2024). Circular economy advances with artificial intelligence and digital twin: Multiple-case study of Chinese industries in agriculture. *Journal of the Knowledge Economy, 16*(1), 2192–2228. https://doi.org/10.1007/s13132-024-02101-w

Bokhari, S. A. A., Park, S. Y., & Manzoor, S. (2025). Digital government transformation through artificial intelligence: The mediating role of stakeholder trust and participation. *Digital, 5*(3), Article 43. https://doi.org/10.3390/digital5030043

Branda, F. (2025). Generated humans, lost judgment: Rethinking knowledge with AI. *AI & Society, 40*(8). https://doi.org/10.1007/s00146-025-02748-2

Carter, D. (2005). Living in virtual communities: An ethnography of human relationships in cyberspace. *Information, Community & Society, 8*(2), 148–167. https://doi.org/10.1080/13691180500146235

Farina, M., Zhdanov, P., Karimov, A., & Lavazza, A. (2022). AI and society: A virtue ethics approach. *AI & Society, 39*(3), 1127–1140. https://doi.org/10.1007/s00146-022-01545-5

Feroz, K., & Kwak, M. (2024). Digital transformation (DT) and artificial intelligence (AI) convergence in organizations. *Journal of Computer Information Systems,* 1–17. https://doi.org/10.1080/08874417.2024.2424372

Fitzgerald, M., Kruschwitz, N., Bonnet, D., & Welch, M. (2014). Embracing digital technology: A new strategic imperative. *MIT Sloan Management Review, 55*(2), 1-12. https://sloanreview.mit.edu/issue/

Foresti, R., Rossi, S., Magnani, M., Guarino Lo Bianco, C., & Delmonte, N. (2020). Smart society and artificial intelligence: Big data scheduling and the global standard method applied to smart maintenance. *Engineering, 6*(7), 835–846. https://doi.org/10.1016/j.eng.2019.11.014

Fox, S., & Griffy-Brown, C. (2022). Artificial intelligence in society: Technology in Society briefing. *Technology in Society, 71,* Article 102130. https://doi.org/10.1016/j.techsoc.2022.102130

Gutorovich, O. V., & Gutorovich, V. N. (2019). Consequences of IT transformations. *Discourse, 5*(4), 42–52. https://philpapers.org/rec/GUTCOI-2

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society, 22*(1), 70–86. https://doi.org/10.1177/1461444819858691

Hanandini, D. (2024). Social transformation in modern society: A literature review on the role of technology in social interaction. *Jurnal Ilmiah Ekotrans & Erudisi, 4*(1), 82–95. https://doi.org/10.69989/j0m6cg84

Holmstrom, J. (2021). From AI to digital transformation: The AI readiness framework. *Business Horizons, 65*(3), 329–339. https://doi.org/10.1016/j.bushor.2021.03.006

John, M. M., Olsson, H. H., & Bosch, J. (2022). Towards an AI-driven business development framework: A multi-case study. *Journal of Software: Evolution and Process, 35*(6). https://doi.org/10.1002/smr.2432

Kazim, F. A. (2021). Digital transformation in communities of Africa. *International Journal of Digital Strategy, Governance, and Business Transformation, 11*(1), 1–23. https://doi.org/10.4018/ijdsgbt.287100

Klein, A. Z. (2022). Ethical issues of digital transformation. *Organizações & Sociedade, 29*(102), 443–448. https://doi.org/10.1590/1984-92302022v29n0020en

Kraus, S., Jones, P., Kailer, N., Weinmann, A., Banegas, N. C., & Tierno, N. R. (2021). Digital transformation: An overview of the current state of the art of research. *SAGE Open, 11*(3), 1–15. https://doi.org/10.1177/21582440211047576

Kumari, T., & Ul Oman, Z. (2024). The modern technology has disrupted today's World: An analytical review of how technology affected quality of human interaction. *International Journal of Computer Trends and Technology, 72*(1), 27–34. https://doi.org/10.14445/22312803/ijctt-v72i1p105

Lucchi, N. (2023). ChatGPT: A case study on copyright challenges for generative artificial intelligence systems. *European Journal of Risk Regulation, 15*(3), 1–23. https://doi.org/10.1017/err.2023.59

Makridakis, S. (2017). The forthcoming artificial intelligence (AI) revolution: Its impact on society and firms. *Futures, 90*(90), 46–60. https://doi.org/10.1016/j.futures.2017.03.006

Matzat, U., & van Ingen, E. (2020). Social inequality and the digital transformation of Western society: What can stratification research and digital divide studies learn from each other? *Soziologie des Digitalen – Digitale Soziologie? 23,* 379–397. https://doi.org/10.5771/9783845295008-379

Melnyk, Yu. B., & Pypenko, I. S. (2020). How will blockchain technology change education future?! *International Journal of Science Annals, 3*(1), 5–6. https://doi.org/10.26697/ijsa.2020.1.1

Mhlanga, D. (2024). Digital transformation of education, the limitations and prospects of introducing the fourth industrial revolution asynchronous online learning in emerging markets. *Discover Education, 3*(1), 1–18. https://doi.org/10.1007/s44217-024-00115-9

Mhlanga, D., Denhere, V., & Moloi, T. (2022). COVID-19 and the key digital transformation lessons for higher education institutions in South Africa. *Education Sciences, 12*(7), Article 464. https://doi.org/10.3390/educsci12070464

Mzwri, K., & Turcsányi-Szabo, M. (2025). The impact of prompt engineering and a generative AI-driven tool on autonomous learning: A case study. *Education Sciences, 15*(2), Article 199. https://doi.org/10.3390/educsci15020199

Pappas, I. O., Mikalef, P., Dwivedi, Y. K., Jaccheri, L., & Krogstie, J. (2023). Responsible digital transformation for a sustainable society. *Information Systems Frontiers, 25*(3), 945–953. https://doi.org/10.1007/s10796-023-10406-5

Patel, K., & McCarthy, M. P. (2000). *Digital transformation: The essentials of e-business leadership*. McGraw-Hill. https://search.worldcat.org/title/digital-transformation-the-essentials-of-e-business-leadership/oclc/464589445?referer=di&ht=edition

Priharsari, D., Abedin, B., Burdon, S., Clegg, S., & Clay, J. (2023). National digital strategy development: Guidelines and lesson learnt from Asia Pacific countries. *Technological Forecasting and Social Change, 196,* Article 122855. https://doi.org/10.1016/j.techfore.2023.122855

Pypenko, I. S., Maslov, Yu. V., & Melnyk, Yu. B. (2020). The impact of social distancing measures on higher education stakeholders. *International Journal of Science Annals, 3*(2), 9–14. https://doi.org/10.26697/ijsa.2020.2.2

Rakibul, M., & Bhuiyan, I. (2022). *Digital transformation and society.* SSRN. https://doi.org/10.2139/ssrn.4604376

Reis, J., Amorim, M., Melão, N., & Matos, P. (2018). Digital transformation: A literature review and guidelines for future research. *Advances in Intelligent Systems and Computing, 745,* 411–421. https://doi.org/10.1007/978-3-319-77703-0_41

Schuster, N., & Kilov, D. (2025). Moral disagreement and the limits of AI value alignment: A dual challenge of epistemic justification and political legitimacy. *AI & Society, 40*(8), 6073–6087. https://doi.org/10.1007/s00146-025-02427-2

Stapleton, L. (2025). AI, society, and the shadows of our desires. *AI & Society, 40*(7), 5109–5113. https://doi.org/10.1007/s00146-025-02484-7

Stolterman, E., Fors, A. C., Truex, D. P., & Wastell, D. (2004). Information technology and the good life. In B. Kaplan, D. P. Truex, & D. Wastell, et al. (Eds.), *Information Systems Research: Relevant Theory and Informed Practice* (pp. 687–693). Kluwer Academic Publishers. https://ifipwg82.org/sites/ifipwg82.org/files/Stolterman.pdf

Stratu-Strelet, D., Gil-Gómez, H., Oltra-Badenes, R., & Oltra-Gutierrez, J. V. (2023). Developing a theory of full democratic consolidation: Exploring the links between democracy and digital transformation in developing eastern European countries. *Journal of Business Research, 157,* Article 113543. https://doi.org/10.1016/j.jbusres.2022.113543

Tana, S., Breidbach, C. F., & Burton-Jones, A. (2023). Digital transformation as collective social action. *Journal of the Association for Information Systems, 24*(6), 1618–1644. https://doi.org/10.17705/1jais.00791

Tang, Q., Kamarudin, S., Rahman, A., & Zhang, X. (2025). Bridging gaps in online learning: A systematic literature review on the digital divide. *Journal of Education and Learning, 14*(1), 161–176. https://eric.ed.gov/?id=EJ1463252

Taufik, D. A. (2025). Artificial intelligence and digital transformation: A study on their impact on industries. *AIRA (Artificial Intelligence Research and Applied Learning), 4(*1), 1–15. https://doi.org/10.1234/aira.v4i1.74

Van Veldhoven, Z., & Vanthienen, J. (2021). Digital transformation as an interaction-driven perspective between business, society, and technology. *Electronic Markets, 32*(2), 629–644. https://doi.org/10.1007/s12525-021-00464-5

Varnosfaderani, S. M., & Forouzanfar, M. (2024). The role of AI in hospitals and clinics: Transforming healthcare in the 21st century. *Bioengineering, 11*(4), Article 337. https://doi.org/10.3390/bioengineering11040337

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems, 28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003

World Economic Forum. (2023). *The future of jobs report 2023: Insight report.* https://www.weforum.org/publications/the-future-of-jobs-report-2023/

Yang, C., Gu, M., & Albitar, K. (2024). Government in the digital age: Exploring the impact of digital transformation on governmental efficiency. *Technological Forecasting and Social Change, 208*(1), Article 123722. https://doi.org/10.1016/j.techfore.2024.123722

Zeleza, P. T., & Okanda, P. M. (2022). Enhancing the digital transformation of African universities: Covid-19 as accelerator. *Journal of Higher Education in Africa, 19*(1), 1–28. https://doi.org/10.57054/jhea.v19i1.1886

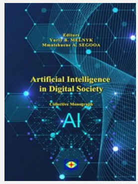
Information about the authors:

Baduza Gugulethu – https://orcid.org/0000-0003-4092-6521; PhD, Dr, Senior Lecturer, Rhodes University, Makhanda, South Africa.

Penxa Lungile – https://orcid.org/0009-0006-2576-5474; PhD, Dr, Lecturer, University of the Western Cape, Cape Town, South Africa.

Ramafi Pelonomi – https://orcid.org/0000-0003-2477-7060; Mcom, Lecturer, University of Witwatersrand, Johannesburg, South Africa.

Artificial Intelligence in Digital Society,
Volume 1, 2026
DOI: 10.26697/9786177089192.2026
ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)



Chapter 4. Artificial Intelligence Adoption and Its Effect on Small and Medium Enterprises' Performance: A Lens of Technology-Organisation-Environment Framework and Ethical Principles

Makelana P. [1]

[1] *Vaal University of Technology, South Africa*

Received: *06.12.2025;* ***Accepted:*** *10.02.2026;* ***Published:*** *10.03.2026*

Abstract

In the digital era, small and medium enterprises (SMEs) adopt AI for business operations to enhance firm performance. The adoption of AI is driven by multiple factors influencing the business landscape, while ethical principles may also impact adoption. The aim of this study is to develop a model that explains factors influencing AI adoption by SMEs to enhance business performance, by integrating the technology-organization-environment (TOE) framework with the diffusion of innovation (DOI) theory and ethical principles. The study employed a quantitative method, using a self-administered questionnaire disseminated among 150 respondents from South African SMEs and analysed using structural equation modelling (SEM). The results showed that compatibility, top management support, organisational readiness, employee capability, customer pressure, vendor support, fairness, accountability, and transparency significantly influence AI adoption, while relative advantage, complexity, high costs, and competitive pressure were less significant. The study concludes that AI adoption is key to enhancing the performance of South African SMEs.

Keywords: artificial intelligence, ethical principles, diffusion of innovations, small and medium enterprises, technology-organisation-environment.

Cite this chapter as:

Makelana, P. (2026). Artificial Intelligence adoption and its effect on small and medium enterprises' performance: A lens of technology -organisation-environment framework and ethical principles. In Y. B. Melnyk & M. A. Segooa (Eds.), *Artificial Intelligence in Digital Society, Vol. 1.* (pp. 55–69). KRPOCH. https://doi.org/10.26697/aids.2026.4

The electronic version of this chapter is complete. It can be found online in the AIDS Archive https://doi.org/10.26697/aids

Introduction

The digital era has created many opportunities and challenges for organisations worldwide (Achieng & Malatji, 2022). Leading this change is artificial intelligence (AI) (Rajaram & Tinguely, 2024; Rana, Pillai, Sivathanu & Malik, 2024). AI enables machines to imitate human intelligence using technologies like natural language processing (NLP), image recognition, machine learning (ML), and deep learning (Badghish & Soomro, 2024), transforming how organisations operate (Sanchez, Calderon & Herrera, 2025).

Research shows AI adoption benefits organisations by boosting creativity and productivity (Wang, Lin, & Shao, 2023; Rajaram & Tinguely, 2024). Across sectors, AI is especially valuable for small and medium enterprises (SMEs) (Badghish & Soomro, 2024), helping them deliver personalised and effective marketing strategies (Mokhtar & Salimon, 2022).

Knowledge Gaps and the Purpose of the Study

The present literature divulges why SMEs need to adopt AI and mainly emphasizes the benefit of AI to SMEs (Rana et al., 2024; Sanchez et al., 2025; Ardito, Filieri, Raguseo & Vitari, 2025; Visuthiphol & Pankham, 2025). A plethora of research on the adoption of AI in the SMEs environment has revealed various factors of the adoption of AI (Mokhtar & Salimon, 2022; Badghish & Soomro, 2024; Hamida, 2025).

Nevertheless, there is limited research on technological, organisational, environmental, and ethical principles influencing AI adoption among SMEs in developing countries. Thus, scholarly research is needed to understand these factors, particularly in South Africa. This study aims to develop a model explaining factors influencing AI adoption in SMEs.

Research Questions

The present study aims to answer the following research questions:

- What are the technological, organisational and environmental factors influencing the adoption of AI in SMEs?
- What are the ethical principles influencing the adoption of AI in SMEs?

Problem Statement

Many countries are benefiting from AI adoption, and South Africa has seen an increase in this trend (Muzuva, Zhou & Zondo, 2024). However, inequality, inadequate infrastructure, and unemployment remain challenges (Vuyani, Gervase-Iwu, Tengeh & Esambe, 2021; Matekenya & Moyo, 2022). The government considers SMEs key drivers of economic growth (Bvuma & Marnewick, 2020; Enaifoghe & Ramsuraj, 2023), yet SMEs have a high failure rate of 70% to 80 % (Tlhagale & Nyoka, 2025; Bolosha, Sinyolo & Ramoroka, 2022) due to limited digital skills, poor access to global markets, and inadequate ICT infrastructure (Bvuma & Marnewick, 2020; Vuyani et al., 2022). AI can help SMEs address these challenges (Wang et al., 2023). However, AI is often perceived as complex and difficult to adopt (Chatterjee Rana, Dwivedi & Baabdullah, 2021;

O'Shaughnessy, Schiff, Varshney, Rozell, & Davenport, 2023), and its impact on SME performance in South Africa remains underexplored (Muzuva, Zhou, & Zondo, 2024).

Literature Review

The Benefits of AI Adoption in SMEs

AI has become a new and strategic trend for all economic sectors, particularly SMEs (Badghish & Soomro, 2024). AI is strategically utilised to share information and long-term relationship building with customers (Khan, Emon & Rahman, 2024), making it a prevalent form of marketing adopted by SMEs (Kedi, Ejimuda, Idemudia & Ijomah, 2024). Therefore, a plethora of research asserts that the adoption of AI by SMEs during crises can also aid in company survival since AI can facilitate closer customer engagement with customers and the supply chain while reducing costs (Chen, Hu, Zhou and Yang, 2023; Rajaram & Tinguely, 2024). Through the adoption of AI, SMEs can engage in two-way communication with customers and interact with each other (Visuthiphol & Pankham, 2025).

Because SMEs often face resource scarcity and limited technical capabilities, AI helps address these challenges by offering cost-effective and user-friendly solutions that support business operations (Mokhtar & Salimon, 2022; Wang et al., 2023). Extant research has identified several advantages of AI adoption for SMEs, including reduced costs, improved customer awareness and knowledge sharing (Hamida, 2025; Maiti, Kayal & Vujko, 2025; Visuthiphol & Pankham, 2025). Furthermore, Badghish and Soomro (2024) found that AI adoption can improve SMEs performance. Although AI improve SMEs performance, it is essential to uphold ethical principles (Omonov & Ahn, 2025).

Ethical AI Principles

Organisations with clear ethical principles are more likely to adopt AI responsibly (Rana et al., 2024). Ethical AI principles include fairness, accountability, and transparency (Omonov & Ahn, 2025). Fairness of AI refers to treating everyone equally. It means AI should not be used in a way that is unfair or harmful to employees (Ashok, Madan, Joha & Sivarajah, 2022). If AI ensures outcomes that are fair, unbiased, and devoid of prejudice, it would likely be adopted by organisations (Shin & Park, 2019).

Accountability in AI refers to the responsibility of organisations to ensure that AI function correctly and produce ethical outcomes (Floridi, Cowls, King & Taddeo, 2020). It includes deciding who is responsible for the decisions made by AI, such as the organisation using the system or the developers who created it (Rana et al., 2024).

Since AI relies on programmed code and existing data, errors or unfair results can occur (Melnyk, 2025; Omonov & Ahn, 2025). Therefore, organisations must take responsibility for how AI is developed and used (Ashok et al., 2022). AI transparency means being clear about how AI works and what data it uses (Floridi et al., 2020). It helps people and organisations understand how the AI makes decisions and whether the results make sense (Rana et al., 2024).

Because AI systems are complex, they can be hard to understand (Shin & Park, 2019). Organizations should clearly explain what data AI uses, how it processes the data, and if the data is suitable, so people can trust and use AI properly (Omonov & Ahn, 2025).

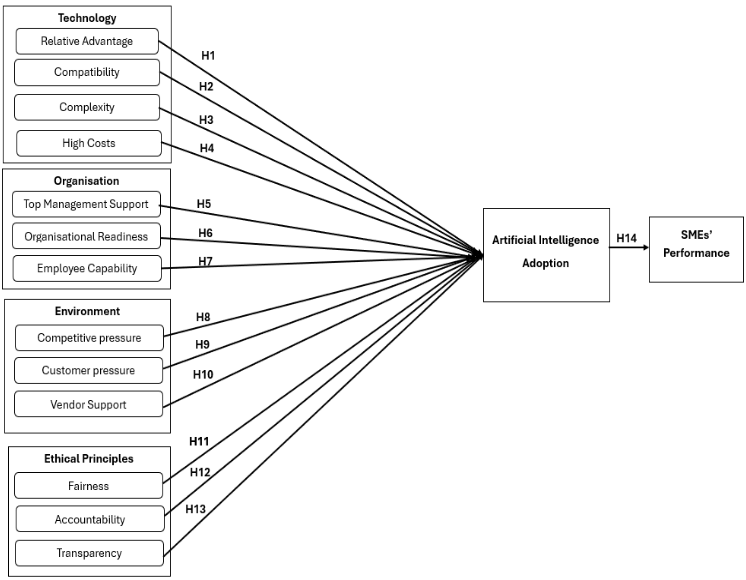
Conceptual Framework and Hypotheses Development

The present study employs technology-organisation-environment (TOE) framework and the diffusion of innovations (DOI) theory to explain AI adoption by SMEs. TOE, introduced by Depietro, Wiarda and Fleischer in 1990 examines why organisations adopt new technologies through three contexts: technology, organisation, and environment. The TOE framework has been applied in Saudi Arabia (Badghish & Soomro, 2024), India (Karan & Angadi, 2025), and Malaysia (Masod & Zakaria, 2023).

The DOI theory, introduced by Rogers in 2003, is a theory used to explain how organisations accept new ideas and new technology. DOI has five attributes, including relative advantage, compatibility, trialability, complexity and observability (Rogers, 2003). DOI has been applied in studies on AI adoption (Badghish & Soomro, 2024; Sanchez et al., 2025).

Figure 4.1

Conceptual Framework



Technology

The adoption of AI in SMEs is influenced by technological factors, particularly relative advantage. SMEs are more likely to adopt new technology if it improves on current systems (Maroufkhani, Tseng, Iranmanesh, Ismail & Khalid, 2020) and aligns with existing business processes (Badghish & Soomro, 2024). However, adoption may fail if AI is seen as costly or very difficult to adopt (Apostoaie, Roman, Maxim & Jijie, 2025). Based on this, we propose the following hypotheses:

H1: Relative advantage influences AI adoption in SMEs.
H2: Compatibility influences AI adoption in SMEs.
H3: Complexity influences AI adoption in SMEs.
H4: High costs influence AI adoption in SMEs.

Organisation

AI adoption in SMEs is also influenced by organisational factors, especially top management support. Siradhana and Arora (2024) and Mathagu (2024) find that strong top management support boosts AI adoption success. Apostoaie et al. (2025) note that SMEs also need financial, technological, and skilled human resources, and that qualified employees are key to successful AI adoption. Based on these findings, the following hypotheses are formulated:

H5: Top management support influences AI adoption in SMEs.
H6: Organisational readiness influences AI adoption in SMEs.
H7: Employee capability influence AI adoption in SMEs.

Environment

A study by Apostoaie et al. (2025) highlight that environmental factors, especially competitive pressure, influence AI adoption in SMEs. Mokhtar and Salimon (2022) support this, noting competitor pressure drives adoption. Badghish and Soomro (2024) add that customer awareness of new technologies can push organisations to adopt AI to meet service needs. Apostoaie et al. (2025) also argue that vendor support and training increase adoption likelihood. Based on this, we propose the following hypotheses:

H8: Competitive pressure influences AI adoption in SMEs.
H9: Customer pressure influences AI adoption in SMEs.
H10: Vendor support influences AI adoption in SMEs.

Ethical Principles

Ethical principles such as fairness, accountability, and transparency are key for AI adoption (Ashok, Madan, Joha & Sivarajah, 2022). Fairness means treating everyone equally and avoiding discrimination (Ashok et al., 2022). AI is more likely to be adopted if it produces fair results (Rana et al., 2024). Accountability involves responsibility for AI outcomes, which can be difficult to assign because they depend on programming, data, and engineers' decisions (Floridi, Cowls, King & Taddeo, 2020).

Ensuring ethical AI behavior builds trust and supports adoption (Floridi et al., 2020; Rana et al., 2024). Transparency requires clarity on how AI works and learns from users (Shin & Park, 2019). It helps organisations understand decision-making and data quality, influencing interaction and adoption (Rana et al., 2024). Based on these claims, we propose the following hypotheses:

H11: Fairness influences AI adoption in SMEs.

H12: Accountability influences AI adoption in SMEs.

H13: Transparency influences AI adoption in SMEs.

AI Adoption and SMEs' Performance

AI applications enable SMEs' facilitates data driven, agile, and proactive decision-making for immediate business impact (Mokhtar & Salimon, 2022; Rajaram & Tinguely, 2024). By adopting AI, SMEs' can leverage customers' engagements in business and thereby improve their market performance (Badghish & Soomro, 2024; Siradhana & Arora, 2024). Accordingly, the present study suggests the following hypothesis:

H14: AI adoption influences SMEs' performance.

Methodology

Research Design and Approach

The present study employed a quantitative approach to empirically test the hypotheses derived from the conceptual framework. A cross-sectional survey was utilised, meaning data was gathered from respondents at a single point in time. This method is appropriate for exploring relationships among variables within a conceptual framework and is widely applied in studies focusing on AI adoption. The study aligns with the positivist paradigm, which posits that social phenomena can be examined objectively and that variable relationships can be measured and validated through statistical analysis (Saunders, Lewis & Thornhill, 2009).

Data Collection and Analysis

This present study targeted SMEs operating in South Africa. To collect data from South African SMEs, a closed-ended questionnaire was developed and physically distributed. A study conducted by Saunders Lewis and Thornhill (2019), asserts that closed-ended questions are more specific and less susceptible to interpretation and verbosity than open-ended questions.

The questionnaire items were measured using a five (5)-point Likert-type scale ranging from 1 to 5, where 1 and 2 represented strongly disagree and disagree, respectively, and 4 and 5 represented agree and strongly agree. A total of 200 questionnaires were distributed to South African SMEs, and 150 were returned. The collected data were analysed using the Statistical Package for the Social Sciences (SPSS) version 28.

Results

Demographic Profile of Respondents

As depicted in Table 4.1, 66.6% of the respondents were male and 33.3% were female. Most respondents were aged 20 to 29 years (46.6%), followed by those aged 30 to 39 years (40%), while 13.3% were aged 50 years and above.

Table 4.1 further indicates that the majority of respondents held a diploma (46.6%), with others holding a B-tech (20%), a master's degree (13%), a PhD (10%), and a matric (10%). Regarding AI adoption, most respondents (86.6%) stated that they had adopted AI, whereas 13.3% reported that they had never adopted it.

Table 4.1

Demographic Profile of Respondents (N=150)

Variables		Frequency	Percentage
Gender	Male	100	66.6
	Female	50	33.3
	Total	150	100.0
Age	20-29 years	70	46.6
	30-39 years	60	40.0
	50 years and above	20	13.3
	Total	150	100.0
Education	Matric	15	10.0
	Diploma	70	46.6
	B-tech	30	20.0
	Master's	20	13.3
	PhD	15	10.0
	Total	150	100.0
Artificial Intelligence Adoption	Yes	130	86.6
	No	20	13.3
	Total	150	100.0

Assessment of Measurement Model

In this section, the measurement model was examined. As noted by Hair, Gudergan, Fischer, Nitzl and Menictas (2019), the measurement model is assessed by using factor loadings (FL), composite reliability (CR), and average variance extracted (AVE). According to Hair et al. (2019), the values of FL, CR and AVE should be greater than 0.7, 0.5, and 0.7, respectively. As in Table 4.2, all the constructs meet the threshold requirements and demonstrate acceptable convergent validity.

Table 4.2
Loadings Reliability and Validity Statistics

Construct	Item	Outer Loading	FL	CR	AVE
Relative Advantage (RA)	RA1	0.816	0.868	0.855	0.925
	RA2	0.819			
	RA3	0.850			
Compatibility (CM)	CM1	0.806	0.762	0.839	0.785
	CM2	0.756			
	CM3	0.825			
Complexity (CX)	CX1	0.915	0.870	0.918	0.875
	CX2	0.786			
	CX3	0.865			
High Costs	HC1	0.925	0.623	0.855	0.925
	HC2	0.856			
	HC3	0.775			
Top Management Support (TMS)	TMS1	0.775	0.778	0.834	0.775
	TMS2	0.782			
	TMS3	0.894			
Organisational Readiness (OR)	OR1	0.894	0.765	0.856	0.905
	OR2	0.744			
	OR3	0.825			
Employee Capability (EC)	EC1	0.855	0.862	0.872	0.835
	EC2	0.775			
	EC3	0.802			
Competitive Pressure (CP)	CP1	0.775	0.716	0.924	0.844
	CP2	0.893			
	CP3	0.755			
Customer Pressure (CSP)	CSP1	0.935	0.885	0.789	0.943
	CSP2	0.819			
	CSP3	0.766			
Vendor Support (VS)	VS1	0.884	0.967	0.815	0.977
	VS2	0.952			
	VS3	0.778			
Fairness (FE)	FE1	0.841	0.909	0.875	0.845
	FE2	0.857			
	FE3	0.910			
Accountability (AC)	AC1	0.812	0.918	0.798	0.757
	AC2	0.856			
	AC3	0.924			
Transparency (TP)	TP1	0.840			
	TP2	0.765			
	TP3	0.816			

Assessment of Structural Model

In this section, SEM was used to test hypotheses (Figure 4.3). Of 14 paths, ten are significant. Table 4.3 shows that compatibility (H2, $p<0.05$), top management support (H5, p<0.05), organisational readiness (H6, p<0.05), employee capability (H7, p<0.05), customer pressure (H9, p<0.05), vendor support (H10, p<0.05), fairness (H11, p<0.05), accountability (H12, p<0.05), transparency (H13, p<0.05), and AI adoption (H4, p<0.05) are supported, significantly influencing AI adoption by SMEs. Nevertheless, Relative advantage (H1, p>0.05), complexity (H3, p>0.05), high costs (H4, p>0.05), and competitive pressure (H8, p>0.05) are not supported, indicating these four factors are insignificant to AI adoption.

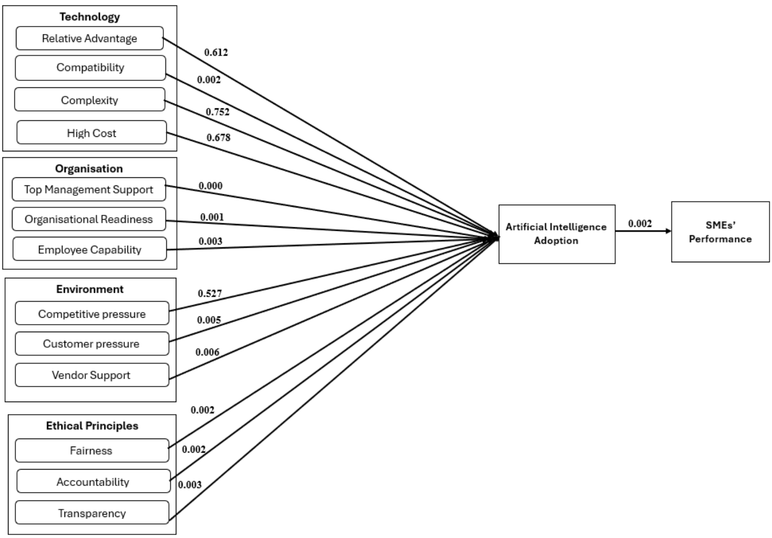
Table 4.3

Hypotheses Testing

Constructs	Std. Beta	T-Value	*p* values	Results
H1 Relative advantage →AI adoption	0.052	0.645	0.612	Not supported
H2 Compatibility →AI adoption	0.132	0.124	0.002	Supported
H3 Complexity →AI adoption	0.273	1.756	0.752	Not supported
H4 High Costs →AI adoption	0.142	0.432	0.678	Not supported
H5 Top Management Support→AI adoption	0.147	1.572	0.000	Supported
H6 Organisational Readiness →AI adoption	0.235	4.552	0.001	Supported
H7 Employee Capability →AI adoption	0.178	2.071	0.003	Supported
H8 Competitive Pressure →AI adoption	0.052	9.357	0.527	Not supported
H9 Customer Pressure →AI adoption	0.346	2.473	0.005	Supported
H10 Vendor Support →AI adoption	0.234	0.451	0.006	Supported
H11 Fairness →AI adoption	0.126	3.124	0.002	Supported
H12 Accountability →AI adoption	0.057	1.756	0.002	Supported
H13 Transparency →AI adoption	0.132	0.432	0.003	Supported
H14 AI adoption →SMEs' performance	0.427	12.357	0.002	Supported

Figure 4.2

Structural Model



Discussion

This study examined factors influencing AI adoption in SMEs. Technological factors showed that relative advantage (H1) had no significant effect, and this is supported by Fu, Silalahi, Yang and Eunike (2024). Compatibility (H2) positively influenced adoption, and this is supported by Bhardwaj, Garg, and Gajpal (2021). Complexity (H3) and high costs (H4) negatively affected adoption, and this is supported by Apostoaie et al. (2025).

Organisational factors including top management support (H5), organisational readiness (H6), and employee capability (H7) positively influenced adoption. This is supported by Siradhana and Arora (2024) and Apostoaie et al. (2025), suggesting SMEs adopt AI when management supports it, resources are available, and employees understand its benefits.

Environmental factors showed that competitive pressure (H8) negatively affects AI adoption, and this is supported by Mokhtar and Salimon (2022), while customer pressure (H9) and vendor support (H10) positively influence AI adoption. This is supported by Badghish and Soomro (2024) and Apostoaie et al. (2025).

Ethical principles including fairness (H11), accountability (H12), and transparency (H13) significantly affect AI adoption, and this is supported by Rana et al. (2024). AI adoption (H14) positively impacts SME performance, and this is supported by Badghish and Soomro (2024).

Theoretical and Practical Contribution

This study develops a model that integrates the TOE framework with AI ethical principles to explain AI adoption and its impact on SME performance. It addresses calls to examine technological, organizational, and environmental factors influencing AI adoption. The study adds a new perspective by incorporating ethical principles at the organizational level. Practically, the model helps SME managers understand lower AI adoption compared to large firms and assess technological, organizational, environmental, and ethical factors influencing AI adoption in South African SMEs.

Conclusion

The rapid growth of AI has encouraged researchers and practitioners to study its role in improving enterprise performance using the TOE framework. This study proposes a model to identify factors influencing AI adoption and its impact on South African SME performance. The results show that compatibility, management support, organisational readiness, employee capability, customer pressure, vendor support, fairness, accountability, and transparency strongly influence AI adoption, while relative advantage, complexity, cost, and competitive pressure are less influential. The study concludes that AI adoption is important for improving South African SME performance.

Ethical Approval

The study obtained ethical clearance from the institution's Ethics Committee (Ref no. FCRE/ICT/2022/03/001 (1).

References

Achieng, M. S., & Malatji, M. (2022). Digital transformation of small and medium enterprises in sub-Saharan Africa: A scoping review. *The Journal for Transdisciplinary Research in Southern Africa, 18*(1), 1–13. https://doi.org/10.4102/td.v18i1.1257

Apostoaie, C. M., Roman, T., Maxim, A., & Jijie, D. (2025). Determinants of AI adoption intention in SMEs. Romanian case study. *Journal of Business Economics and Management, 26*(2), 277–296. https://doi.org/10.3846/jbem.2025.23650

Ardito, L., Filieri, R., Raguseo, E., & Vitari, C. (2025). Artificial intelligence adoption and revenue growth in European SMEs: Synergies with IoT and big data analytics. *Internet Research, 35*(4), 1066–2243. https://doi.org/10.1108/INTR-02-2024-0195

Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for Artificial Intelligence and Digital technologies. *International Journal of Information Management, 62*, Article 102433. https://doi.org/10.1016/j.ijinfomgt.2021.102433

Badghish, S., & Soomro, Y. A. (2024). Artificial intelligence adoption by SMEs to achieve sustainable business performance: Application of technology–organization–environment framework. *Sustainability (Switzerland), 16,* 1–24. https://doi.org/10.3390/su16051864

Bhardwaj, A. K., Garg, A., & Gajpal, Y. (2021). Determinants of blockchain technology adoption in supply chains by small and medium enterprises (SMEs) in India. *Mathematical Problems in Engineering,* 1–14. https://doi.org/10.1155/2021/5537395

Bolosha, A., Sinyolo, S., & Ramoroka, K. (2022). Factors influencing innovation among small, micro and medium enterprises (SMMEs) in marginalized settings: Evidence from South Africa. *Innovation and Development, 13*(3), 583–601. https://doi.org/10.1080/2157930X.2022.2092681

Bvuma, S., & Marnewick, C. (2020). An information and communication technology adoption framework for small, medium and micro-enterprises operating in townships South Africa. *Southern African Journal of Entrepreneurship and Small Business Management, 12*(1), 1–12. https://doi.org/10.4102/sajesbm.v12i1.318

Chatterjee, S., Rana, N. P., Dwivedi, Y. K., & Baabdullah, A. (2021). Understanding AI adoption in manufacturing and production firms using an integrated TAM-TOE model. *Technological Forecasting and Social*

Change, 170, Article 120880. https://doi.org/10.1016/j.techfore.2021.120880

Chen, Y., Hu, Y., Zhou, & Yang, S. (2023). Investigating the determinants of performance of artificial intelligence adoption in hospitality industry during COVID-19. *International Journal of Contemporary Hospitality Management, 35*(8), 2868–2889. https://doi.org/10.1108/IJCHM-04-2022-0433

Darwish, A., Hassanien, A. E., Elhoseny, M., Kumar, A., & Khan, S. (2019). The impact of the hybrid platform of internet of things and cloud computing on healthcare systems : Opportunities, challenges, and open problems. *Journal of Ambient Intelligence and Humanized Computing, 10*(10), 4151–4166. https://doi.org/10.1007/s12652-017-0659-1

Depietro, R., Wiarda, E., & Fleischer, M. (1990). The context for change: Organization, technology and environment. In *The processes of technological innovation*. Lexington, MA: Lexington Books. https://doi.org/10.7251/emc2201237t

Enaifoghe, A., & Ramsuraj, T. (2023). Examining the function and contribution of entrepreneurship through small and medium enterprises as drivers of local economic growth in South Africa. *African Journal of Inter/Multidisciplinary Studies,* 5(1), 1–11. https://doi.org/10.51415/ajims.v5i1.1172

Floridi, L., Cowls, J., King, T., & Taddeo, M. (2020). How to design AI for social good: Seven essential factors. *Science and Engineering Ethics, 26*(3), 1771–1796. https://doi.org/10.1007/s11948-020-00213-5

Fu, C. J., Silalahi, A. D. K., Yang, L. W., & Eunike, I. (2024). Advancing SME performance: A novel application of the technological-organizational-environment framework in social media marketing adoption. *Cogent Business & Management, 11*(1), 1–25. https://doi.org/10.1080/23311975.2024.2360509

Hair, J. F., Ringle, C. M., Gudergan, S. P., Fischer, A., Nitzl, C., & Menictas, C. (2019). Partial least squares structural equation modeling-based discrete choice modeling: An illustration in modeling retailer choice. *Business Research, 12*(1), 115–142. https://doi.org/10.1007/s40685-018-0072-4

Hamida, A. G. (2025). Adoption of artificial intelligence technology by SMEs: Impact on customized e-marketing strategies and online purchase. *International Conference on Technology Enabled Economic Changes (InTech),* 1319–1324. https://doi.org/10.1109/intech64186.2025.11198535

Karan, B., & Angadi, G. R. (2025). Understanding school readiness factors in relation to the incorporation of artificial intelligence using TOE framework: An empirical evidence from India. *TechTrends, 69*(1), 38–59. https://doi.org/10.1007/s11528-024-01020-6

Kedi, W.E., Ejimuda, C., Idemudia, C., & Ijomah, T. (2024). AI Chatbot integration in SME marketing platforms: Improving customer interaction and service efficiency. *International Journal of Management & Entrepreneurship Research,* *6*(7), 2332–2341. https://doi.org/10.51594/ijmer.v6i7.1327

Khan, T., Emon, M. H., & Rahman, S. (2024). Marketing strategy innovation via AI adoption: A study on Bangladeshi SMEs in the context of industry 5.0. *2024 6th International Conference on Sustainable Technologies for Industry 5.0, STI 2024,* 1–6. https://doi.org/10.1109/STI64222.2024.10951050

Maiti, M., Kayal, P., & Vujko, A. (2025). A study on ethical implications of artificial intelligence adoption in business: Challenges and best practices. *Future Business Journal, 11*(1), 1–12. https://doi.org/10.1186/s43093-025-00462-5

Maroufkhani, P., Tseng, M. L., Iranmanesh, M., Ismail, W. K. W., & Khalid, H. (2020). Big data analytics adoption: Determinants and performances among small to medium-sized enterprises. *International Journal of Information Management,* *54*(July), Article 102190. https://doi.org/10.1016/j.ijinfomgt.2020.102190

Masod, M.Y., & Zakaria, S. (2023). Artificial intelligence adoption in the manufacturing sector: Challenges and strategic framework. *International Journal of Research and Innovation in Social Science, VII*(2454), 1175–1189. https://doi.org/10.47772/IJRISS

Matekenya, W., & Moyo, C. (2022). Innovation as a driver of SMME performance in South Africa: A quantile regression approach. *African Journal of Economic and Management Studies*, *13*(3), 452–467. https://doi.org/10.1108/AJEMS-06-2021-0306

Mathagu, S. (2024). Artificial intelligence in small and medium enterprises – An empirical analysis of critical factors. *Premier Journal of Science, 1*, Article 100009. https://doi.org/10.70389/pjs.100009

Melnyk, Y. B. (2025). Should we expect ethics from artificial intelligence: The case of ChatGPT text generation. *International Journal of Science Annals, 8*(1), 5–11. https://doi.org/10.26697/ijsa.2025.1.5

Mokhtar, S. S. M., & Salimon, M. G. (2022). SMEs' adoption of artificial intelligence-chatbots for marketing communication: A conceptual framework for an emerging economy. In Adeola, O., Hinson, R. E., Sakkthivel, A. M. (Eds.), *Marketing Communications and Brand Development in Emerging Markets: Volume II. Palgrave Studies of Marketing in Emerging Economies* (pp. 25–53). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-95581-6_2

Muzuva, M., Zhou, H., & Zondo, R. W. (2024). Has generative AI become of age: Assessing its impact on the productivity of SMEs in South Africa.

International Journal of Research in Business and Social Science, 13(7), 2147–4478. https://doi.org/10.20525/ijrbs.v13i7.3576

O'Shaughnessy, M. R., Schiff, D. S., Varshney, L. R., Rozell, C.J., & Davenport, M. (2023). What governs attitudes toward artificial intelligence adoption and governance? *Science and Public Policy, 50*(2), 161–176. https://doi.org/10.1093/scipol/scac056

Omonov, M. S., & Ahn, Y. (2025). Towards smart public administration: A TOE-based empirical study of AI chatbot adoption in a transitioning government context. *Administrative Sciences, 15*(8), 1–29. https://doi.org/10.3390/admsci15080324

Rajaram, K., & Tinguely, P. (2024). Generative artificial intelligence in small and medium enterprises: Navigating its promises and challenges. *Business Horizons, 67,* 629–648. https://doi.org/10.1016/j.bushor.2024.05.008

Rana, N. P., Pillai, R., Sivathanu, B., & Malik, N. (2024). Assessing the nexus of Generative AI adoption, ethical considerations and organizational performance. *Technovation, 135*(June), Article 103064. https://doi.org/10.1016/j.technovation.2024.103064

Sanchez, E., Calderon, R., & Herrera, F. (2025). Artificial intelligence adoption in SMEs: Survey based on TOE–DOI framework, primary methodology and challenges. *Applied Sciences, 15,* 1–43. https://doi.org/10.3390/app15126465

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students* (5th ed.). Pearson Education Limited. https://anyflip.com/ftteb/yphw/basic/

Saunders, M. N. K., Bristow, A., Thornhill, A., & Lewis, P. (2019). Understanding research philosophy and approaches to theory development. In M. N. K. Saunders, P. Lewis, & A. Thornhill (Eds.), *Research Methods for Business Students* (8th ed., pp. 128–171). Harlow: Pearson Education. https://oro.open.ac.uk/66370/

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior, 98*, 277–284. https://doi.org/10.1016/j.chb.2019.04.019

Siradhana, N. K., & Arora, R. (2024). Examining the influence of artificial intelligence implementation in HRM practices using T-O-E model. *Vision: The Journal of Business Perspective.* https://doi.org/10.1177/09722629241231458

Tlhagale, F. K., & Nyoka, C. (2025). The high unemployment rate and the high failure rate of black-owned small to medium enterprises in South Africa: The paradox. *African Journal of Innovation and Entrepreneurship, 4*(1), 357–374. https://doi.org/10.31920/2753-314x/2025/v4n1a15

Visuthiphol, S., & Pankham, S. (2025). Artificial intelligence-enabled decision making in social media adoption for sustainable digital business in Thai

SMEs. *Decision Making: Applications in Management and Engineering, 8*(1), 333–347. https://doi.org/10.31181/dmame8120251366

Vuyani, R., Gervase-Iwu, C., Tengeh, R. K., & Esambe, E. (2021). SMEs, economic growth, and business incubation conundrum in South Africa: A literature appraisal. *Journal of Management and Research, 8*(2), 214–251. https://doi.org/https://doi.org/10.29145/jmr/82/08

Wang, X., Lin, X., & Shao, B. (2023). Artificial intelligence changes the way we work: A close look at innovating with chatbots. *Journal of the Association for Information Science and Technology, 74*(3), 339–353. https://doi.org/10.1002/asi.24621

Information about the author:

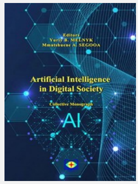
Makelana Phenuel – https://orcid.org/0000-0003-0986-1117; Doctor of Computing, Lecturer, Department of Computer Sciences, Vaal University of Technology, Vanderbijlpark, South Africa.

PART III

ARTIFICIAL INTELLIGENCE-BASED CHATBOTS AND INTELLIGENT AGENTS

Check for updates

Artificial Intelligence in Digital Society,
Volume 1, 2026
DOI: 10.26697/9786177089192.2026
ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)



Chapter 5. Artificial Intelligence-Driven Chatbots and Intelligent Agents for Monitoring, Evaluation, and Organisational Learning: A Review of Techniques and Trends

Kgopa A. T. [1], **Msweli N. T.**[1]

[1] *University of South Africa, South Africa*

Received: *11.12.2025;* ***Accepted:*** *10.02.2026;* ***Published:*** *10.03.2026*

Abstract

Artificial Intelligence (AI)-driven chatbots and intelligent agents are increasingly deployed to support monitoring, evaluation, and organisational learning. Advances in large language models, retrieval-augmented generation, and multi-agent architectures have expanded conversational AI. Despite growing adoption, existing research remains fragmented across technical, educational, and organisational domains, limiting holistic understanding of their design, impact, and governance. This gap creates challenges for organisations seeking evidence-based guidance for implementation. The purpose of this study is to examine and synthesise existing literature on the design, adoption, and impact of AI-based chatbots and intelligent agents within the contexts of monitoring, evaluation, and internal organisational operations. This study presents a systematic literature review and bibliometric analysis of peer-reviewed studies (2021-2025). The review analyses publication trends, core techniques, and application areas of AI-driven chatbots and intelligent agents. Findings reveal rapid growth and increasing focus on organisational learning and evaluation use cases. The study contributes a consolidated synthesis of techniques, benefits, and challenges, identifies research gaps, and offers directions for future research and evidence-based adoption of conversational AI in organisational environments.

Keywords: artificial intelligence-driven chatbots, intelligent agents, artificial intelligence, ChatGPT, organisational learning.

Cite this chapter as:

Kgopa, A. T., & Msweli, N. T. (2026). Artificial intelligence-driven chatbots and intelligent agents for monitoring, evaluation, and organisational learning: A review of techniques and trends. In Y. B. Melnyk & M. A. Segooa (Eds.), *Artificial Intelligence in Digital Society, Vol. 1.* (pp. 71–86). KRPOCH. https://doi.org/10.26697/aids.2026.5

The electronic version of this chapter is complete. It can be found online in the AIDS Archive https://doi.org/10.26697/aids

Introduction

Artificial intelligence (AI) has rapidly transformed human-computer interaction through advances in machine learning, deep learning, and natural language processing, leading to the emergence of sophisticated conversational systems such as chatbots and intelligent agents (Lee & Li, 2023; Martins et al., 2022). Initially developed as rule-based question-answering tools, these systems have evolved into adaptive and context-aware agents capable of supporting decision-making, learning, and complex organisational tasks (Marroquin & Senadji, 2025).

In recent years, organisations have increasingly integrated AI-driven chatbots into monitoring and evaluation (M&E) and organisational learning processes to enable real-time data interpretation, automated reporting, feedback generation, and continuous improvement (Chen & Gasco-Hernandez, 2024; Hutson & Plate, 2023). The introduction of large language models (LLMs), retrieval-augmented generation (RAG), and multi-agent architectures has further enhanced the ability of conversational AI to support personalised learning, adaptive assessments, and evidence-based organisational decision-making (Marroquin & Senadji, 2025; Burov et al., 2025).

Despite these advances, existing research remains fragmented across technical, educational, and organisational perspectives, offering limited integrated insight into the design, evaluation, governance, and long-term impact of conversational AI systems (Al-Sharafi et al., 2023; Gkinko & Elbanna, 2023). Many organisations adopt chatbots without a comprehensive understanding of ethical risks, performance alignment, and sustainability within organisational learning ecosystems (Bartosiak & Modlinski, 2022; Melnyk & Pypenko, 2025; Qiao et al., 2022). Therefore, this study systematically synthesises the literature on AI-driven chatbots and intelligent agents to clarify their roles, impacts, and challenges in monitoring, evaluation, customer support, and organisational learning contexts. The study is guided by the following objectives:

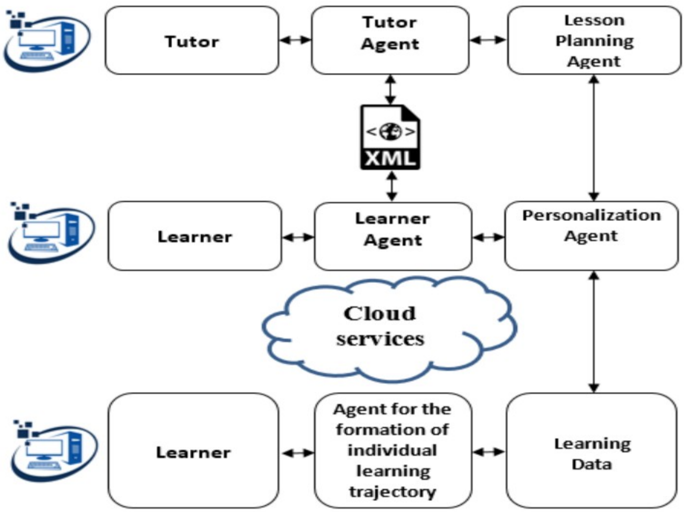
- To examine and synthesise existing literature on the design, adoption, and application of AI-driven chatbots and intelligent agents in monitoring, evaluation, customer support, and organisational learning contexts.
- To analyse research trends and patterns using bibliometric techniques to identify publication growth, collaboration networks, dominant sources, and key application domains related to conversational AI.
- To identify gaps, challenges, and future research directions to inform the development of integrated, ethical, and effective conversational AI frameworks for organisational monitoring, evaluation, and continuous learning.

Related Work

Recent scholarship shows AI-driven chatbots and intelligent agents are moving from “Q&A tools” to organisational infrastructure for continuous learning, monitoring, and evaluation. Marroquin and Senadji (2025) frame generative conversational agents as workplace learning technologies, highlighting their value for on-demand, contextual support, while noting gaps in peer learning, learning tracking, and integration with performance systems key requirements for monitoring and evaluations.

A strong architectural strand is multi-agent learning environments. Burov et al. (2025) proposes intelligent agent-managers for personal learning environments, where tutor/learner agents interact with planning and personalisation services (as illustrated in the attached framework). Conceptually, it can be summarised as shown in Figure 5.1.

Figure 5.1
Architecture of Distance Learning Multi-Agent Management



Note. From "Using intelligent agent-managers to build personal learning environments in the e-learning system", by O. Yu. Burov et al., 2025, *Proceedings of the 7th International Workshop on Augmented Reality in Education*, p. 127 (https://ceur-ws.org/Vol-3918/).

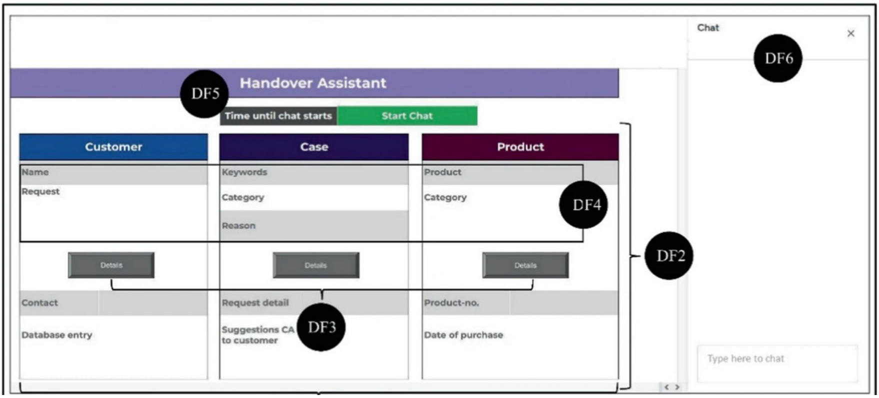
For embedding conversational AI into organisational workflows, Klievtsova et al. (2023) synthesise "conversational process modelling," showing how dialogue interfaces can capture, adapt, and operationalise business processes useful for monitoring compliance and performance. Extending this, Klievtsova et al. (2025) advance conversationally actionable process model creation, enabling process models that can be executed and queried through conversation, supporting auditable evaluation cycles.

Governance, adoption, and human factors are equally prominent. Gkinko and Elbanna (2023) provide a taxonomy of workplace chatbot users, demonstrating that outcomes depend on both design (e.g., social presence) and organisational context, shaping user emotions and appropriation patterns critical for sustained M&E uptake. In education, Al-Sharafi et al. (2023) show that knowledge-management factors strongly influence sustainable chatbot use, implying that organisational learning benefits require deliberate support for knowledge acquisition and application, not only chatbot capability.

Operationally, hybrid service designs matter. Poser et al. (2022) propose effective handover from conversational agents to human employees, addressing service failures and ensuring continuity an essential control mechanism when chatbots support monitoring/reporting:

Figure 5.2

Web-Based Handover Assistant Chatbot



Note. From "Don't throw it over the fence! Toward effective handover from conversational agents to service employees", by M. Poser et al., 2022, *Human-Computer Interaction. User Experience and Behavior. HCII 2022. Lecture Notes in Computer Science, 13304* (https://doi.org/10.1007/978-3-031-05412-9_36). Copyright 2022 by Springer Nature Switzerland AG.

Despite extensive research on AI-driven chatbots and intelligent agents, the literature reveals limited empirical evidence on their integrated use for M&E and continuous organisational learning. Existing studies often focus on technical design, user experience, or isolated learning outcomes, with insufficient attention to longitudinal impact, governance, ethical oversight, and alignment with organisational performance systems. Therefore, the purpose of this study is to examine and synthesise existing literature on the design, adoption, and impact of AI-based chatbots and intelligent agents within the contexts of monitoring, evaluation, customer support, and internal organisational operations.

Methodology

The study follows a dual approach, systematic literature reviews (SLR) and bibliometric analysis. SLR seek to identify publications that contain material of relevance to a research question or objectives and synthesise the outcomes of those publications, while bibliometric analysis focus on measurable publication patterns (e.g., publication counts, keyword co-occurrence, co-authorship trends, relevant sources). A PRISMA framework (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) is applied to improve the reporting quality, maintain transparency, reduce bias and improve the documentation of review protocol (Munn et al., 2018).

Eligibility Criteria

The eligibility criteria were set as studies and reports published in the last 5 years, from 2021 to 2025. The restrictions applied included, peer review journal articles, and conference papers. Only articles reported in the English language and in final publication stage were eligible for inclusion.

Information Sources and Search

AI and chatbot technologies are applied across multiple disciplines, and research on these topics is published and indexed in multidisciplinary databases. SCOPUS and Web of Science capture studies from a wide range of fields and are fully compatible with Biblioshiny used for bibliometric analysis. Therefore, both databases were used reducing the risk of database-specific bias and improving the completeness of the literature capture. The search string below was applied to both databases, and the retrieved records were subsequently merged, deduplicated, and screened to identify eligible publications for inclusion in the study: ("AI-driven chatbots" OR "intelligent agents" OR "conversational agents") AND ("organisational learning" OR "organizational learning") AND PUBYEAR > 2020 AND PUBYEAR < 2026 AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cp")) AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (PUBSTAGE, "final")).

Data Extraction

The initial search yielded 215 publications, which formed the dataset for the bibliometric analysis. During this process, all records meeting the basic inclusion criteria (peer-reviewed articles and conference papers, English language, relevance to AI/chatbots and intelligent systems) were retained to map publication trends, sources, and themes using Biblioshiny.

Afterwards a rigorous screening and eligibility process was applied to identify studies suitable for in-depth qualitative analysis. The PRISMA process is summarised in the Figure 5.3.

Figure 5.3

Data Extraction Process

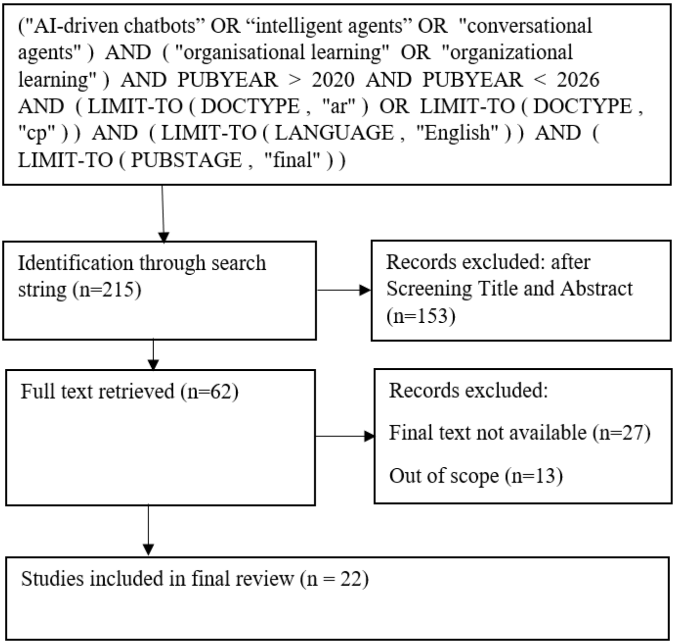


Table 5.1

Inclusion and Exclusion Criteria

Inclusion criteria	Exclusion criteria
Peer-reviewed publications	Non-Peer-reviewed publications
Publications published after December 2020	Studies published before 2021
Full text Journal and Conference articles	Non-English studies
Studies in their final stage of publications	
Studies in English	

Quality Assessment Criteria

According to García-Peñalvo (2022), it is important to evaluate the quality of the articles chosen for review. This study used the criteria presented in Table 5.2 to evaluate the quality of all the articles included in the review.

The evaluation was performed independently by the researchers. All the QA questions are measured with a rating of 1–3 (0 - Not good, 0.5 - Good, and 1 - Very good).

Table 5.2

Quality Assessment Criteria

Quality Assessment	Criteria
QA1	Inclusion and Exclusion criteria (Does the study meet it?)
QA2	Credible source (Is the study published in a recognised source?)
QA3	Relevant to research aim/question (Does the study involve AI-driven chatbots or intelligent agents?)
QA4	Evaluation (does the paper presents experimental or simulation-based performance evaluation and quantitatively/qualitatively analysed?)
QA5	Outcomes (Are the outcomes of the study aligned with it aim)

Reporting on the Evaluation Process

The purpose of the review process is to establish whether each publication is appropriate for the systematic review or not. The pre-defined checklist was designed to check the relevant aspects of the selected publication. With regards to QA2, articles published in journals were scored 1, while conference papers scored 0.5.

Table 5.3

Results after Quality Assessment

Publication	Type	QA1	QA2	QA3	QA4	QA5	Score
Marroquin & Senadji, 2025	Article	1	1	0.5	0.5	1	4
Burov et al., 2025	Conf. paper	1	0.5	1	0	0.5	3
Tsoi & Stronen, 2024	Conf. paper	1	1	1	0	0.5	3.5
Bartosiak & Modlinski, 2022	Article	1	1	1	1	1	5
Sachdeva et al., 2024	Article	1	1	0.5	0.5	1	4
Lee & Li, 2023	Article	1	1	1	1	1	5
Mukherjee & Chittipaka, 2022	Article	1	1	1	1	1	5
Al-Sharafi et al., 2023	Article	1	1	1	1	1	5
Chen & Gasco-Hernandez, 2024	Article	1	1	1	1	1	5
Poser et al., 2022	Article	1	1	1	0	1	4
Sofiyah et al., 2024	Article	1	1	1	1	1	5
Qiao et al., 2022	Conf. paper	1	0.5	1	1	1	4.5
Poser et al., 2022	Conf. paper	1	0.5	1	1	1	4.5
Flandrin et al., 2021	Conf. paper	1	0.5	1	1	1	4.5
Singh et al., 2021	Conf. paper	1	0.5	1	1	0.5	4
Alotaibi et al., 2022	Conf. paper	1	0.5	1	0	1	3.5
Huang et al., 2024	Article	1	1	1	1	1	5
Gkinko & Elbanna, 2023	Article	1	1	1	1	1	5
Martins et al., 2022	Article	1	1	1	0	1	4
Dube et al., 2024	Conf. paper	1	0.5	1	1	1	4.5
Terblanche & Tau, 2025	Article	1	1	1	1	1	5
Maragno et al., 2023	Article	1	1	1	0	1	4

Results and Discussion

This section provides the dataset results from bibliometric analysis. Table 5.4 provides an overview of the bibliometric characteristics of research on AI-driven chatbots and intelligent agents between 2021 and 2025. The dataset comprises 215 documents drawn from 150 sources, reflecting a broad and multidisciplinary research base (Flandrin et al., 2021; Huang et al., 2024; Maragno et al., 2023; Singh et al., 2021). An annual growth rate of 16.95% indicates rapid and sustained expansion of the field, while the low average document age (1.68 years) highlights its recent and evolving nature. The average of 20.27 citations per document suggests strong scholarly impact despite the field's youth. High keyword diversity (over 2,100 combined keywords) points to conceptual richness and thematic breadth. Authorship patterns show a highly collaborative research culture, with 733 authors, an average of 3.64 co-authors per document, and nearly 34% international collaboration. Journal articles dominate outputs, confirming the field's academic maturity.

Table 5.4

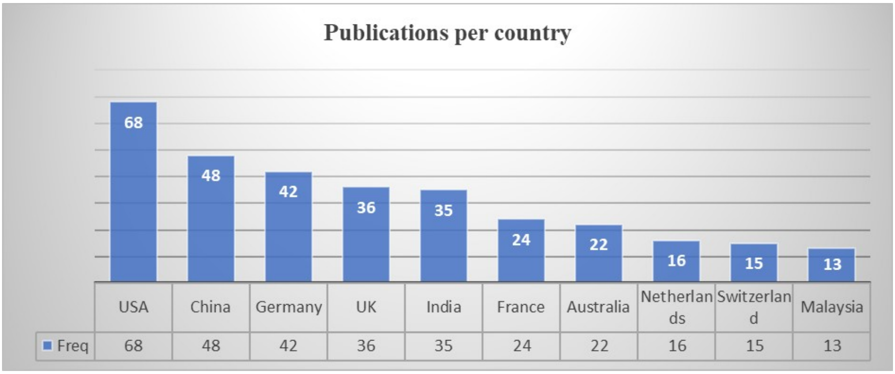
Main Information

Description	Results
Main information about data:	
Timespan	2021:2025
Sources (Journals, Books, etc.)	150
Documents	215
Annual growth rate, %	16.95
Document average age	1.68
Average citations per doc	20.27
References	1971
Document contents:	
Keywords plus (ID)	1174
Author's keywords (DE)	937
Authors:	
Authors	733
Authors of single-authored docs	13
Authors collaboration:	
Single-authored docs	13
Co-authors per doc	3.64
International co-authorships, %	33.95
Document types:	
Article	165
Conference paper	50

Figure 5.4 shows that research on AI-driven chatbots and intelligent agents is geographically diverse but dominated by leading countries.

Figure 5.4

Publication per Country



The United States leads with 68 publications, reflecting strong investment and research capacity in AI technologies. China and Germany follow, highlighting both rapid technological development and strong academic ecosystems. The United Kingdom and India also contribute substantially, indicating active engagement from both developed and emerging economies (Sachdeva et al., 2024). European countries such as France, the Netherlands, and Switzerland show consistent contributions, while Australia and Malaysia demonstrate growing regional participation. Overall, the distribution suggests global interest, with research activity concentrated in technologically advanced and innovation-driven countries.

Key Developments in AI-Driven Chatbot

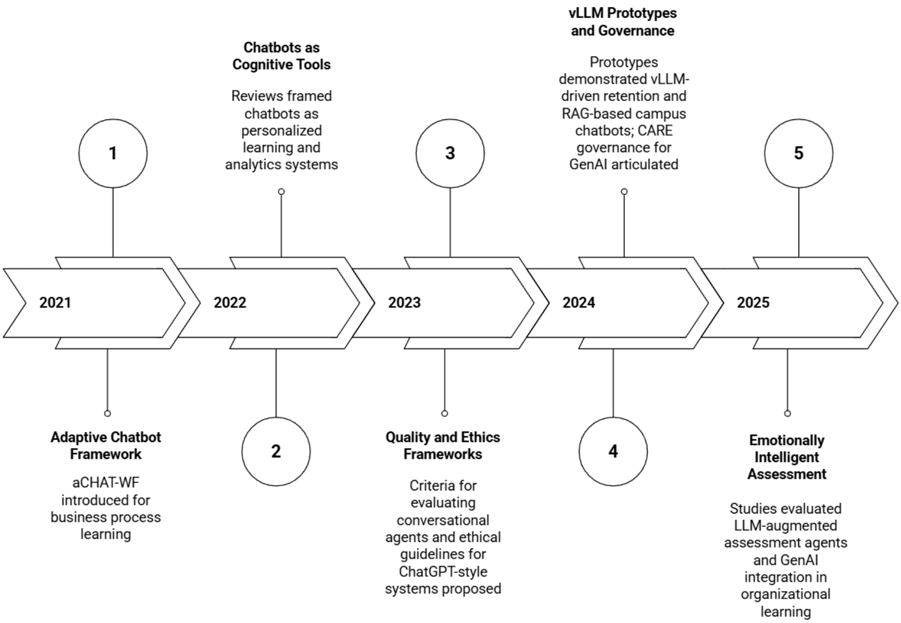
Figure 5.5 illustrates the progressive evolution of AI-driven chatbots between 2021 and 2025, highlighting a shift from basic automation to intelligent, learning-oriented systems (Flandrin et al., 2021; Qiao et al., 2022). In 2021-2022, research focused on adaptive chatbot frameworks that supported business process learning and personalised learning analytics, positioning chatbots as cognitive tools rather than simple interfaces (Lee & Li, 2023; Mukherjee & Chittipaka, 2022; Wilkinson et al., 2017).

By 2023, increased attention was given to quality, ethics, and governance, with studies proposing evaluation criteria to address bias, transparency, and responsible AI use in conversational systems (Bartosiak & Modlinski, 2022; Gkinko & Elbanna, 2023).

The 2024 phase marks a technological leap, characterised by vLLM prototypes and Retrieval-Augmented Generation (RAG), enabling context-aware monitoring, reporting, and evidence-based organisational learning (Chen & Gasco-Hernandez, 2024; Hutson & Plate, 2023; Sofiyah et al., 2024). By 2025, research converges on emotionally intelligent and LLM-augmented assessment agents integrated into organisational learning ecosystems, emphasising reflection, decision support, and continuous learning (Terblanche & Tau, 2025; Burov et al., 2025). The figure reflects increasing maturity, sophistication, and organisational embeddedness of conversational AI.

Figure 5.5

Key Developments in AI-Driven Chatbot (2021-2025)



Techniques in Building AI-Chatbots and Intelligent Agents

Table 5.5 summarizes core AI techniques, common architectures, and evaluation approaches used to build chatbots and agents for monitoring, evaluation, and organizational learning. It highlights the role of large language models, retrieval augmentation, dialogue management, agent modules, and evaluation frameworks.

Modern conversational AI systems combine core NLP and machine learning methods use aforementioned techniques to ground interactions in institutional documents and deliver reliable, domain-specific answers (Alotaibi et al., 2022; Singh et al., 2021). At the same time, advances in agent and conversational architectures separate dialogue flow, user modelling, and content management, enabling adaptive interactions that can be configured by non-technical domain experts.

Table 5.5
Core Techniques in Building AI-Chatbots and Intelligent Agents

Technique	Role in M&E and Organisational learning
Large language models and vLLMs	Natural language understanding and generation for tutoring, question and answers (Q&A), and adaptive feedback (uses both generation and scoring) (Dube et al., 2024; Klievtsova et al., 2023).
Retrieval-Augmented Generation (RAG)	Domain grounding and factual response generation from organizational corpora for assessments and reporting (Hutson & Plate, 2023).
Dialogue management and intent classification	Multi-turn control, adaptive sequencing of learning items, and user modeling for assessments and learning paths (Chen, 2024).
Agent architectures and multi-agent systems	Autonomous monitoring, notifications, and personal learning environments with role separation (author/learner/manager) (Burov et al., 2025).
Planning, rollouts, and reinforcement approaches	Proactive response planning and progression-aware actions to optimize dialog outcomes and task success (Bartosiak & Modlinski, 2022).
Evaluation and benchmark frameworks	Specialized pedagogical benchmarks, lifecycle quality criteria, and automated self-play or bot-bot methods for scalable evaluation (Klievtsova et al., 2023).

Applications Use Cases and Benefits
Deployment of AI-driven chatbots in post training or continuous learning, the intelligent systems apply spaced repetition, adaptive question generation, and personalized revision to improve knowledge retention after training sessions. In educational organisations, they present opportunities for conversation based assessments that can be used to evaluate student responses, and deliver tailored feedback, consequently reducing the academics workload by automating question prompts and scoring assistance (Al-Sharafi et al., 2023; Pypenko, 2024).

Conversational tools are also able to convert organisation documents, or spreadsheets into interactive query-driven interfaces allowing institutional teams to engage more with data (Burov et al., 2025). Unlike before, where employees used manual filtering or analysis to create static reports, now they can ask natural-language questions to dynamically explore performance indicators, monitor trends, and generate evaluation metrics on demand.

Organisation Domain Teaching and Workflow Training

Adaptive chatbots can support domain-specific teaching and workflow training by guiding users through business processes step by step and sequencing learning content based on the user's role, experience level, and progress (Klievtsova et al., 2023, 2025). By dynamically adjusting explanations, examples, and prompts, AI-chatbots enable personalised, just-in-time learning while supporting non-technical authors to configure conversation styles and training flows without requiring programming skills. With regards to knowledge management and communities of practice, conversational AI, can be integrated to facilitate experience sharing and informal knowledge exchange among practitioners (Tsoi & Stronen, 2024). By enabling users to query organisational knowledge in context, such as policies, best practices, lessons learned, and expert insights, conversational tools help surface tacit and explicit knowledge at the point of need, strengthening collective learning and continuous improvement.

Conclusion, Limitations and Future Studies

The power of AI is augmenting the capabilities of existing technological tools, enabling them to perform more intelligent, adaptive, and context-aware tasks. Recently, chatbots have been widely deployed across various sectors, with functionalities extending beyond simple question-answering.

This review reveals deployments of AI chatbots and intelligent systems in performance assessments, institutional reporting, front line support, and knowledge management. With these deployments, organisations get to experience benefits include customization of responses, faster data to insight, and improved scalability, with opportunities of having intelligent personal assistance. The study acknowledges the following limitations, the review was limited to studies published in English and sourced from two peer-reviewed academic databases (Scopus and Web of Science), excluding relevant findings from other sources. Even though search strategy was used, selection bias may still exist due to variations in database indexing, leading to exclusion of relevant studies with inaccessible full texts.

Future reviews could address these limitations by incorporating additional databases and considering multilingual studies to provide a more comprehensive and representative synthesis of the evidence. Future research should focus on the following three directions. Firstly, the exploration of hybrid architectures combining vLLMs with deterministic retrieval and rule layers to improve factual accuracy and auditability for M&E tasks. Secondly, human centric governance frameworks (such as CARE) to balance automation with accountability, responsiveness, and empowerment of users. Lastly, empirical longitudinal studies on organizational decision-making improvements, and cost benefit across deployments to establish evidence of impact.

References

Al-Sharafi, M. A., Al-Emran, M., Iranmanesh, M., Al-Qaysi, N., Iahad, N. A., & Arpaci, I. (2023). Understanding the impact of knowledge management factors on the sustainable use of AI-based chatbots for educational purposes using a hybrid SEM-ANN. *Interactive Learning Environments, 31*(10), 7491–7510. https://doi.org/10.1080/10494820.2022.2075014

Alotaibi, M., Alotaibi, M., Alamri, L., Alkadi, D., Alsahali, S., Aljameel, S., & Youldash, M. (2022). CAPEs advisory: A conversational agent based on NLP techniques for professional examinations advisory. *Proceedings of the Future Technologies Conference, 1288,* 755–768. https://doi.org/10.1007/978-3-030-63128-4_56

Bartosiak, M. L., & Modlinski, A. (2022). Fired by an algorithm? Exploration of conformism with biased intelligent decision support systems in the context of workplace discipline. *Career Development International, 27*(6/7), 601–615. https://doi.org/10.1108/CDI-06-2022-0170

Burov, O.Yu., Pasko, N. B., Viunenko, O. B., Agadzhanova, S. V, & Ahadzhanov-Honsales, K. H. (2025). Using intelligent agent-managers to build personal learning environments in the e-learning system. *7th International Workshop on Augmented Reality in Education,* 125–133. https://ceur-ws.org/Vol-3918/paper296.pdf

Chen, T., & Gasco-Hernandez, M. (2024). Uncovering the results of AI chatbot use in the public sector: Evidence from US state governments. *Public Performance & Management Review, 48*(6), 1331–1356. https://doi.org/10.1080/15309576.2024.2389864

Chen, Y. (2024). Enhancing language acquisition: The role of AI in facilitating effective language learning. *3rd International Conference on Humanities, Wisdom Education and Service Management (HWESM 2024),* 593–600. https://doi.org/10.2991/978-2-38476-253-8_71

Dube, M., Mutunhu Ndlovu, B., & Dube, S. (2024). Factors influencing the adoption of AI chatbots by non-governmental organizations. In *7th European Industrial Engineering and Operations Management Conference* (pp. 639–651). https://doi.org/10.46254/EU07.20240155

Flandrin, P., Hellemans, C., Van Der Linden, J., & Van De Leemput, C. (2021). Smart technologies in hospitality: effects on activity, work design and employment. A case study about chatbot usage. *Proceedings of the 17th "Ergonomie et Informatique Avancée" Conference, 2,* 1–11. https://doi.org/10.1145/3486812.3486838

García-Peñalvo, F. J. (2022). Developing robust state-of-the-art reports: Systematic literature reviews. *Education in the Knowledge Society, 23,* Article E28600. https://doi.org/10.14201/eks.28600

Gkinko, L., & Elbanna, A. (2023). The appropriation of conversational AI in the workplace: A taxonomy of AI chatbot users. *International Journal of Information Management, 69,* Article 102568. https://doi.org/10.1016/J.IJINFOMGT.2022.102568

Huang, H. W., Teng, D. C. E., & Tiangco, J. A. N. Z. (2024). The impact of AI chatbot-supported guided discovery learning on pre-service teachers' learning performance and motivation. *Journal of Science Education and Technology,* 1–15. https://doi.org/10.1007/S10956-024-10179-9/TABLES/7

Hutson, J., & Plate, D. (2023). Enhancing institutional assessment and reporting through conversational technologies: Exploring the potential of AI-powered tools and natural language processing. *DS Journal of Artificial Intelligence and Robotics, 1*(1), 11–22. https://doi.org/10.59232/air-v1i1p102

Klievtsova, N., Benzin, J. V., Kampik, T., Mangler, J., & Rinderle-Ma, S. (2023). Conversational process modelling: State of the art, applications, and implications in practice. In Di Francescomarino, C., Burattin, A. (Eds.), *Lecture Notes in Business Information Processing, Vol. 490*. Springer. https://doi.org/10.1007/978-3-031-41623-1_19

Klievtsova, N., Kampik, T., Mangler, J., & Rinderle-Ma, S. (2025). Conversationally actionable process model creation. In Comuzzi, M., Grigori, D., Sellami, M. (Eds.), *Lecture Notes in Computer Science, Vol. 15506*. Springer. https://doi.org/10.1007/978-3-031-81375-7_3

Lee, K. W., & Li, C. Y. (2023). It is not merely a chat: Transforming chatbot affordances into dual identification and loyalty. *Journal of Retailing and Consumer Services, 74,* Article 103447. https://doi.org/10.1016/j.jretconser.2023.103447

Maragno, G., Tangi, L., Gastaldi, L., & Benedetti, M. (2023). AI as an organizational agent to nurture: effectively introducing chatbots in public entities. *Public Management Review, 25*(11), 2135–2165. https://doi.org/10.1080/14719037.2022.2063935

Marroquin, E. M., & Senadji, B. (2025). Activity theory as framework for analysis of workplace learning technologies: the case of generative AI conversational agents. *The International Journal of Information and Learning Technology, 42*(4), 353–365. https://doi.org/10.1108/IJILT-07-2024-0141

Martins, I., Andrade, D., & Tumelero, C. (2022). Increasing customer service efficiency through artificial intelligence chatbot. *Revista de Gestao, 29*(3), 238–251. https://doi.org/10.1108/REGE-07-2021-0120

Melnyk, Y. B., & Pypenko, I. S. (2025). Implementing of artificial intelligence in a higher educational ecosystem. *International Journal of Science Annals, 8*(1), 13–20. https://doi.org/10.26697/ijsa.2025.1.1

Mukherjee, S., & Chittipaka, V. (2022). Analysing the adoption of intelligent agent technology in food supply chain management: An empirical evidence. *FIIB Business Review, 11*(4), 438–454. https://doi.org/10.1177/23197145211059243

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., Mcarthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medivcal Research Methodology, 18*(141). https://doi.org/10.1186/s12874-018-0611-x

Poser, M., Hackbarth, T., & Bittner, E. A. C. (2022). Don't throw it over the fence! Toward effective handover from conversational agents to service employees. *International Conference on Human-Computer Interaction, 13304 LNCS,* 531–545. https://doi.org/10.1007/978-3-031-05412-9_36

Pypenko, I. S. (2024). Benefits and challenges of using artificial intelligence by stakeholders in higher education. *International Journal of Science Annals, 7*(2), 28–33. https://doi.org/10.26697/ijsa.2024.2.7

Qiao, Q., Wu, W., & Li, Y. (2022). Enhancing consumer usage of AI-chatbots: The role of perceived humanness, social presence, and social interactivity. *Proceedings of the 8th International Conference on Information Management,* 97–103. https://doi.org/10.1109/ICIM56520.2022.00025

Sachdeva, A., Kim, A., & Dennis, A. R. (2024). Taking the chat out of chatbot? Collecting user reviews with chatbots and web forms. *Journal of Management Information Systems, 41*(1), 146–177. https://doi.org/10.1080/07421222.2023.2301175

Singh, H., Cascini, G., & McComb, C. (2021). Comparing virtual and face-to-face team collaboration: Insights from an agent-based simulation. *Proceedings of the ASME Design Engineering Technical Conference, 6,* V006T06A022. https://doi.org/10.1115/DETC2021-66043

Sofiyah, F. R., Dilham, A., Hutagalun, A. Q., Yulinda, Y., Lubis, A. S., & Marpaung, J. L. (2024). The chatbot artificial intelligence as the alternative customer services strategic to improve the customer relationship management in real-time responses. *International Journal of Economics and Business Research, 27*(5). https://doi.org/10.1504/IJEBR.2024.10064925

Terblanche, N., & Tau, T. (2025). Article Industry and Higher Education. *Industry and Higher Education, 39*(3), 279–290. https://doi.org/10.1177/09504222241287090

Tsoi, J. C. H., & Strønen, F. (2024). Integration of conversational AI capabilities in knowledge management processes for higher education. *Proceedings of the European Conference on Knowledge Management, ECKM,* 2024-Septe, 1026–1033. https://doi.org/10.34190/eckm.25.1.2659

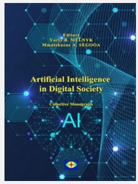
Wilkinson, A., Pettifor, A., Rosenberg, M., Halpern, C. T., Thirumurthy, H., Collinson, M. A., & Kahn, K. (2017). The employment environment for youth in rural South Africa: A mixed-methods study. *Development Southern Africa, 34*(1), 17–32. https://doi.org/10.1080/0376835X.2016.1259986

Information about the authors:
Kgopa Alfred Thaga – https://orcid.org/0000-0001-5455-7064; PhD (Informatics), Dr, Senior Lecturer, University of South Africa, Roodepoort, South Africa.
Msweli Nkosikhona Theoren – https://orcid.org/0000-0003-4709-0763; PhD (Information Systems), Dr, Senior Lecturer, University of South Africa, Roodepoort, South Africa.

Check for updates

Artificial Intelligence in Digital Society,
Volume 1, 2026

DOI: 10.26697/9786177089192.2026
ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)



Chapter 6. The Use of Artificial Intelligence-Based Chatbots to Promote the Sustainability of South African Small and Medium Enterprises in the Digital Era

Makelana P. [1]

[1] *Vaal University of Technology, South Africa*

Received: *08.12.2025;* ***Accepted:*** *10.02.2026;* ***Published:*** *10.03.2026*

Abstract

The advancement of artificial intelligence (AI) applications, such as AI chatbots, enables small and medium enterprises (SMEs) to enhance competitiveness, operational performance, and digital marketing capabilities. However, utilisation among SMEs in developing countries remains limited. Addressing this gap, the study applies technology-organization-environment (TOE), technology acceptance model (TAM), and diffusion of innovation (DOI) frameworks to identify factors influencing AI chatbot utilisation among SMEs in South Africa. A quantitative method was employed: a self-administered questionnaire was distributed to 300 SMEs, and data were analyzed using SEM. Results showed relative advantage, compatibility, top management support, organisational readiness, ethical AI regulation, perceived usefulness, and ease of use were highly significant; whilst security is less significant. The study makes a contribution by developing a model that explains factors influencing AI chatbot utilisation among SMEs.

Keywords: artificial intelligence chatbots, diffusion of innovations, small and medium enterprises, technology-organisation-environment, technology acceptance model.

Cite this chapter as:

Makelana, P. (2026). The use of artificial intelligence-based chatbots to promote the sustainability of South African small and medium enterprises in the digital era. In Y. B. Melnyk & M. A. Segooa (Eds.), *Artificial Intelligence in Digital Society, Vol. 1.* (pp. 87–101). KRPOCH. https://doi.org/10.26697/aids.2026.6

The electronic version of this chapter is complete. It can be found online in the AIDS Archive https://doi.org/10.26697/aids

Introduction

The digital era has created both opportunities and challenges for organisations around the world (George, 2024; Marganaha, 2024). Artificial intelligence (AI) is at the centre of this change (Wang, Lin & Shao, 2023; Badghish & Soomro, 2024; Hamida, 2025). One important example of AI is AI chatbots (Kedi, Ejimuda, Idemudia & Ijomah, 2024). Organisations use chatbots to provide personalised services, automate tasks, and communicate with customers (Wang, Lin & Shao, 2023). According to a report by Grand View Research (2025), the global chatbot market was USD 7.76 billion in 2024 and is expected to grow to USD 27.29 billion by 2030. A report published by Gartner (2023) predicts that by 2026, more than 80% of organisations will use AI through apps or programming models (Rana, Pillai, Sivathanu & Malik, 2024). This fast growth of AI chatbots offers small and medium enterprises (SMEs) the chance to improve their operations and provide better customer service.

Identified Gaps and the Purpose of the Study

AI is vital for organisational performance and survival (Muzuva, Zhou & Zondo, 2024; Wang et al., 2023). Research has examined AI adoption factors and implementation in countries such as Germany, China, United kingdom, and Malaysia (Ulrich & Frank, 2021; Liang & Hongtao, 2023; Mathagu, 2024; Roszelan & Shahron, 2025). Nevertheless, studies on AI chatbot use in SMEs, especially in developing countries like South Africa, remain limited (Shekgola & Modiba, 2025).

Research Questions

The present study addresses the following questions as follows:
- What are the factors influencing the utilisation of AI chatbots in SMEs?
- Does the utilisation of AI chatbots influence the performance of SMEs?

Literature Review

Small and Medium Enterprises (SMEs) in South Africa

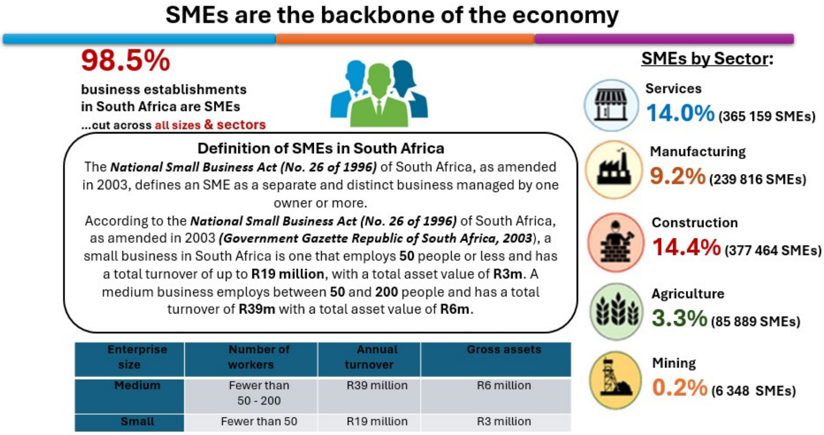
The National Small Business Act (NSB) No. 26 of 1996, amended in 2003, defines SMEs as organisations with 50 to 200 employees, an annual turnover of up to R39 million, and gross assets of up to R6 million (Madzimure, Mafini & Dhurup, 2020). Figure 6.1 illustrates this definition of SMEs.

The Importance of SMEs to the South African Economy

SMEs are key drivers of economic growth, playing a vital role in reducing poverty, and creating jobs (Bvuma & Marnewick, 2020). They make up about 98.5% of businesses, contribute 39% 57% to the gross domestic product (GDP), and provide 60% of employment (Mhlongo & Daya, 2023). SMEs are crucial for achieving the National Development Plan (NDP) targets of higher GDP and job creation by 2030 (Matekenya & Moyo, 2022). Despite their importance, SMEs in South Africa face numerous challenges. Figure 6.2 illustrates employment targets aligned with NDP goals.

Figure 6.1

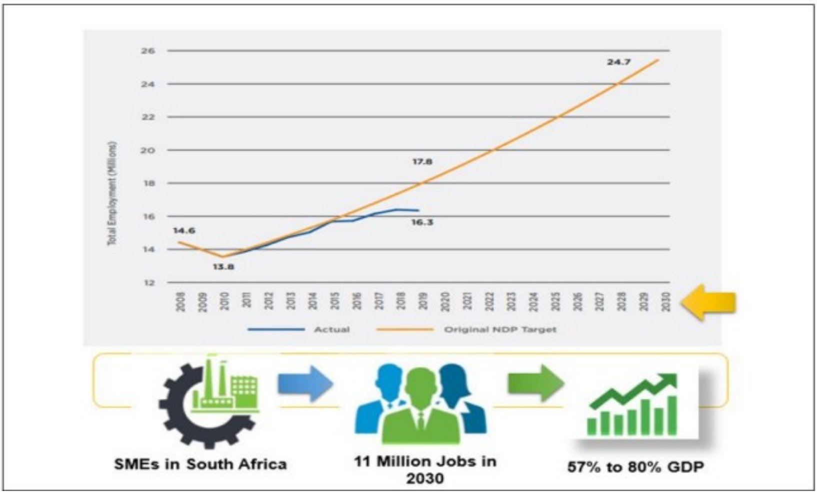
Definition of SMEs in South Africa (Adapted)



Note. Adapted from "Adoption of fourth industrial revolution 4.0 among Malaysian small and medium enterprises (SMEs)", by A. Shahzad et al., 2023, *Humanities and Social Sciences Communications, 10*, Article 693, p. 4 (https://doi.org/10.1057/s41599-023-02076-0). Copyright 2023 by Springer.

Figure 6.2

South Africa's Employment Targets



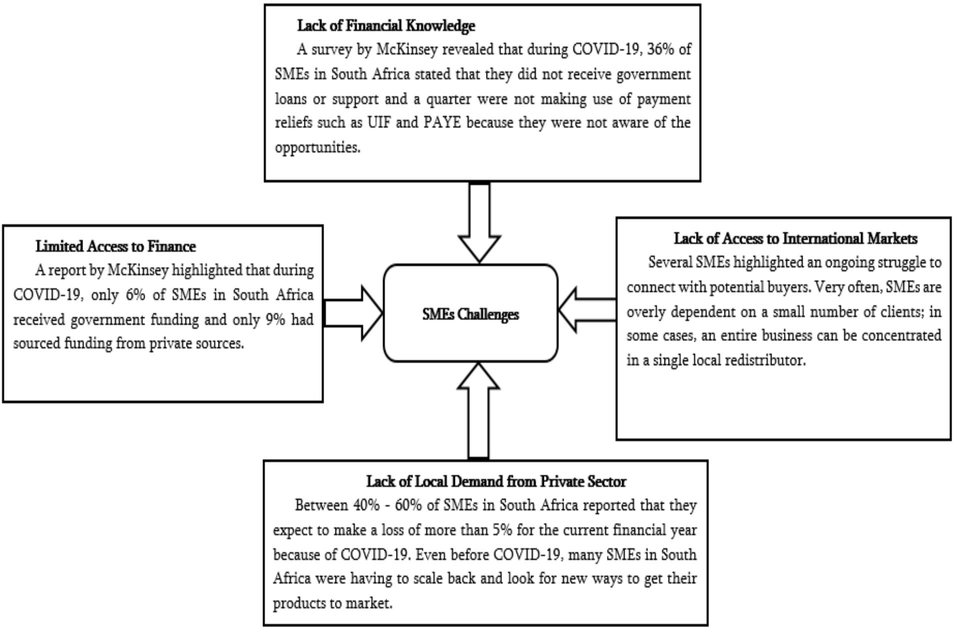
Note. From "Economic progress towards the national development plan's vision 2030", by National Planning Commission, 2020 (https://www.nationalplanningcommission.org.za). Copyright 2020 by NPC.

Challenges Faced by SMEs in South Africa

Many scholars have explored challenges faced by SMEs globally and in South Africa. For instance, Achieng and Malatji (2022) conducted a scoping review on digital transformation in sub-Saharan African SMEs, highlighting limited financial support and insufficient digital skills. Similarly, Etim and Daramola (2020) note that South African SMEs face lack of access to global markets, limited finance, and low ICT awareness. Figure 6.3 illustrates key challenges affecting SMEs in South Africa.

Figure 6.3

SMEs Challenges in South Africa (Adapted)



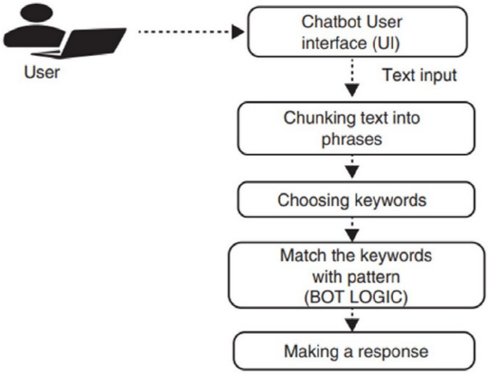
Note. Adapted from "Digital transformation of small and medium enterprises in sub-Saharan Africa: A scoping review", by M. S. Achieng & M. Malatji, 2022, *The Journal for Transdisciplinary Research in Southern Africa, 18*(1) (https://doi.org/10.4102/td.v18i1.1257). Copyright 2022 by AOSIS (Pty) Ltd.

The Importance of Using AI chatbots in SMEs

In the era of digital transformation, AI has emerged as a pivotal marketing tool for organizations of all sizes, especially for SMEs (Laki & Miklosik, 2025). The word "bot" in "Chatbots" is short for "robot", implying that chatbots are computer programs or systems designed to simulate human conversation (Alboqami, 2023). Figure 6.4 shows how chatbots operate.

Figure 6.4

How Chatbots Operate



Note. From "SMEs' adoption of artificial intelligence-chatbots for marketing communication: A conceptual framework for an emerging economy", by S. S. M. Mokhtar & M. G. Salimon, 2022, *Marketing Communications and Brand Development in Emerging Markets: Volume II.* (https://doi.org/10.1007/978-3-030-95581-6_2). Copyright 2022 by the Author(s), under exclusive license to Springer Nature Switzerland AG.

A chatbot is an AI-powered tool that interacts with customers to understand their needs (Melnyk & Pypenko, 2023). Sharma, Singh, Islam, and Dhir (2024) note that AI chatbots can revolutionize SMEs by boosting competitiveness, operational performance, and digital marketing.

Similarly, Laki and Miklosik (2025) argue that AI chatbots help SMEs automate tasks, optimize marketing strategies, and cut operational costs. Supporting this, Kedi et al. (2024) observe that AI chatbots also assist SMEs in sharing information with customers and employees.

Conceptual Framework and Hypotheses Development

The present study employs the technology-organisation-environment (TOE) framework, technology acceptance model (TAM), and diffusion of innovation (DOI) theory to explain SMEs' use of AI chatbots.

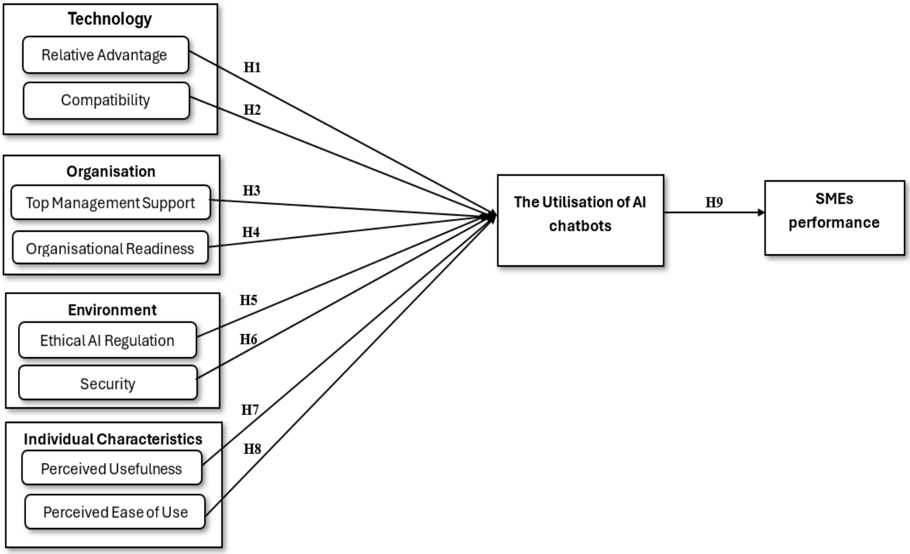
TOE, introduced by Depietro, Wiarda, and Fleischer in 1990, explains why organisations adopt new technologies through technology, organisation, and environment and has been applied on AI adoption (Badghish & Soomro, 2024; Almashawreh, Talukder, Charath & Khan, 2024).

DOI initially introduced by Rogers in 2003, explains adoption via relative advantage, compatibility, trialability, complexity, and observability (Badghish & Soomro, 2024; Sanchez, Calderon & Herrera, 2025). TAM introduced by Davis in

1989, examines user behaviour through perceived ease of use and usefulness and has been applied on AI adoption (Erraoui & Amine, 2024). Figure 6.5 shows AI chatbots utilisation is influenced by technological, organisational, environmental, and individual characteristics.

Figure 6.5

Conceptual Framework



Technology

The utilisation of AI chatbots is influenced by technological factors such as relative advantage and compatibility. Relative advantage refers to the perceived superiority of a new technology over existing ones (Rogers, 2003) and has been shown to positively influence technology adoption (Bhardwaj, Garg & Gajpal, 2021). Compatibility with existing organisational systems and processes further increases adoption among SMEs (Badghish & Soomro, 2024). Based on these claims, we propose the following hypotheses:

H1: Relative advantage influences the utilisation of AI chatbots.

H2: Compatibility influences the utilisation of AI chatbots.

Organisation

AI chatbot adoption is influenced by organisational factors, notably top management support, which is senior leaders' backing of technology use (Mathagu, 2024) and increases implementation success (Siradhana and Arora, 2024). Organisational readiness to adopt new technology, requires SMEs to have sufficient financial, technological, and skilled human resources (Badghish & Soomro, 2024). Consequently, the following hypotheses are proposed:

H3: Top management support influences the utilisation of AI chatbots.

H4: Organisational readiness influences the utilisation of AI chatbots.

Environment

The use of AI chatbots is influenced by technological factors, particularly ethical AI regulation (Omonov & Ahn, 2025). Ethical AI regulation refers to laws and frameworks that ensure AI use aligns with ethical principles (Cajueiro & Celestino, 2025). Ethical principles support transparency, fairness, accountability, and effective operation of AI chatbots, supported by strong security measures (Omonov & Ahn, 2025). Consequently, the following hypotheses are formulated:

H5: Ethical AI regulation influences the utilisation of AI chatbots.

H6: Security influences the utilisation of AI chatbots.

Individual Characteristics

Individual traits, especially perceived usefulness and ease of use, affect AI chatbot use. Perceived usefulness is how much a user thinks a technology improves performance, while ease of use is how much it reduces effort (Davis, 1989). These factors are related, as easier technologies are often seen as more useful (Bhardwaj et al., 2021). Users are more likely to adopt AI chatbots if they see them as helpful and effortless (Omonov & Ahn, 2025). Based on this, we propose the following hypotheses:

H7: Perceived usefulness of AI chatbots influences the utilisation of AI chatbots.

H8: Perceived ease of use influences the utilisation of AI chatbots.

The Utilisation of AI chatbots and SMEs performance

Considering SMEs' resource constraints and limited technical capabilities, AI chatbot utilisation overcomes these limits and improves business processes (Sharma, Singh, Islam & Dhir, 2024). Similarly, a study by Omonov & Ahn (2025) found that AI chatbots enhance organisational performance. Therefore, the following hypothesis is formulated:

H9: The utilisation of AI chatbots influences SMEs performance.

Methodology

Research Design and Approach

SMEs conducting business in Gauteng province, South Africa, were primarily targeted. The present study employed a quantitative approach to empirically test the hypotheses derived from the conceptual framework.

A cross-sectional survey was utilised, meaning data was gathered from respondents at a single point in time. The study aligns with the positivist paradigm, which posits that social phenomena can be examined objectively and that variable relationships can be measured and validated through statistical analysis (Saunders, Lewis & Thornhill, 2019).

Data Collection and Analysis

To collect data from SMEs, a closed-ended questionnaire was developed and physically distributed. According to Saunders et al. (2019), closed-ended questions are more specific and less prone to interpretation and verbosity than open-ended ones. Items were measured on a 5-point Likert scale (1 = strongly disagree, 2 = disagree, 4 = agree, 5 = strongly agree). Of 500 questionnaires distributed, 300 were returned. Data were analysed using SPSS version 28.

Results

As depicted in Table 6.1, the percentage of male respondents was 63.0% and 37.7% were female. Age wise, the study results show that 74.3% of respondents were from the 20 to 29 age group, 21.3% from the 30 to 39 group and 4.7% were 50 years and above. Furthermore, Table 6.1 illustrates that 56.7% of respondents have a BTech degree, 19.3% have a master's degree, 15.3% have a diploma, and 2.0% have a PhD.

Table 6.1

Demographic Profile

Variables		Frequency	Percentage
Gender	Male	187	63.3
	Female	113	37.7
	Total	300	100.0
Age	20-29 years	222	74.3
	30-39 years	64	21.3
	50 years and above	14	4.7
	Total	300	100.0
Education	Matric	20	6.7
	Diploma	46	15.3
	B-tech	170	56.7
	Master's	58	19.3
	PhD	6	2.0
	Total	300	100.0

Assessment of Measurement Model

The measurement model was examined using factor loadings (FL), composite reliability (CR), and average variance extracted (AVE) (Hiar, Ringle, Gudergan, Fischer, Nitzl & Menictas, 2019).

FL, CR, and AVE should exceed 0.7, 0.5, and 0.7, respectively. Table 6.2 shows all constructs meet these thresholds, confirming acceptable convergent validity.

Table 6.2
Loadings Reliability and Validity Statistics

Construct	Item	Outer Loading	FL	CR	AVE
Relative Advantage (RA)	RA1	0.725	0.778	0.955	0.725
	RA2	0.826			
	RA3	0.950			
Compatibility (CP)	CP1	0.766	0.862	0.739	0.885
	CP2	0.856			
	CP3	0.925			
Top Management Support (TMS)	TMS1	0.715	0.970	0.778	0.988
	TMS2	0.886			
	TMS3	0.795			
Organisational Readiness	OR1	0.855	0.773	0.755	0.825
	OR2	0.856			
	OR3	0.775			
Ethical AI Regulation	EAR1	0.875	0.778	0.834	0.775
	EAR2	0.954			
	EAR3	0.774			
Security	SC1	0.854	0.765	0.856	0.905
	SC2	0.855			
	SC3	0.725			
Perceived Usefulness	PU1	0.792	0.752	0.772	0.925
	PU2	0.835			
	PU3	0.712			
Perceived Ease of Use	PEU1	0.885	0.916	0.854	0.755
	PEU2	0.793			
	PEU3	0.785			
Utilisation of AI chatbots	UAC1	0.775	0.778	0.849	0.753
	UAC2	0.819			

Assessment of Structural Model

In this section, the structural equation model (SEM) was used to test the hypotheses. Figure 6.6 presents the simplified structural model with hypothesized relationships among latent variables. Figure 6.6 shows that eight of the nine paths are significant.

As shown in Table 6.3, eight hypotheses including relative advantage (H1, $p<0.05$), compatibility (H2, $p<0.05$), top management support (H3, $p<0.05$), organisational readiness (H4, $p<0.05$), ethical AI regulation (H5, $p<0.05$), perceived usefulness (H7, $p<0.05$), perceived ease of use (H8, $p<0.05$) and AI chatbot utilisation (H9, $p<0.05$) are accepted, indicating these factors significantly affect utilisation. Security (H6, $p>0.05$) is rejected.

Figure 6.6

Structural Model

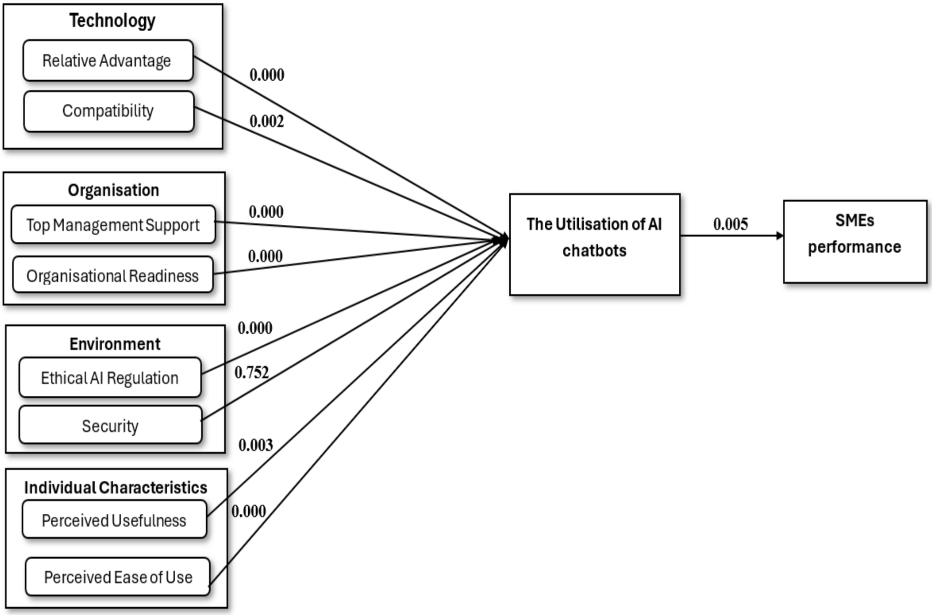


Table 6.3

Hypotheses Testing

Constructs	Std. Beta (β)	T-Values	*p* values	Results
H1 Relative advantage→ utilisation of AI chatbots	0.178	0.245	0.000	Accepted
H2 Compatibility→ utilisation of AI chatbots	0.142	0.344	0.002	Accepted
H3 Top Management Support→ utilisation of AI chatbots	0.373	1.286	0.000	Accepted
H4 Organisational Readiness→ utilisation of AI chatbots	0.442	0.142	0.000	Accepted
H5 Ethical AI Regulation → utilisation of AI chatbots	0.347	1.772	0.000	Accepted
H6 Security→ utilisation of AI chatbots	0.135	2.552	0.752	Rejected
H7 Perceived Usefulness→ utilisation of AI chatbots	0.268	1.071	0.003	Accepted
H8 Perceived Ease of Use→ utilisation of AI chatbots	0.112	4.357	0.000	Accepted
H9 Utilisation of AI chatbots→ SMEs performance	0.274	0.173	0.005	Accepted

Discussion

The present study aimed to find and explain the factors that affect the use of AI chatbots by SMEs in South Africa. For technological factors, relative advantage has a positive effect on AI chatbot use ($p=0.000<0.05$; H1 accepted). This means the benefits of AI chatbots encourage SMEs, especially in South Africa, to use them (Siradhana & Arora, 2024; Badghish & Soomro, 2024).

Compatibility was also positive and significant ($p=0.002<0.05$; H2 accepted), showing AI works well with existing SME systems and is easy to adopt (Omonov & Ahn, 2025).

For organisational factors, top management support was significant ($p=0.000<0.05$; H3 accepted), showing companies adopt new technology when management supports it (Siradhana & Arora, 2024).

Organisational readiness was significant ($p=0.000<0.05$; H4 accepted), meaning use is higher when resources and skills are available.

In terms of environmental factors, ethical AI regulation was significant ($p=0.000<0.05$; H5 accepted), meaning SMEs follow rules.

Security was not significant ($p=0.752>0.05$; H6 rejected), suggesting weak protections make SMEs less likely to use AI (Omonov & Ahn, 2025).

For individual characteristics, perceived usefulness ($p=0.003<0.05$; H7 accepted) and perceived ease of use ($p=0.000<0.05$; H8 accepted) positively affected use.

Finally, AI chatbots had a significant impact on SMEs' performance ($p=0.005<0.05$; H9 accepted), in line with earlier studies (Badghish & Soomro, 2024).

Theoretical and Practical Contribution

The study makes a theoretical contribution by developing a model that integrates TOE, TAM, and DOI to explain AI chatbot utilisation and its impact on SME performance.

It also makes a practical contribution by helping SME managers understand limited AI chatbot use and assess factors influencing adoption, particularly in developing countries. The model addresses technological, organisational, and environmental factors and individual characteristics.

Conclusion

The growth of AI has driven scholars to examine its impact on organisational performance using TOE, TAM, and DOI. This study proposes a model to explain factors influencing AI chatbot utilisation among SMEs in South Africa. The study results show that relative advantage, compatibility, top management support, organisational readiness, ethical AI regulation, perceived usefulness, and perceived ease of use significantly influence chatbot utilisation, while security is less significant. The study concludes that AI chatbots are essential for improving SME performance.

Ethical Approval

The study obtained ethical clearance from the institution's Ethics Committee (Ref no. FCRE/ICT/2022/03/001 (1).

References

Achieng, M. S., & Malatji, M. (2022). Digital transformation of small and medium enterprises in sub-Saharan Africa: A scoping review. *The Journal for Transdisciplinary Research in Southern Africa, 18*(1), 1–13. https://doi.org/10.4102/td.v18i1.1257

Alboqami, H. (2023). Factors affecting consumers adoption of AI-based chatbots: The role of anthropomorphism. *American Journal of Industrial and Business Management, 13,* 195–214. https://doi.org/10.4236/ajibm.2023.134014

Almashawreh, R., Talukder, M., Charath, & Khan, M. (2024). AI adoption in Jordanian SMEs: The influence of technological and organizational orientations. *Global Business Review,* 1–29. https://doi.org/10.1177/09721509241250273

Badghish, S., & Soomro, Y. A. (2024). Artificial intelligence adoption by SMEs to achieve sustainable business performance: Application of technology–organization–environment framework. *Sustainability (Switzerland), 16,* 1–24. https://doi.org/10.3390/su16051864

Bhardwaj, A. K., Garg, A., & Gajpal, Y. (2021). Determinants of blockchain technology adoption in supply chains by small and medium enterprises (SMEs) in India. *Mathematical Problems in Engineering,* 1–14. https://doi.org/10.1155/2021/5537395

Bvuma, S., & Marnewick, C. (2020). An information and communication technology adoption framework for small, medium and micro-enterprises operating in townships South Africa. *Southern African Journal of Entrepreneurship and Small Business Management, 12*(1), 1–12. https://doi.org/10.4102/sajesbm.v12i1.318

Cajueiro, D. O., & Celestino, V. R. R. (2025). A comprehensive review of artificial intelligence regulation: Weighing ethical principles and innovation. *Journal of Economy and Technology, 4,* 77–91. https://doi.org/10.1016/j.ject.2025.07.001

Chatterjee, S., Rana, N. P., Dwivedi, Y. K., & Baabdullah, A. (2021). Understanding AI adoption in manufacturing and production firms using an integrated TAM-TOE model. *Technological Forecasting and Social Change, 170,* Article 120880. https://doi.org/10.1016/j.techfore.2021.120880

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly: Management Information Systems, 13*(3), 319–339. https://doi.org/10.2307/249008

Depietro, R., Wiarda, E., & Fleischer, M. (1990). The context for change: Organization, technology and environment. In *The processes of technological innovation*. Lexington, MA: Lexington Books. https://doi.org/10.7251/emc2201237t

Erraoui, S., & Amine, A. (2024). Proposal of technology acceptance model: Adoption of artificial intelligence in Moroccan SMEs. *European Journal of Economic and Financial Research, 8*(6), 79–94. https://doi.org/10.46827/ejefr.v8i6.1858

Etim, E., & Daramola, O. (2020). The informal sector and economic growth of South Africa and Nigeria: A comparative systematic review. *Journal of Open Innovation: Technology, Market, and Complexity, 6*(4), 1–26. https://doi.org/10.3390/joitmc6040134

Gartner. (2023, October 11). Gartner says more than 80% of enterprises will have used generative AI APIs or deployed generative AI-enabled applications by 2026. https://www.gartner.com/en/newsroom/press-releases/2023-10-11-gartner-says-more-than-80-percent-of-enterprises-will-have-used-generative-ai-apis-or-deployed-generative-ai-enabled-applications-by-2026

George, A. (2024). Digital transformation in business: Opportunities, challenges, and implications. *Partners Universal Innovative Research Publication (PUIRP), 2(*6), 46–54. https://doi.org/10.5281/zenodo.14599717

Grand View Research. (2025). *South Africa social media analytics market size & outlook.* https://www.grandviewresearch.com/horizon/outlook/social-media-analytics-market/south-africa

Hair, J. F., Ringle, C. M., Gudergan, S. P., Fischer, A., Nitzl, C., & Menictas, C. (2019). Partial least squares structural equation modeling-based discrete choice modeling: An illustration in modeling retailer choice. *Business Research, 12*(1), 115–142. https://doi.org/10.1007/s40685-018-0072-4

Hamida, A. G. (2025). Adoption of artificial intelligence technology by SMEs: Impact on customized e-marketing strategies and online purchase. *International Conference on Technology Enabled Economic Changes (InTech),* 1319–1324. https://doi.org/10.1109/intech64186.2025.11198535

Kedi, W.E., Ejimuda, C., Idemudia, C., & Ijomah, T. (2024). AI Chatbot integration in SME marketing platforms: Improving customer interaction and service efficiency. *International Journal of Management & Entrepreneurship Research, 6*(7), 2332–2341. https://doi.org/10.51594/ijmer.v6i7.1327

Laki, K., & Miklosik, A. (2025). Leveraging AI-powered social media platforms to enhance customer engagement and drive sales growth in Uganda's SMEs. *Marketing Science & Inspirations, 20*(3), 46–58. https://doi.org/10.46286/msi.2025.20.3.5

Liang, L. Y., & Hongtao, L. (2023). The factors influencing the adoption of AI in e-commerce by SMEs in Shandong province. *International Journal of*

Research and Innovation in Applied Science, 10(6), 60–66. https://doi.org/10.51584/IJRIAS

Madzimure, J., Mafini, C., & Dhurup, M. (2020). E-procurement, supplier integration and supply chain performance in small and medium enterprises in South Africa. *South African Journal of Business Management, 51*(1), 2078–5585. https://doi.org/10.4102/SAJBM.V51I1.1838

Marganaha, H. (2024). Business development in the digital age. *International Journal of Advanced Science and Computer Applications, 3*(2), 1–4. https://doi.org/10.47679/ijasca.v3i2.55

Matekenya, W., & Moyo, C. (2022). Innovation as a driver of SMME performance in South Africa: A quantile regression approach. *African Journal of Economic and Management Studies*, *13*(3), 452–467. https://doi.org/10.1108/AJEMS-06-2021-0306

Mathagu, S. (2024). Artificial intelligence in small and medium enterprises – An empirical analysis of critical factors. *Premier Journal of Science, 1*, Article 100009. https://doi.org/10.70389/pjs.100009

Melnyk, Y. B., & Pypenko, I. S. (2023). The legitimacy of artificial intelligence and the role of ChatBots in scientific publications. *International Journal of Science Annals, 6*(1), 5–10. https://doi.org/10.26697/ijsa.2023.1.1

Mhlongo, T., & Daya, P. (2023). Challenges faced by small, medium and micro enterprises in Gauteng: A case for entrepreneurial leadership as an essential tool for success. *Southern African Journal of Entrepreneurship and Small Business Management, 15*(1), 1–12. https://doi.org/10.4102/sajesbm.v15i1.591

Mokhtar, S. S. M., & Salimon, M. G. (2022). SMEs' adoption of artificial intelligence-chatbots for marketing communication: A conceptual framework for an emerging economy. In Adeola, O., Hinson, R. E., Sakkthivel, A. M. (Eds.), *Marketing Communications and Brand Development in Emerging Markets: Volume II. Palgrave Studies of Marketing in Emerging Economies* (pp. 25–53). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-95581-6_2

Muzuva, M., Zhou, H., & Zondo, R. W. (2024). Has generative AI become of age: Assessing its impact on the productivity of SMEs in South Africa. *International Journal of Research in Business and Social Science, 13*(7), 2147–4478. https://doi.org/10.20525/ijrbs.v13i7.3576

National Planning Commission. (2020, December). *Economic progress towards the national development plan's vision 2030.* https://www.nationalplanningcommission.org.za/publications_reports

Omonov, M. S., & Ahn, Y. (2025). Towards smart public administration: A TOE-based empirical study of AI chatbot adoption in a transitioning government context. *Administrative Sciences, 15*(8), 1–29. https://doi.org/10.3390/admsci15080324

Rogers, E. M. (2003). *Diffusion of innovations* (3rd ed.). The Free Press. https://www.scribd.com/document/341985724/Everett-m-Rogers-Diffusion-of-Innovations?language_settings_changed=English

Roszelan, A. I., & Shahron, M. (2025). Readiness for artificial intelligence adoption in Malaysian manufacturing companies. *Journal of Information Technology Management, 17*(1), 1–13. https://doi.org/10.22059/jitm.2025.99920

Sanchez, E., Calderon, R., & Herrera, F. (2025). Artificial intelligence adoption in SMEs: Survey based on TOE–DOI framework, primary methodology and challenges. *Applied Sciences, 15,* 1–43. https://doi.org/10.3390/app15126465

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students* (5th ed.). Pearson Education Limited. https://anyflip.com/ftteb/yphw/basic/

Shahzad, A., Zakaria, S.A., Kotzab, H., Makki, M.A., Hussain, A., & Fischer, J. (2023). Adoption of fourth industrial revolution 4.0 among Malaysian small and medium enterprises (SMEs). *Humanities and Social Sciences Communications, 10*, Article 693. https://doi.org/10.1057/s41599-023-02076-0

Sharma, S., Singh, G., Islam, N., & Dhir, A. (2024). Why do SMEs adopt artificial intelligence-based chatbots? *IEEE Transactions on Engineering Management, 71,* 1773–1786. https://doi.org/10.1109/TEM.2022.3203469

Shekgola, M., & Modiba, M. (2025). Utilising an AI chatbot to support smart digital government for Society 5.0 in South Africa. *South African Journal of Information Management, 27*(1), 1–10. https://doi.org/10.4102/sajim.v27i1.1983

Siradhana, N. K., & Arora, R. (2024). Examining the influence of artificial intelligence implementation in HRM practices using T-O-E model. *Vision: The Journal of Business Perspective.* https://doi.org/10.1177/09722629241231458

Ulrich, P., & Frank, V. (2021). Relevance and adoption of AI technologies in German SMEs – Results from survey-based research. *25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems,* 2152–2159. https://doi.org/10.1016/j.procs.2021.08.228

Wang, X., Lin, X., & Shao, B. (2023). Artificial intelligence changes the way we work: A close look at innovating with chatbots. *Journal of the Association for Information Science and Technology, 74*(3), 339–353. https://doi.org/10.1002/asi.24621

Information about the author:

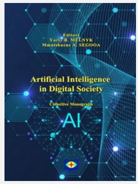
Makelana Phenuel – https://orcid.org/0000-0003-0986-1117; Doctor of Computing, Lecturer, Department of Computer Sciences, Vaal University of Technology, Vanderbiljpark, South Africa.

PART IV

ARTIFICIAL INTELLIGENCE IN PRACTICE: INNOVATION, INTEGRATION, AND MANAGEMENT

Artificial Intelligence in Digital Society,
Volume 1, 2026
DOI: 10.26697/9786177089192.2026
ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)

Check for updates



Chapter 7. Artificial Intelligence as the Engine of Invention: Revolutionizing Production, Decisions, and Consumer Value

Bvuma S. [1], **Sathekge M. S.** [1]

[1] *University of Johannesburg, South Africa*

Received: *26.10.2025;* ***Accepted:*** *10.02.2026;* ***Published:*** *10.03.2026*

Abstract

The chapter discusses three overlapping areas of transformative effects of artificial intelligence (AI) smart manufacturing, augmented decision-making, and personalized consumer experiences. In the manufacturing industry, AI can be used to facilitate predictive maintenance, intelligent supply chains, and human-robot cooperation and improve efficiency and resilience. Cognitive automation, predictive analytics, and scenario planning AI are used to supplement human judgment in the decision-making process, enhance accuracy without losing human control. To consumers, AI is giving them hyper-personalized experiences through recommendation systems, behavioral analytics as well as conversational interfaces, and raises ethical issues of privacy and filter bubbles. To be implemented effectively, it must include data quality, workforce up-skilling, governance and responsible innovation.

Keywords: artificial intelligence, smart manufacturing, augmented decision-making, personalized experiences, ethics.

Cite this chapter as:
Bvuma, S., & Sathekge, M. S. (2026). Artificial intelligence as the engine of invention: Revolutionizing production, decisions, and consumer value. In Y. B. Melnyk & M. A. Segooa (Eds.), *Artificial Intelligence in Digital Society, Vol. 1.* (pp. 103–117). KRPOCH. https://doi.org/10.26697/aids.2026.7
The electronic version of this chapter is complete. It can be found online in the AIDS Archive https://doi.org/10.26697/aids

Introduction

The modern artificial intelligence environment is a paradigm shift in the manner in which organizations generate value, streamline business processes and interact with consumers. In contrast to the previous technological revolutions that made existing processes more automated, artificial intelligence (AI) is making possible whole new operational paradigms and even business models. This change is best reflected in three areas that are interrelated: smart manufacturing, where AI leads to the Industry 4.0 effort; augmented decision-making, where intelligent systems improve the human cognitive experience; and personalized consumer experiences, where AI develops individualized interactions on a scale never before seen (Jin et al., 2021). The combination of machine learning, advanced analytics, Internet of Things technologies and cognitive computing has formed an ecosystem where intelligent systems can not only perform tasks, but also learn, adapt and make previously unavailable insights. This chapter looks at how these AI applications are altering competitive landscapes, transforming the meaning of excellence and setting new rules of human-machine cooperation in the digital economy (Ficili et al., 2025).

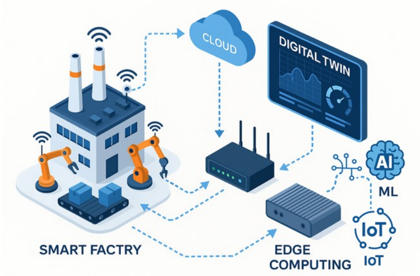
Smart Manufacturing and Industry 4.0

The AI-Enabled Production Paradigm

The final step of Industry 4.0 is the adoption of artificial intelligence in the manufacturing industry, as cyber-physical systems, IoT sensors and smart algorithms coming together resulting in self-optimizing and autonomous production environments. This is not just the automation but also the predictive potential, adaptability, and real-time decision-making that radically changes the production of goods (Radanliev et al., 2021). The core of smart manufacturing is the digital twin, a virtual system simulation fed real-time data by IoT sensors. AI algorithms use this data to forecast outcomes and optimize operations without disrupting production. Aerospace manufacturers, for instance, use digital twins to predict component failures weeks in advance, making maintenance predictive, which decreases unexpected downtime (Rechkemmer et al., 2025).

Figure 7.1

Digital Twin



Predictive Maintenance and Quality Optimisation

Predictive maintenance is one of the most significant uses of AI in industry. Conventional maintenance plans flip-flop between two extremes reactive maintenance that responds to failures once they happen which leads to expensive downtimes and preventive maintenance that is fixed on schedules which tend to replace components at an early stage. Artificial intelligence-based predictive maintenance is not confined to these constraints and utilizes sensor data patterns in terms of vibration, temperature changes, acoustic emissions, and operational parameters to identify insidious irregularities that lead to equipment breakdown (Li & Li, 2025). Machine learning and deep learning generate probabilistic failure predictions using historical and real-time sensor data. These systems identify specific equipment degradation trends, allowing maintenance teams to intervene optimally. Carmakers using this approach have reduced unplanned downtime by up to forty percent and cut maintenance expenses by twenty to thirty percent. (Karkaria et al., 2024). There has also been computer vision and machine learning revolutionizing quality control. Historical inspection systems are based on human operators or rule-based systems, which are not very adaptable. AI-based visual inspection systems use convolution neural networks that have been trained on thousands of defect images to detect defects with higher accuracy and consistency than humans (Jankauski et al., 2022). AI-based visual inspection detects defects in semiconductor wafers and automotive parts. Sophisticated systems also enable root cause analysis by correlating defects with upstream process parameters for continuous improvement.

Intelligent Supply Chain Management

Optimization of supply chains with AI will help manage the complexity of the contemporary global networks in which thousands of suppliers, logistics providers, and distribution points can interact dynamically. Machine learning algorithms are used to analyse the trends of demand in the past, market trends, weather conditions, economic factors, and other social media indicators to come up with very precise demand forecasting (Pypenko & Melnyk, 2021; Raj (2025). Precise forecasts enable optimized inventory, production scheduling, and purchasing, eliminating stock-outs and oversupply. Reinforcement learning (RL) is highly effective in supply chain optimization, as AI agents learn best strategies through simulation and trial and error. Unlike traditional methods, RL agents adapt to uncertainty and dynamical situations, maximizing long-term goals while adhering to constraints (Li et al., 2022). AI-driven supply chains offer impressive resilience, quickly proposing countermeasures during disruptions. Manufacturers with these systems recovered faster and maintained higher service levels than competitors.

Collaborative Robotics and Human-Machine Integration

The new technology of AI-controlled collaborative robots, or cobots, is a paradigm shift of the previous industrial automation. Cobots do not need to be isolated behind safety supplies as the conventional industrial robots do; instead, they collaborate with human workers and can be used in conjunction with their dexterity and judgment. AI allows these systems to sense the surrounding with computer vision, adjust to changes in components and positioning and act safely in the presence of human beings (Cohen et al., 2025). Sophisticated collaborative robots (cobots) use reinforcement or imitation learning to master complex manipulation skills. They evolve through practice rather than extensive programming, enabling them to execute intricate, variable tasks, like electronics assembly, which are typically challenging to automate (Shrivastava et al., 2025). Human-cobot collaboration has more than just task allocation as a way of increasing productivity. AI systems examine workflows to determine the most appropriate task separation that implies that specific, repetitive operations are assigned to robots, and judgment-based and variable work is left to humans. This synergy will not only increase the throughput, but also make workers happier as they will no longer have to perform monotonous tasks and will experience less physical burden.

Figure 7.2
The Collaborative Factory

Note. From “The collaborative factory: How “Cobots” and AI are redefining the 2025 assembly line”, by D. Kim, 2025, *Made-in-China* (https://insights.made-in-china.com/The-Collaborative-Factory-How-Cobots-and-AI-are-Redefining-the-2025-Assembly-Line_AaDThbrcVEHo.html). Copyright 2025 by Focus Technology Co., Ltd.

Implementation Challenges and Integration Considerations

There are strong reasons to believe that AI implementation in manufacturing settings is associated with significant challenges despite the strong advantages. The old systems pose a serious challenge to integration because most production plants are using many decades old equipment and software. Such systems are not always connected and their data is not easily accessible, likely to be used in AI.

Retrofitting IoT sensors and creating data pipelines are costly in terms of investment and operational risk (Vössing et al., 2022). Another important challenge is the data standardization. The manufacturing organizations usually store data in a heterogeneous format in different systems, which include enterprise resource planning systems, manufacturing execution systems, programmable logic controllers and quality databases. The cleaning, normalization, and integration of this data to be used by AI applications is a very laborious task. Also, the relevance and quality of historical data is frequently not sufficient to build resilient machine learning models (Aldoseri et al., 2023). The most challenging aspect is workforce transformation. The introduction of AI in the manufacturing industry demands the inclusion of staff members who are knowledgeable about the manufacturing processes and the field of data science, and such skill set is rather hard to find. Companies should invest in an all-encompassing up-skilling program, where engineers and operators are trained on data literacy, data thinking and the basics of AI. At the same time, they should deal with labor concerns regarding automation's use in the workplace, reassuring workers that AI will complement and not substitute their skills (Engström et al., 2024).

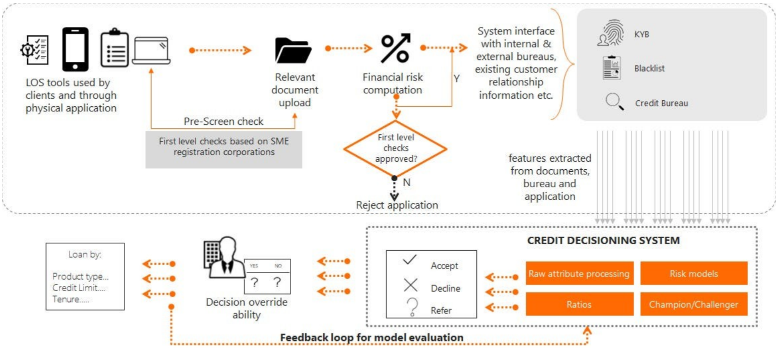
Augmented Decision-Making and Intelligent Automation

The Paradigm of Human-AI Collaboration

The most advanced uses of AI in decision-making acknowledge that the ideal results are not achieved when human judgment is substituted but rather enhanced (Pypenko, 2023). This paradigm recognizes the fact that humans and AI systems have complementary capabilities: humans are good at contextual (contextual) understanding and moral judgment and creative problem-solving, whereas AI systems can process large volumes of information, detect subtle patterns, and be consistent across a wide range of decisions (Tasente, 2025). Such a philosophy is reflected in the form of decision support systems that combine various AI technologies, such as the use of machine learning to identify patterns and extract information, natural language processing to extract information, and predictive analytics to model scenarios, and present the insights to human decision-makers. These systems do not produce independent decisions but suggest evidence-based ones, underline the factors that matter, and also quantify uncertainties, which allow the human to make more informed decisions (Singh et al., 2025). This is demonstrated in AI-enhanced credit decisioning in the financial services. Hundreds of variables are analyzed by machine learning models: transaction history, behavioral patterns, alternative information to determine creditworthiness more precisely than traditional scoring models. Nevertheless, in many cases, human officers make final lending decisions using contextual and qualitative data. This blend of AI analysis and human judgment decreases default rates while ensuring fairness (Abi, 2025).

Figure 7.3

SMB Lending Banks System



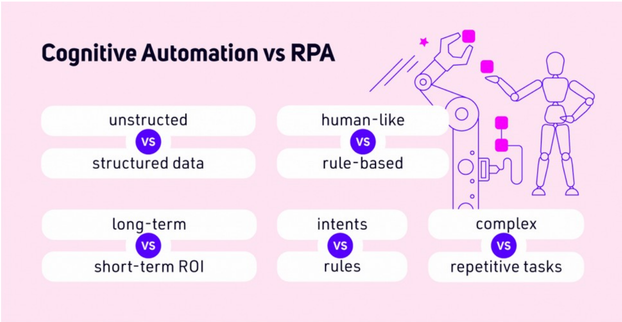
Note. From “Deep learning for recommender systems: A Netflix case study”, by H. Steck et al., 2021, *AI Magazine, 42*(3) (https://doi.org/10.1609/aimag.v42i3.18140). Copyright 2021 by John Wiley and Sons.

Cognitive Automation and Robotic Process Automation

Robotic process automation has developed to be simple rule-based execution of tasks to be cognitive in the execution of unstructured information and variation adaptation. Classical RPA is an efficient technique in highly structured and repetitive jobs such as data entry and report creation. The majority of business processes are however unstructured, meaning they are emails, documents, images that come in various format and content. Cognitive automation builds on the RPA and adds AI to it: natural language processing to comprehend text, computer vision to read between the lines and machine learning to address exceptions (Chennupati, 2025). Cognitive automation excels in intelligent document processing, handling millions of varying documents1. These self-taught systems use machine learning, NLP, and computer vision to classify and extract pertinent data2. Banks using this technology have cut document processing by seventy percent, improved accuracy, and freed employees for exception handling and customer service (Pingili, 2025). A combination of cognitive automation and decision support form end-to-end intelligent process automation. The AI systems in insurance claims processing search and extract data on the claim forms and additional documentation, cross-reference the policy particulars and medical record, evaluate the fraud indicators, approximate the values of claims based on past trends, and direct complex cases to the relevant specialists. Human adjusters deal with subtle cases and final decisions, but AI significantly speeds up process routine and identifies risk areas that could go undetected (Windmann et al., 2024).

Figure 7.4

Cognitive Automation



Note. From "What is cognitive automation and how does it differ from robotic process automation?" by A. Rzeźniczak, 2022, *TUATARA* (https://tuatara.pl/blog/cognitive-automation-rpa/).

Predictive Analytics and Strategic Planning

Statistical forecasting models based on AI have reinvented the concept of strategic planning as organizations are now able to predict the market changes, competitor actions, and operational risks with a level of precision that has never been witnessed before. These models combine various data streams, economic indicators, social media mood, competitor activity, weather conditions, geopolitical events to come up with probabilistic forecasts at different time horizons (Csaszar et al., 2024). With architectures built around deep learning, specifically recurrent neural networks and transformer models have proven to be extremely effective in time series forecasting. These neural networks are useful in contrast to the traditional statistical techniques which tend to assume linear relations and data stationary, these neural networks are able to learn complicated, nonlinear dynamics and adjust to structural shifts in data trends. Those retailers that use these models to predict demand claim to achieve a fifteen to twenty-five percent higher precision than their traditional methods, directly as a result of decreased stock outs and decreased inventory. The AI has also been used in scenario planning. Strategic choices are made by organizations in conditions of deep uncertainty: disruption of technologies, changes in regulation, market development. Scenario generators on AI rely on past trends and simulation methods, exploring large spaces of possibilities and finding plausible futures and their consequences (Ranjan & Kettani, 2025). The systems assist in the stress-test strategies of the executive, early warning signatures, and contingency plans. In the recent economic turmoil, AI-enabled scenario planning proved to be more strategic and did not lead to performance deterioration in organizations.

Balancing Automation and Human Judgment

Although AI is truly impressive, automated decision-making can be extremely dangerous. Machine training systems have the potential to reproduce any kind of bias contained in their training data, cannot make predictions reliably when going out of their experience, and may maximize a parsimonious set of goals without looking at the bigger picture. The poor decision-making is possible because of the automation bias that is a human predisposition to prefer algorithmic suggestions despite conflicting information (Horowitz & Kahn, 2024). Effective Human-AI interaction requires careful design. Transparency systems should explain AI suggestions in plain language for critical assessment. Confidence pointers indicate AI certainty, guiding appropriate skepticism. Finally, override capabilities preserve human agency, enabling dismissal of AI suggestions when contextual factors warrant it (Vössing et al., 2022). Organizations should also need to have governance systems that stipulate how AI should be used in making decisions. Decisions that involve the ethical aspect or have serious consequences and require human final authority are often those that need analytical assistance by AI. Regular, predetermined choices that have specific goals and can be measured can be assigned to AI that is controlled by human decision-making (Kandikatla & Radeljic, 2025). This model of governance will make AI accountable and give it the advantages of efficiency.

AI-Driven Personalized Consumer Experiences

The Architecture of Personalization

Personalization is no longer a primitive form of segmentation and basic rules of recommendation that are soon turning into advanced AI that can generate personal experiences to millions of consumers at a time. The baseline is the capacity to handle a large amount of behavioral information through browsing history, purchase history, what they read, search history, social interactions and derive actionable conclusions regarding personal preferences, needs and contexts (Patil, 2025). Recommendation engines are the most visible personalization application1. They use collaborative and content-based filtering, integrated with deep learning, to learn rich user and item representations2. At Netflix, neural networks predict subscriber likes, considering hundreds of variables, with over eighty percent of viewing activity driven by these systems (Steck et al., 2021). In addition to recommendations, AI can be used to dynamically personalize whole user experiences. E-commerce sites can change the representation of the product, price, offers, and communication according to specific traits and activities. Financial services applications tailor interfaces, accentuate features of interest, and provide recommendations that are proactive and in line with the financial goals and situations of the users (Kanaparthi, 2024). These adaptive experiences maximize the engagement, satisfaction, and business results at the same time.

Natural Language Interfaces and Conversational AI

Natural language processing has made conversational interfaces that communicate with consumers via text or voice in more human like manner. Neural language models that have been trained using large text corpora are used to understand queries, produce contextually relevant responses, and support coherent conversations by virtual assistants, chatbots, and voice interfaces (Shrivastava et al., 2025). These systems have progressed significantly in sophistication by transformer architecture and large language models. These models are sensitive to small linguistic cues, are able to deal with ambiguous queries, remember context across conversations turns and can produce fluent natural responses. Contemporary conversational AI applications address customer support queries, offer tailored suggestions, transacting, and giving expertise advice on various fields (Maxiom, 2024). In medical care, conversational agents can be used to identify initial symptoms, through AI, give medication reminders, offer psychological assistance, and answer general medical inquiries. These systems use medical knowledge graph and capability to give precise information and identify cases that need human clinical judgment. Research has shown that patients can be greatly satisfied with AI-based health assistants in their everyday interactions and leave human clinicians to work with complicated cases (Chaudhry & Debi, 2023). Financial institutions use conversational AI for customer service and fraud detection. These systems utilize voice biometrics, answer account questions, simplify complex products, and offer spending insights. Combining chatbots with AI servers creates a seamless experience, fully serving customers through natural conversation.

Behavioral Analytics and Predictive Personalization

Personalization now extends beyond explicit user requests to anticipating underlying needs and wants. Behavioral analytics uses machine learning to detect subtle patterns in user interactions, such as hesitations and browsing sequences, to infer intentions and preferences. This allows for predictive personalization. For instance, content streaming services use AI to analyze viewing patterns, pause behavior, and time to recommend specific content and even inform investment in new original content tailored to predicted audience segments. In the retail sector, behavioral analytics dynamically personalizes the entire shopping experience based on customer intelligence. AI systems process click streams and determine whether someone is going to buy, how sensitive the price will be, whether a person is in danger of leaving, and who to cross-sell with (Kanaparthi, 2024). This intelligence fuels tailored email promotions, customized web experiences, promotions and tailored product selections. According to retailers, enhancement of conversion rates by twenty to forty percent occurs due to extensive behavioral personalization.

Privacy, Filter Bubbles, and the Ethics of Personalization

Hyper-personalization causes serious moral dilemmas that companies need to manage in order to keep consumers loyal and benefiting society. The most important issue is privacy. The personalization systems demand a large amount of data collection and analysis, which poses a conflict with privacy expectations of users. Whereas consumers value the value of personalized experiences, several complain of the uneasiness of data collection and algorithmic profiling when it comes to personalization (Cai & Mardani, 2023). Rules such as the General Data Protection Regulation and the California Consumer Privacy Act impose visibility, consent, and control of the personal data of users. Organizations need to apply the privacy-by-design principles, including limiting data gathering to that amount of information that is absolutely required, anonymizing the data when possible, and ensuring meaningful transparency of the way the data is used. In a similar fashion, the different techniques of differential privacy offering mathematical noise to secure individual privacy and retain collective insights can provide good solutions to responsible personalization (Ježová, 2020). Another issue is the filter bubble effect in which personalization systems form copy chambers by showing mainly content that conforms to already existing preferences. Too much personalization can restrict the experience of different views, new discoveries, and random experiences that can enhance human experience. Any recommendation system that is optimized on the basis of engagement measures only will provide a thin content diet that will reinforce existing opinions and preferences (Tasente, 2025). To deal with this challenge, the trade off between personalization, diversity and exploration is necessary. Explicit exploration objectives can be included in recommendation systems which sometimes propose content not recommended based on predicted preferences to expand exposure. The interfaces can emphasize the algorithmic personalization where the user can set the level of personalization, or can navigate outside suggestions. Organizations should not only think about engagement optimization but also about more far-reaching effects on user welfare and discourse on the societal level.

Methodology

This research chapter uses a rigorous methodology to investigate modern AI applications in manufacturing, decision support, and consumer experience. The analysis is based on a systematic literature search, including peer-reviewed publications, industry reports, and technical documentation from 2020-2025, prioritizing quantifiable findings and implementation obstacles.

Case studies examined mature AI implementations across the automotive, aerospace, financial services, retail, healthcare, and technology sectors. Organization selection was based on implementation maturity and recorded performance effects to identify successful factors and consistent issues. Empirical evidence, such as downtime reduction, accuracy improvement, and cost savings percentages, was synthesized from publicly available performance measures and industry benchmarking to form a complete picture of AI trends and organizational effects.

Recommendation

- The focus of the organizations must be on augmentation instead of replacing the human capabilities by designing AI systems that complement human abilities instead of automating the tasks. This entails investing in up-skilling human resources initiatives, developing clear AI regulations, and defining workflows that are collaborative with both the human judgment and machine intelligence working together in a manner that is productive.
- The quality of the data must be high and standardized to implement AI successfully. To prevent this, organizations need to invest in data integration hubs, have stringent data governance policies and retrofit old systems with IoT sensors and connectivity. Artificial intelligence implementation should be preceded by data quality initiatives that will guarantee the reliability of training models and predictions.
- Companies should take the lead in terms of privacy, algorithm discrimination and transparency. This involves the use of privacy-by-design practices, the routine auditing of algorithms, the meaningful user control of personalization, and balancing optimization measures with greater societal concerns to
- Instead of trying to achieve total transformation at the same time, organizations are advised to extract high-impact use cases, conduct pilot projects, measure outcomes rigorously and scale successful projects. The strategy minimizes the risk of implementation, shows value early, develops organizational capability gradually, and allows learning based on initial experience.
- AI implementation would also involve cooperation of technical teams, domain experts, and business leaders. The companies are advised to develop cross-functional groups, adopt common terms and goals, apply the agile approach, and establish active communication among data scientists, engineers, operational members, and the executive management team during the implementation cycle.

Conclusion

The artificial intelligence revolution, spanning intelligent production, augmented decision-making, and customized consumer experiences, marks a fundamental paradigm shift in value creation within the digital economy. AI not only automates tasks but also fosters new operational models and human-machine collaboration.

In manufacturing, AI realizes the Industry 4.0 vision by enabling intelligent production – forecasting maintenance, optimizing quality, adapting supply chains, and facilitating human-robot interaction – leading to competitive advantages like lower costs and accelerated innovation. However, successful implementation requires managing challenges in system integration, data quality, and workforce transformation.

Augmented decision-making highlights AI's power to boost human capacity, not substitute it, by combining machine analysis with human judgment for superior outcomes. This demands governance frameworks to prevent excessive automation dependence. Similarly, large-scale individualized consumer experiences change expectations but introduce ethical concerns, particularly regarding privacy and autonomy.

Responsible AI deployment must balance business goals with consumer welfare and societal values. The future of Human-AI integration hinges on enhancing human abilities, ensuring reasonable correct ability, and prioritizing responsible innovation to realize AI's common benefits.

References

Abi, N. R. (2025). Machine learning for credit scoring and loan default prediction using behavioral and transactional financial data. *World Journal of Advanced Research and Reviews*, *26*(3), 884–904. https://doi.org/10.30574/wjarr.2025.26.3.2266

Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-thinking data strategy and integration for artificial intelligence: Concepts, opportunities, and challenges. *Applied Sciences*, *13*(12), Article 7082. https://doi.org/10.3390/app13127082

Cai, H., & Mardani, A. (2023). Research on the impact of consumer privacy and intelligent personalization technology on purchase resistance. *Journal of Business Research*, *161*, Article 113811. https://doi.org/10.1016/j.jbusres.2023.113811

Chaudhry, B. M., & Debi, H. R. (2023). User perceptions and experiences of an AI-driven conversational agent for mental health support. *mHealth*, *10*, Article 22. https://doi.org/10.21037/mhealth-23-55

Chennupati, N. (2025). Cognitive RPA: A framework for hybridizing artificial intelligence with robotic process automation in enterprise systems. *European Journal of Computer Science and Information Technology*, *13*(19), 64–78. https://doi.org/10.37745/ejcsit.2013/vol13n196478

Cohen, Y., Biton, A., & Shoval, S. (2025). Fusion of Computer Vision and AI in Collaborative Robotics: A review and future prospects. *Applied Sciences*, *15*(14), Article 7905. https://doi.org/10.3390/app15147905

Csaszar, F. A., Ketkar, H., & Kim, H. (2024). Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors. *Strategy Science*. https://doi.org/10.1287/stsc.2024.0190

Engström, A., Pittino, D., Mohlin, A., Johansson, A., & Mirzaei, N. E. (2024). Artificial intelligence and work transformations: integrating sensemaking and workplace learning perspectives. *Information Technology and People*, *37*(7), 2441–2461. https://doi.org/10.1108/itp-01-2023-0048

Ficili, I., Giacobbe, M., Tricomi, G., & Puliafito, A. (2025). From sensors to data intelligence: Leveraging IoT, cloud, and Edge computing with AI. *Sensors*, *25*(6), Article 1763. https://doi.org/10.3390/s25061763

Horowitz, M. C., & Kahn, L. (2024). Bending the automation bias curve: A study of human and AI-based decision making in national security contexts. *International Studies Quarterly*, *68*(2). https://doi.org/10.1093/isq/sqae020

Jankauski, M., Schwab, R., Casey, C., & Mountcastle, A. (2022). Insect wing buckling influences stress and stability during collisions. *Journal of Computational and Nonlinear Dynamics*, *17*(11). https://doi.org/10.1115/1.4055309

Ježová, D. (2020). Principle of privacy by design and privacy by default. In *Institute of Comparative Law ; University of Pécs Faculty of Law ; Josip Juraj Strossmayer University of Osijek, Faculty of Law eBooks* (pp. 127–139). https://doi.org/10.18485/iup_rlr.2020.ch10

Jin, X., Ke, Y., & Chen, X. (2021). Credit pricing for financing of small and micro enterprises under government credit enhancement: Leverage effect or credit constraint effect. *Journal of Business Research*, *138*, 185–192. https://doi.org/10.1016/j.jbusres.2021.09.019

Kanaparthi, V. (2024). AI-based personalization and trust in digital finance. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2401.15700

Kandikatla, L., & Radeljic, B. (2025, October 10). *AI and Human Oversight: A Risk-Based Framework for Alignment*. arXiv.org. https://arxiv.org/abs/2510.09090

Karkaria, V., Chen, J., Luey, C., Siuta, C., Lim, D., Radulescu, R., & Chen, W. (2024). A digital twin framework utilizing machine learning for robust predictive maintenance: enhancing tire health monitoring. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2408.06220

Kim, D. (2025, July 18). The collaborative factory: How "Cobots" and AI are redefining the 2025 assembly line. *Made-in-China.com*. https://insights.made-in-china.com/The-Collaborative-Factory-How-Cobots-and-AI-are-Redefining-the-2025-Assembly-Line_AaDThbrcVEHo.html

Li, W., & Li, T. (2025). Comparison of deep learning models for predictive maintenance in industrial manufacturing systems using sensor data. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-08515-z

Li, Y., Pan, Q., He, X., Sang, H., Gao, K., & Jing, X. (2022). The distributed flowshop scheduling problem with delivery dates and cumulative payoffs. *Computers & Industrial Engineering*, *165*, Article 107961. https://doi.org/10.1016/j.cie.2022.107961

Maxiom. (2024, May 2). Discover the pioneering era of conversational AI in LLMs. *Maxiom Technology*. https://www.maxiomtech.com/conversational-ai-in-llms/?utm

Patil, D. (2025). Artificial intelligence for personalized marketing and consumer behaviour analysis: Enhancing engagement and conversion rates. https://doi.org/10.2139/ssrn.5057436

Pingili, N. R. (2025). AI-driven intelligent document processing for banking and finance. *International Journal of Management & Entrepreneurship Research*, *7*(2), 98–109. https://doi.org/10.51594/ijmer.v7i2.1802

Pypenko, I. S. (2023). Human and artificial intelligence interaction. *International Journal of Science Annals,* *6*(2), 54–56. https://doi.org/10.26697/ijsa.2023.2.7

Pypenko, I. S., & Melnyk, Yu. B. (2021). Principles of digitalisation of the state economy. *International Journal of Education and Science, 4*(1), 42–50. https://doi.org/10.26697/ijes.2021.1.5

Radanliev, P., De Roure, D., Nicolescu, R., Huth, M., & Santos, O. (2021). Artificial intelligence and the internet of things in industry 4.0. *CCF Transactions on Pervasive Computing and Interaction*, *3*(3), 329–338. https://doi.org/10.1007/s42486-021-00057-3

Raj, A. (2025, September 27). *Supply chain intelligence and analytics software through put AI*. ThroughPut Inc. https://throughput.world/

Ranjan, R. P., & Kettani, Z. (2025, October 10). *Scenario planning for managing AI disruption risk: A 3C-AI framework*. California Management Review. https://cmr.berkeley.edu/2025/10/scenario-planning-for-managing-ai-disruption-risk-a-3c-ai-framework/

Rechkemmer, D., Korth, M., May, M. C., & Lanza, G. (2025). Development of a concept for the design of user-friendly simulation models. *Procedia CIRP*, *132*, 110–115. https://doi.org/10.1016/j.procir.2025.01.019

Rzeźniczak, A. (2022, August 11). *What is cognitive automation and how does it differ from robotic process automation?* TUATARA. https://tuatara.pl/blog/cognitive-automation-rpa/

Shrivastava, N., Tewari, P., Sujatha, S., Bogireddy, S. R., Varshney, N., & Sharma, V. (2025). Natural language processing for conversational AI: Chatbots and virtual assistants. In *2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation* (pp. 1–6). https://doi.org/10.1109/iatmsi64286.2025.10984818

Singh, S., Chang, Q., & Yu, T. (2025). Hierarchical learning for robotic assembly tasks leveraging learning from demonstration. *Advanced Robotics Research, 1*(1), Article 2400024. https://doi.org/10.1002/adrr.202400024

Steck, H., Baltrunas, L., Elahi, E., Liang, D., Raimond, Y., & Basilico, J. (2021). Deep learning for recommender systems: A Netflix case study. *AI Magazine*, *42*(3), 7–18. https://doi.org/10.1609/aimag.v42i3.18140

Tasente, T. (2025). Understanding the dynamics of filter bubbles in social media communication: A literature review. *Vivat Academia*, 1–21. https://doi.org/10.15178/va.2025.158.e1591

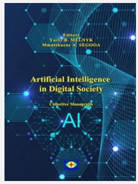
Vössing, M., Kühl, N., Lind, M., & Satzger, G. (2022). Designing transparency for effective Human-AI collaboration. *Information Systems Frontiers*, *24*(3), 877–895. https://doi.org/10.1007/s10796-022-10284-3

Windmann, A., Wittenberg, P., Schieseck, M., & Niggemann, O. (2024). Artificial intelligence in industry 4.0: A review of integration challenges for industrial systems. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2405.18580

Information about the authors:
Bvuma Stella – https://orcid.org/0000-0001-8351-5269; PhD in Information Technology Management; Professor, Director, University of Johannesburg, Johannesburg, South Africa.
Sathekge Machiniba Sylvia – https://orcid.org/0009-0001-9410-3267; Doctor of Business Administration, Doctor, Professor of Practice, University of Johannesburg, Johannesburg, South Africa.

Artificial Intelligence in Digital Society,
Volume 1, 2026

DOI: 10.26697/9786177089192.2026
ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)



Chapter 8. Navigating Governance, Ethics, and Data Security Risks in Artificial Intelligence Adoption

Sathekge M. S. [1], **Bvuma S.** [1]

[1] *University of Johannesburg, South Africa*

Received: *25.10.2025;* ***Accepted:*** *10.02.2026;* ***Published:*** *10.03.2026*

Abstract

The fast and viral uptake of Generative AI (GenAI) and large foundation models (LFMs) in the corporate worlds is a major, but ill-managed, change in organizational security, risk, and governance. Although GenAI involves an overwhelming number of advantages, its implementation creates a new layer of data security threats that traditional ICT security models were not created to cover. The chapter gives a critical review of the situation in governance today and the particular ethical and technical issues that come along with the integration of GenAI. It is analyzed to explain GenAI Security Threats, such as model poisoning and prompt injection, and the highly significant problem of data leakage and exposure of intellectual property (IP). It also explores the Ethical Gaps, which say that inexplicable bias may yield the results of discrimination or generate shadow vulnerability, which could not be audited. The main contribution is the suggestion of a Socio-Technical Governance Framework that incorporates human control, Explainable AI (XAI), and constant security surveillance into the GenAI deployment pipeline. Actionable Best Practices of data sanitization, model validation and defining clear lines of accountability in AI-driven decisions support this framework. This chapter is meant to inform technology leaders and policymakers by expressing the need to have a proactive risk-based approach to ensure GenAI is exploited safely and in a manner that is responsible in the digital society.

Keywords: generative AI, data security, corporate governance, ethical AI, risk management, model poisoning, data leakage explainable AI.

Cite this chapter as:

Sathekge, M. S., & Bvuma, S. (2026). Navigating governance, ethics, and data security risks in artificial intelligence adoption. In Y. B. Melnyk & M. A. Segooa (Eds.), *Artificial Intelligence in Digital Society, Vol. 1*. (pp. 118–131). KRPOCH. https://doi.org/10.26697/aids.2026.8

The electronic version of this chapter is complete. It can be found online in the AIDS Archive https://doi.org/10.26697/aids

Introduction

Context: The Generative AI Discontinuity and Corporate Risk

The adoption of Generative AI (GenAI) and Large Foundation Models (LFMs) into businesses at an accelerated pace and with viral pace can be deemed as one of the largest technological changes of the decade, quickly transforming into an experimental novelty and becoming an essential utility in a business enterprise.

The GenAI is full of potential, and it leads to tangible productivity increases and will change the fundamental operations of code generation, research, and content creation. But because of this pace of adoption driven by business demand of a competitive advantage, it has brought about a major disjuncture in the structure of organizational security, risk, and governance.

The problem is in its structure: GenAI brings a new dimension of data security threats that traditional Information and Communication Technology (ICT) security frameworks were not supposed to deal with (Kendzierskyj et al., 2024). The fact that AI systems are not fully similar to conventional software is based on the fact that threats are no longer targeted at the code but at the very learning and adaptive nature of the model. This requires an interdisciplinary approach that is holistic and in tandem with the future of AI-driven transformation (Radanliev et al., 2025).

Problem Focus and Chapter Goal

A major gap between institutional regulation and the safeguarding of technological innovation and risk has been caused by the rate of GenAI adoption exceeding the maturity of institutional protections. Traditional Information and Communication Technology (ICT) security models were not made to be responsive to the underlying vulnerabilities of probabilistic, content-generating systems (Radanliev et al., 2025).

The chapter of this book argues that this governance gap results in unacceptable corporate exposure in three interrelated areas:

1. Data Security Threats: New attack vectors such as model poisoning and prompt injection compromise model integrity and enable sensitive information to be exfiltrated. This is worsened by the fact that there has been a growing threat of information leakage and intellectual property (IP) exposure as proprietary corporate data are more and more passing to and processing by GenAI tools (Sidorkin, 2025).

2. Ethical and Fairness Gaps: With the complexity and the obscurity of the inner mechanics of the LLMs, it is challenging to analyze the mechanisms to unintentionally reproduce training data-driven biases (Bano et al., 2023; Melnyk, 2025). When applied in the high-stakes corporate security or access control, the outcomes can be discriminatory with shadow vulnerability that cannot be audited in traditional ways.

3. Accountability Deficit: The lack of clear, strictly implemented policies on governance issues surrounding the use of data, along with the self-directed nature

of the work of Generative AI, grossly contributes to the loss of accountability of the locus in cases of error or harm. Such ambiguity significantly increases the chances of the violation of the regulatory requirements (Mandava, 2025).

The Chapter Goal will be the critical synthesis of these challenges and, consequently, to formulate the need to adopt a proactive, risk-based approach to GenAI security governance.

Chapter Contribution and Structure

The fundamental input of this piece of work is the suggestion of a Socio-Technical Governance Framework. This framework is intended to address the governance gap by combining human oversight and the concepts of XAI, including transparency and interpretability, and the ongoing security monitoring as a part of the GenAI deployment pipeline directly. This is a direct way of addressing the causes of the problem of the black box, where the grounding of governance on human-understandable processes is done. The further parts continue with a critical synthesis: initially, the descriptions of the research method and the conceptual basis (Section 2) and the Ethical Gaps Analysis (Section 3) which require systemic change. The chapter then proceeds to introduce the detailed Socio-Technical Governance Framework (Section 4) and the related Actionable Best Practices (Section 5) to be implemented, and finally, is concluded with strategic suggestions on the part of technology leaders and policy makers.

Conceptual Foundation, Research Approach and Methodology

Research Approach and Methodology

The study involves deductive research in which the researcher attempts to derive the topic and the research questions through inductive reasoning. The research employs a deductive research methodology where the investigator tries to come up with the topic and the research questions by using inductive inferences.

The basis of this chapter is a critical synthesis and analysis of authoritative, publicly published reports and a literature review conducted by peer-reviewed sources. The method was chosen to quickly internalise the most recent discoveries, regulatory outlooks, and risk evaluations in regards to GenAI, which tend to be uncovered initially in authoritative industry white papers and specialised, conferences because of the faster rate at which the technology advances.

The search strategy was based on locating the literature published during the past five years (2021-2025), whereby much attention was paid to the publications that were published after 2023 when the GenAI explosion has taken effect after the introduction of LFMs on a large scale. The major search terms were Generative AI, Data Security, Corporate Governance, Ethical AI, Risk Management, Model Poisoning, Data Leakage, XAI. This methodology was important to make sure that the threats found and mitigation measures suggested are extremely relevant and up-to-date.

Conceptual Foundation

According to Chen and Metcalf (2024), Socio-Technical assumes the existence of a company where society and technologies are integrated into a single system. Significant, neither is it possible to conceive the social without the technical, nor the technical without the social. The Socio-Technical Governance Framework, as suggested is deeply based on the Socio-Technical Systems (STS) theory. STS aims to regard organizations, processes, and technology as not separate entities, but as one system. This is necessary within the framework of the Digital Society since the governance failures of the GenAI are not a matter of technical deficiency (e.g., the deficiency of the code) but the deficiency of the systemic engagement between the algorithmic functionality and the human policy, oversight, and culture.

Using an STS lens offers the conceptual rationale behind the structure of the framework that requires convergence of:

- *Technical Pillar*: The introduction of techniques such as XAI that deal with the technical black box issue by introducing transparency and interpretability.
- *Social Pillar*: Implementing the human control and obligatory forms of accountability including the governance boards, auditor trails that would verify the compliance of AI-made decisions to the ethical and regulatory policy.

The framework therefore maintains that addressing the risks of model poisoning and bias cannot be achieved only by methods of improved filtering algorithms, and it is important to bring clarity in human roles and enforceable policies in terms of access to data and decision review which entrench the principle that technology and human action are mutually constitutive in the determination of a security and ethical outcome. To account for the results of any technology, including AI it is necessary to concentrate on the in-between space between these two pillars, which is also complicated (Chen and Metcalf, 2024).

GenAI Security Threats and Technical Risks

Threats to Model Integrity

GenAI systems are compromised with advanced attacks that attack the data used to train the model and query information.

A. Model Poisoning

Model poisoning refers to malice inputs or maliciously labeled data to the training pipeline to cause the model to adopt tainted behavior. This is a direct contravention of integrity and reliability of the model.

- Targeted Attacks (Backdoors): A backdoor attack is a branch of data poisoning in which the malicious activity is not triggered unless a certain trigger is met or a specific phrase is used (Souly, et al., 2025). These attacks are meant to cause the model to misbehave under a certain trigger that is not explicit in the input, but overall functions well. This is what makes the attack difficult to notice when going through regular validation. According to research, a minimum of 250 documents can be used to reliably backdoor LLMs, despite the overall size of the model itself (Souly, et al., 2025).

- Non-Targeted Attacks (Availability): In such attacks, big amounts of noise data or incorrect data are injected into the overall model and affect its functionality and resilience. It is aimed at leading to massive performance decrease, which causes inaccurate outputs as well as a higher error rate, which jeopardizes the utility and credibility of the model (Hubinger, et al., 2024).

B. Prompt Injection

This is an immediate injection where malicious input is developed by attackers to either disrupt or disorient an AI system. Prompt Injection is the highest-ranked security risk, which is classified as one of the most severe vulnerabilities in the applications of LLM. It is the type of attack in which the developers of the original system provide maliciously designed, harmful inputs to the model, and they override the original system instructions or system prompt. (Redbot Security, 2025).

- Mechanism: Prompt injection takes advantage of the fact that the LLM has no reliable way of differentiating trusted system code and untrusted data, which are sent in by a user. (Sidorkin, 2025).

- Consequences: An effective attack may force the model to do a malicious act e.g. coughing confidential information, cross-site scripting, or unwanted code (Bowen et al., 2025). This is especially hazardous in systems that rely on Retrieval-Augmented Generation (RAG), in which the prompts can trigger the model to retrieve and disclose sensitive information in internal, and also a vector-database knowledge sources (Redbot Security, 2025; Souly, et al., 2025).

Threats to Data and Confidentiality

GenAI as a concept poses extreme risks to corporate information confidentiality and intellectual property (IP), and gives rise to the liability of non-regulatory compliance (Ranjan and Kettani, 2025).

A. Data Leakage and Intellectual Property (IP) Exposure

One of the most pressing concerns is the accidental utilization of the proprietary information to train the models that are public or unsecured (Sidorkin, 2025):

- Insecure Data Ingress/Egress: There has not been a defined, enforced governing policy concerning how proprietary data is inputted into GenAI tools (ingress) or how outputs are managed (egress) that is intolerable.

- Model Memorization: Models can unintentionally memorize certain data in the training corpus during training. In the case of this memorized data, this may be hacked by attackers using certain query methods in case it is proprietary or a source of Personally Identifiable Information (PII).

- Shadow AI Risk: Unmanaged employees use public GenAI services the exposure to unknown and uncharted IP risk on the organization directly through employees who feed the model with confidential data to perform tasks such as summarization or code completion.

B. Regulatory Consequences

These technical failures are converted into direct regulatory risks. The absence of strong security protocols with regard to the management of data is a direct conflict with the new global AI standards and data protection regulations. C leakage of the sensitive information on the grounds of model compromise may result in both expensive regulatory non-conformity and grave reputational losses (Chesterman, 2025).

Ethical Gaps: Bias, Fairness, and Accountability

The technical risks of GenAI are associated with deep ethical loopholes of bias, fairness and accountability. These present regulatory risks, which normally harm reputation and trust more than breaches.

The Challenge of Unexplainable Bias

The major ethical issue of the deployed GenAI is the expression of the bias based on the training data. Such systemic weaknesses can be promoted and exaggerated by models that are trained on biased, historically prejudiced, or unrepresentative data and in their outputs and decisions (Radanliev et al., 2025; Ranjan and Kettani, 2025).

- Bias in Security Decision-Making: When GenAI becomes part of a high-stake system, say, threat detection, access control, or employee monitoring, any bias, even when not explainable, may cause discriminatory results, including the marginalization of certain demographics.
- The Opacity Problem: LFMs are often complex systems whose decision-making processes are so opaque that it makes them black boxes. It is so invisible that the human operators or auditors cannot effectively mitigate it since it is incredibly hard to establish the reason behind a specific decision or how a bias is introduced. Such transparency will compromise the fundamental values of fairness and equity in operations (Mandava, 2025).

The Accountability Deficit and Shadow Vulnerabilities

This does not mean that the responsibility is absent, a problem with GenAI is that its inherent opaqueness always creates a lack of accountability, making it difficult to distinguish who should take responsibility and who should not in a situation where a system delivers a biased or damaging result.

Accountability in a business setting entails being traceable and assigning responsibility to algorithmic activities (Mersah et al., 2025; Ranjan & Kettani, 2025; Sidorkin, 2025).

These can lead to:

- Diffused Responsibility: It is hard to figure out who should take responsibility over autonomous GenAI decisions: the author of the decision is the developer, the curator, the team, or the supervisor? (Janssen, 2025).
- Shadow Vulnerabilities: The lack of accountability give rise to shadow vulnerabilities that are invisible risks that are systemic and cannot be audited. They pose the threat of continuous damages and non-compliance with regulations. The governance should entail accountability at the board level of all GenAI risks (Mandava, 2025; Sidorkin, 2025).

The Imperative for Explainable AI (XAI)

The need to integrate XAI is essential so that responsible governance is achieved, and technical tools are offered to address the issues of connecting the incomprehensible model results to human understanding (Mandava, 2025).

These are:

- Achieving Transparency and Interpretability: Creating Transparency and Interpretability XAI algorithms such as SHAP (quantifies feature contribution) or LIME (local surrogate models) can offer human-understandable components to individual AI decisions (Shankar, 2025). This will enable security analysts to verify alerts as well as ethics officers debugging models before and after deployment.
- Supporting Accountability: XAI contributes to antecedence and auditability of results through clear-cut evidence and supports the Socio-Technical Governance Framework in aspects of integrating technical transparency and human controls (Chen and Metcalf, 2024).

The Socio-Technical Governance Framework

The technical and ethical loopholes require an alternative approach to governance other than conventional policy. To cope with the complexity of GenAI, this chapter suggests a Socio-Technical Governance Framework that would combine human regulation with system transparency to handle systemic failures.

The Rationale

The main objective of this framework is to create a binding connection between corporate policy (the social pillar) and algorithmic functionality (the technical pillar).

The framework is designed in such a way that continuous monitoring, explainability and human judgment are incorporated into the lifecycle of GenAI deployment, not disengaged compliance inspections. The given practice is essential to developing reliable AI in stakes-based settings (Chen and Metcalf, 2024).

Framework Presentation

Figure 8.1 represents the operational scheme of the Socio-Technical Governance Framework.

Figure 8.1

Socio-Technical Governance Framework

Tables 8.1-8.4 illustrate the specific elements of the framework core, linkages and pillars.

Central Core: The GenAI Lifecycle

This is the point where governance must be applied.

Table 8.1

The Gen AI Lifecycle

Component	Description	Relevance
GenAI Deployment Pipeline (Policy & Accountability)	The central process encompassing model training, deployment, inference, and operation	Represents the system being governed; all risks (poisoning, injection) occur here

The Linkages (The Mechanisms)

These components span both pillars and are necessary for communication and transparency.

Table 8.2

The Mechanism

Component	Description	Technical Pillar Function	Social Pillar Function
Explainable AI (XAI) Principles	Implementation of interpretability tools (e.g., LIME, SHAP).	Provides technical transparency and helps debug bias.	Enables human validation and justification of AI-driven decisions.
Transparency and Continuous Security Monitoring	Real-time monitoring of model inputs, outputs, and performance.	Detects and flags adversarial attacks (prompt injection, data leakage).	Triggers human oversight, remediation, and reporting to the Governance Board .

The Technical Pillar (Mitigation and Transparency)

This pillar focuses on the system's defensive and diagnostic capabilities.

Table 8.3

The Mitigation and Transparency

Component	Description	Goal / Mitigation	Source
Data Sanitization and Vetting	Strict policies on data ingress; use of anonymization and source checks.	Mitigates Model Poisoning and Data Leakage.	Mandava (2025)
Layered Input/Output Filtering	Mechanisms to sanitize user prompts (input) and vet model responses (output).	Mitigates Prompt Injection and prevents unauthorized data exfiltration.	Sidorkin (2025)
Adversarial Robustness Testing	Mandatory Red Teaming and stress testing.	Proactive identification of vulnerabilities before deployment.	Bowen, et al. (2025)
Immutable Audit Trails	Technical mechanism to log all AI decisions and associated XAI rationale.	Supports the Accountability principle.	Mersah, et al. (2025)

The Social Pillar (Oversight and Accountability)

This pillar focuses on the institutional and human structures that enforce ethical policy and ensure compliance.

Table 8.4

The Oversight and Accountability

Component	Description	Goal / Mitigation	Source
AI Governance Board (AGB)	Cross-functional executive committee (Legal, Ethics, IT, Senior Management).	Sets high-level policy, conducts risk assessments, and retains ultimate decision authority.	Taeihagh (2025)
Defined Human Oversight Roles	Clear assignment of responsibility (e.g., Human-in-the-Loop, Human-over-the-Loop).	Addresses the Accountability Deficit and prevents discriminatory outcomes.	Kandikatla & Radeljić (2025)
Ethics and Remediation Policy	Formal protocol for pausing, remediating, or retiring a GenAI system.	Ensures systems align with corporate fairness and ethical commitments.	Chesterman (2025)
Regulatory Compliance & Reporting	Formal processes for meeting standards (e.g., POPIA, GDPR, EU AI Act).	Mitigates Regulatory Non-Compliance risk.	Radanliev et al. (2025)

Actionable Best Practices for Implementation

This part delivers the practical guide that was promised in the abstract giving steps on how to implement it. The Socio-Technical Governance Framework requires the shift to the realms of theory and best practices that are warehoused on a real-life agenda, and these practices should encompass the implementation of security and ethics into the Generative AI (GenAI) lifecycle. Such practices are the working units of the framework providing technical strength and institutional responsibility against the mentioned dangers of poisoning, injecting, and losing the data.

Securing the Data Pipeline (Mitigating Poisoning/Leakage)

The initial defence against Model Poisoning and Data Leakage/IP Exposure is protecting the data by which the training process and inference is done.

The practices are aimed at protecting the ingress (input) and egress (output) points of the data:

- Data Sanitization and Minimization: Organizations need to strictly screen and sanitise all training data considering the whole corpus as potentially untrusted. This includes implementing data minimization, processing only the data required to train, and applying such methods as tokenization and data masking to replace or cover Personally Identifiable Information (PII) and sensitive Intellectual Property (IP) (OWASP, 2024; Sidorkin, 2025).

- Source Vetting and Continuous Integrity Checks: To mitigate the possibility of Model Poisoning, third-party data sources as well as internal data sources should be regularly checked by integrity tests and formally vetted. It is a method that protects against the introduction of malicious samples, which can be performed with high precision and hides silently, reliably breaking models of any scale (Souly, et al., 2025; Sidorkin, 2025).

- Enforced Data Ingress/Egress Policies: There should be explicit policies of governance that determine how proprietary data can be ingested into GenAI tools (ingress) and how sensitive summary or code can be egressed (egress). This directly prevents the threat of Shadow AI and unintentional IP leakage. (OWASP, 2024; Mandava, 2025).

Model Validation and Adversarial Testing (Mitigating Prompt Injection)

To make the technical pillar of the framework solid, the persistent validation and testing cannot be limited to the common functional quality assurance but should consider the special weak points of GenAI.

- Layered Input Validation (Prompt Injection Defense): Any input by the user should be considered untrusted. This should be implemented in a layered manner with both rule based filters and AI based classifiers that can both identify and neutralise malicious or obfuscated instructions before they can reach the core LLM. It is the most significant technical defense against timely injection. (OWASP, 2024; Taeihagh, 2025).

- Continuous Adversarial Robustness Testing (Red Teaming): Organizations will have to institutionalize focused Red Teaming actions, such that security specialists simulate adversarial attacks, in particular, the ability of the model to resist sophisticated prompt injection and data exfiltration attacks. This constant validation procedure keeps the defenses up to date with the changing attack vectors (Radanliev et al., 2025).
- Output Filtering and Sanitization: The response generated by the model should be inspected to detect some evidence of illicit information e.g. confidential internal information, PII or malicious instructions. This is a final safeguard against data leakage resulting from a successful prompt injection attack (Sidorkin, 2025).

Operationalizing Accountability and Oversight

These practices codify the social pillar of the framework by creating the clear lines of responsibility required to handle the ethical risk and compliance with regulations.

- Establish Mandatory Audit Trails: One of the conditions that is not negotiable is to create immutable and auditable records of every decision, classification, or action made depending on the GenAI system. This technical necessity is a prerequisite to empower the humanistic value of traceability and accountability, where all the results are justifiable and checked (Radanliev et al., 2025).
- Define and Enforce Lines of Responsibility: The AI Governance Board should establish and implement obvious lines of responsibility, clarifying the ultimate accountable human manager or executive to the consequences of a deployed GenAI system. This minimizes the risk of diffused responsibility in the case where systems are unintentionally biased in any way or that they yield discriminatory results (Kandikatla and Radeljić, 2025; Janssen, 2025).
- Mandate Remediation Mechanisms: The organization should specify transparent, tested standards to shut down, remediate, or discontinue an AI system the moment its ongoing security monitoring or XAI analysis of the result reveals that it is introducing or strengthening bias, modifying data integrity or breaching ethical policy. This offers a required safety valve to ensure the public trust and avoid disastrous breakdowns (Ranjan & Kettani, 2025; Taeihagh, 2025).

Conclusion and Recommendations

As can be seen by the results of this chapter, the risks presented by Generative AI (GenAI) are not just technical issues but governance failures in their entirety. In order to ensure that the innovative power of GenAI will be used safely and responsibly, technology leaders and policymakers should take a risk-based proactive approach. The proposed Socio-Technical Governance Framework contains the following recommendations, which are based on the regulatory governance of Digital Society.

Recommendations for Technology Leaders (Implementers)

1. Mandate Explainable AI XAI Integration for High-Risk Systems: Leadership should mandate XAI integration of high-stakes GenAI security deployments to offer transparency, which can allow human verification and debugging of bias to overcome accountability shortcomings.

2. Institutionalize Adversarial Testing: Red Teaming and adversarial stress testing should be institutionalized and continue continually. This is a necessary practice to detect dynamic vulnerabilities to timely inject and maintain model resilience to model poisoning.

3. Enforce Immutable Audit Trails: To realize accountability, technology departments should deploy technologies that generate immutable auditable records of every decision made or impacted by GenAI. This will make any ethical or security lapses attributable to certain activities in support of the social pillar of accountability.

4. Implement Zero-Trust Data Ingress/Egress Policies: Technology leaders need to have and enforce express policies on data ingress and egress using GenAI tools, such as using data sanitization techniques and Data Loss Prevention (DLP) to reduce the possibility of data leakage and IP exposure.

Recommendations for Policymakers (Regulation and Oversight)

1. Establish Clear Board-Level Accountability: The policies should specify accountability of the results of high-risk AI systems to a particular executive or a board-level committee. This eliminates diffusion of responsibility and also makes sure that risks of not adhering to the regulations are met at the topmost level.

2. Harmonize XAI Requirements: To achieve algorithmic fairness and explainability, policymakers ought to employ minimal requirements on XAI interpretability in the regulated industries. These standards need specifically to deal with the way that companies need to show that their systems of GenAI have been tested and fixed in case of unexplainable bias.

3. Mandate Transparency in Data Provenance Regulatory frameworks ought to demand more disclosure around the provenance and composition of the training data applied by LFMs in the effort to assist organizations to alleviate the threat of model poisoning and shadow vulnerabilities.

Conclusion

This chapter has critically examined the GenAI governance landscape, and it has been shown that the adoption of LFMs has occurred at a rapid, heavily viral pace, which resulted in a considerable discontinuity in the organization security, risk and ethics. The review established that the current ICT security models are poorly prepared to address the particular threats of model poisoning and timely injecting, the lack of transparency in these systems generates an ethical dilemma of bias, fairness and accountability. The main contribution of the given work is the suggested Socio-Technical Governance Framework. This model offers a solid framework of controlling the unavoidable merging of human resourcefulness, business data, and machine generated innovation. The framework guarantees that the ethical standards and security protocols are directly embedded in the deployment life-cycle, through a systematic approach to the combination of the Technical Pillar (XAI and continuous monitoring) and the Social Pillar (human oversight and defined accountability). Such an active and risk-sensitive stance is required to reduce the risk of the high cost of non-compliance with regulatory policies and reputation losses, which means that GenAI will be incorporated safely and responsibly into the new digital society.

References

Bano, M., Zowghi, D., Shea, P., & Ibarra, G. (2023). *Investigating responsible AI for scientific research: An empirical study*. arXiv. https://doi.org/10.48550/arXiv.2312.09561

Bowen, D., Murphy, B., Cai, W., Khachaturov, D., Gleave, A., & Pelrine, K. (2025). Scaling trends for data poisoning in LLMs. *Proceedings of the AAAI Conference on Artificial Intelligence*, *39*(26), 27206–27214. https://doi.org/10.1609/aaai.v39i26.34929

Chen, B. J., & Metcalf, J. (2024, May 28). *Explainer: A sociotechnical approach to AI policy*. Data & Society Research Institute. https://datasociety.net/library/a-sociotechnical-approach-to-ai-policy/

Chesterman, S. (2025). Good models borrow, great models steal: Intellectual property rights and generative AI. *Policy and Society*, *44*(1), 23–37. https://doi.org/10.1093/polsoc/puae040

Hubinger, E., Denison, C., Mu, J., Lambert, M., Tong, M., MacDiarmid, M., ... & Maxwell, T. C. (2024). *Sleeper agents: Training deceptive LLMs that persist through safety training*. arXiv. https://doi.org/10.48550/arXiv.2401.05566

Janssen, M. (2025). Responsible governance of generative AI: Conceptualizing GenAI as complex adaptive systems. *Policy and Society*, *44*(1), 38–51. https://doi.org/10.1093/polsoc/puae039

Kandikatla, L., & Radeljić, B. (2025, October 10). *AI and human oversight: A risk-based framework for alignment*. arXiv. https://doi.org/10.48550/arXiv.2510.09090

Kendzierskyj, S., Jahankhani, H., & Hussien, O. (2024). Space governance frameworks and the role of AI and quantum computing. In H. Jahankhani (Ed.), *Space law and policy* (pp. 1–39). Springer. https://doi.org/10.1007/978-3-031-62228-1_1

Mandava, S. (2025). Explainable data governance using XAI techniques to enhance traceability, transparency, and accountability in AI systems. *Applied Data Science and Analysis*, *2025*(1). https://doi.org/10.58496/ADSA/2025/001

Melnyk, Y. B. (2025). Should we expect ethics from artificial intelligence: The case of ChatGPT text generation. *International Journal of Science Annals, 8*(1), 5–11. https://doi.org/10.26697/ijsa.2025.1.5

Mersah, M. A., Yigezu, M. G., Tonja, A. L., Shakil, H., Iskander, S., Kolesnikovs, O., & Kalita, J. (2025). Explainable AI: XAI-guided context-aware data augmentation. *Expert Systems with Applications*, *289*(128364). https://doi.org/10.1016/j.eswa.2025.128364

OWASP. (2024). *OWASP top 10 for LLM applications 2025*. https://owasp.org/www-project-top-10-for-large-language-model-applications/assets/PDF/OWASP-Top-10-for-LLMs-v2025.pdf

Radanliev, P., Santos, O., & Ani, U. D. (2025). Generative AI cybersecurity and resilience. *Frontiers in Artificial Intelligence*, *8*(1568360), 1–18. https://doi.org/10.3389/frai.2025.1568360

Ranjan, R. P., & Kettani, Z. (2025). Scenario planning for managing AI disruption risk: A 3C-AI framework. *California Management Review*. https://cmr.berkeley.edu/2025/10/scenario-planning-for-managing-ai-disruption-risk-a-3c-ai-framework/

Redbot Security. (2025, October 30). *Prompt-injection-attacks-ai-security-2025*. https://redbotsecurity.com/prompt-injection-attacks-ai-security-2025/

Shankar, V. (2025). Machine learning for Linux kernel optimization: Current trends and future directions. *International Journal of Computer Sciences and Engineering*, *13*(3), 56–64. https://doi.org/10.26438/ijcse/v13i3.5664

Sidorkin, A. M. (2025). AI platforms security. *AI-EDU Arxiv*, *2025*(1). https://journals.calstate.edu/ai-edu/article/view/5444

Souly, A., Rando, J., Chapman, E., Davies, X., Hasircioglu, B., Shereen, E., ... & Kirk, R. (2025, October 8). *Poisoning attacks on LLMs require a near-constant number of poison samples*. arXiv. https://doi.org/10.48550/arXiv.2510.07192

Taeihagh, A. (2025). Governance of generative AI. *Policy and Society*, *44*(1), 1–22. https://doi.org/10.1093/polsoc/puaf001

Tedeneke, A. (2023, June 26). *World Economic Forum launches AI Governance Alliance focused on responsible generative AI*. World Economic Forum. https://www.weforum.org/press/2023/06/world-economic-forum-launches-ai-governance-alliance-focused-on-responsible-generative-ai/

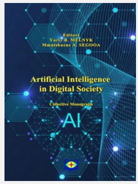
Information about the authors:

Sathekge Machiniba Sylvia – https://orcid.org/0009-0001-9410-3267; Doctor of Business Administration, Doctor, Professor of Practice, University of Johannesburg, Johannesburg, South Africa.

Bvuma Stella – https://orcid.org/0000-0001-8351-5269; PhD in Information Technology Management; Professor, Director, University of Johannesburg, Johannesburg, South Africa.

Artificial Intelligence in Digital Society, Volume 1, 2026
DOI: 10.26697/9786177089192.2026
ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)

Check for updates



Chapter 9. Harnessing Smart Artificial Intelligence for Industrial 4.0: A South African Case Study of Manufacturing Industry

Mogoale P. M. [1], **Pretorius A. B.** [1], **Mogase R. C.** [1], **Segooa M. A.** [1]

[1] *Tshwane University of Technology, South Africa*

Received: *12.12.2025;* ***Accepted:*** *10.02.2026;* ***Published:*** *10.03.2026*

Abstract

The global trend towards Industry 4.0 has raised demand to incorporate technology in the manufacturing industry. This new paradigm requires cyber-physical systems, the Internet of Things (IoT), and Artificial Intelligence (AI) to enhance the efficiency and competitiveness of traditional industrial methods. Industry 4.0 incorporates Smart Artificial Intelligence (SAI) to enhance efficiency, digitalise production, and automate the intelligent processing of commodities. Despite the benefits SAI technology carries, many South African industries struggle to realise its full potential due to resource and financial constraints. This chapter discusses the challenges of SAI adoption and how the manufacturing sector leverages SAI to enhance its productivity and competitiveness. A systematic literature review was conducted. ScienceDirect publications from 2022-2025 period were reviewed. Only the review and research papers, focusing on the SA manufacturing industry, were considered. The findings reveal how the use of SAI in South Africa (SA) is hindered, thereby constraining innovation and productivity. SAI promotes manufacturing in the country; inadequate infrastructure, a lack of funding are the biggest obstacles to implementing SAI in SA. A contribution about how SAI technology is leveraged in the SA manufacturing industry was established, advancing knowledge that may inform industry leaders.

Keywords: smart artificial intelligence, South Africa, smart manufacturing, industry 4.0, systematic literature review.

Cite this chapter as:

Mogoale, P. M., Pretorius, A. B., Mogase, R. C., & Segooa, M. A. (2026). Harnessing smart artificial intelligence for industrial 4.0: A South African case study of manufacturing industry. In Y. B. Melnyk & M. A. Segooa (Eds.), *Artificial Intelligence in Digital Society, Vol.1.* (pp. 132–145). KRPOCH. https://doi.org/10.26697/aids.2026.9

The electronic version of this chapter is complete. It can be found online in the AIDS Archive https://doi.org/10.26697/aids

Introduction

Industry 4.0 (I4.0), a Fourth Industrial Revolution initiative, is transforming the manufacturing sector into a more competitive environment by employing technologies such as smart artificial intelligence (SAI), the Internet of Things (IoT), and cyber-physical systems to enhance productivity and efficiency (Adams, 2023). SAI technologies are crucial in this transformation, enabling enhanced automation, predictive analytics, and optimal resource management in complex industrial processes (Akoh, 2024). SAI is generally acknowledged as a crucial technology for advancing the future development of industrial 4.0 manufacturing (Papadimitriou et al., 2024). Consequently, Industry 4.0 emphasises improved efficiency, digitised manufacturing operations, and the systematic processing of intelligent goods (Philbeck & Davis, 2018; Pypenko & Melnyk, 2021). Despite the benefits, many manufacturing companies struggle to implement SAI technologies due to a lack of necessary resources and knowledge (Espina-Romero et al., 2024). This context highlights the need for targeted interventions to bridge the digital divide and enhance technological capacity in SA, particularly to address shortcomings.

Manufacturing Sector in SA

The manufacturing sector is a vital component of South Africa's economy, significantly contributing to its advancement and wealth (Maphisa et al., 2024). Almost 11,400 VAT-licensed firms contribute to South Africa's manufacturing sector, establishing it as a vital component of the national economy (Ngepah et al., 2024). This sector ranks as the fourth-largest industry in SA (Maisiri & Van Dyk, 2021). This chapter categorises the manufacturing industry based on the Manufacturing, Engineering and Related Services industry Education and Training Authority (MerSETA).

Purpose

The purpose of this chapter is to evaluate the approaches in which different manufacturing industries in SA utilise SAI to improve competitiveness and productivity. The objectives of the chapter are: 1) to analyse current SAI adoption within the South African manufacturing sector, and 2) to recommend a shared understanding of the SAI technologies most suitable for the South African manufacturing industry.

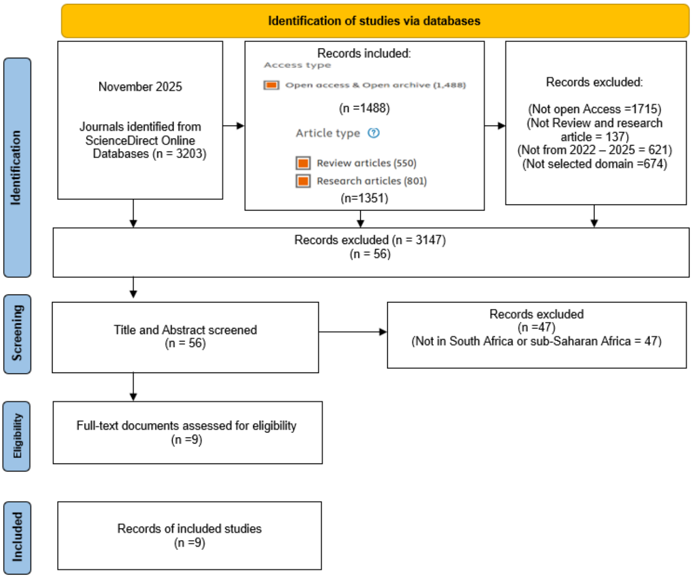
Research Methodology

A systematic literature review (SLR) was utilised as a method to evaluate existing literature. Focuses on studies that discuss SAI technologies in the SA manufacturing settings and were published during the previous five years. SLR was selected as the methodology due to its crucial role in academic research, which can synthesise key theoretical foundations and empirical results within a specific discipline, identify potential research opportunities, and formulate new theories, all aimed at enhancing knowledge (Webster & Watson, 2002). This review is essential for gaining a comprehensive understanding of the current approaches to AI-driven smart manufacturing technologies in industrial settings.

The credibility and relevance of the review were evaluated in this chapter using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), a standard for systematic reviews and meta-analyses. The PRISMA methodology facilitated a peer-reviewed, systematic approach for article selection, search techniques, data extraction, and data analysis procedures (Page et al., 2021). To improve the clarity and comprehensiveness of reporting in systematic reviews, PRISMA provides a visual representation of a systematic review through a four-phase flowchart, as illustrated in Figure 9.1.

Figure 9.1

Identification of the Selected Studies Adapted from PRISMA 2020



Note. From "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews", by M. J. Page et al., 2021, *BMJ, 372,* Article 71. (https://doi.org/10.1136/bmj.n71). Copyright 2021 BMJ Publishing Group Ltd.

Identification of Studies

Only relevant studies from 2022 to 2025 from ScienceDirect were included. ScienceDirect is a digital library that provides access to peer-reviewed journals, books, and articles across various scientific and technological fields. Due to the chapter's word limit, only one database was used for the literature analysis. Although other databases may be considered for future studies.

The search keyword was "Artificial Intelligence" OR "AI" AND "manufacturing sector" OR "industry" AND/OR "South Africa". The preliminary search yielded 3,203 publications in various categories. Filtering for open-access articles yielded 1,488 results. Narrowing the search to just peer-reviewed research and review publications yielded 1353 results. Limiting the search to five-year publications found 730 relevant publications. Only computer science, technology, social science, engineering, energy, humanities, market complexity, technological forecasting, and agriculture research were included, resulting in 56 papers. At least one publication type was chosen to address variations.

Screening of Titles and Abstracts

The selection was determined by screening titles and abstracts that were retrieved and imported into EndNote. The retrieved papers were analysed by screening titles and abstracts that conformed to the identified search term. Only papers contextualised in SA or sub-Saharan Africa were chosen after screening. Sub-Saharan Africa refers to the region of the continent located south of the Sahara Desert, including countries such as SA and others. Consequently, studies within the sub-Saharan region with a focus on SA are viable for analysis.

Eligibility of Studies

A total of nine papers were subsequently confirmed feasible for analysis. A summary of the selected eligible studies is presented in Figure 9.1.

Inclusion Studies

The articles selected for analysis are presented in Table 9.1.

Table 9.1

List of Articles Selected for Analysis

Paper_ ID	**Title and Author**	**Articles function with search parameters**		
		AI	Industry 4.0 or Manufacturing	SA or Sub-Sahara
P1	A critical review of the enablers and constraints of artificial intelligence in the South African public sector (Baloyi et al., 2025)	X		X
P2	Artificial intelligence and industry 4.0 and 5.0: a bibliometric study and research agenda (Fosso-Wamba & Guthrie, 2024)	X	X	
P3	The effects of digital transformation on innovation and productivity: Firm-level evidence of South African manufacturing micro and small enterprises (Gaglio et al., 2022)		X	X

P4	Synthesising the potential of artificial intelligence in the fulfilment of sustainable development goals in South Africa: An ethical perspective (Mapungwana & Chadyiwa, 2025)	X	X	
P5	Intelligent manufacturing eco-system: A post COVID-19 recovery and growth opportunity for manufacturing industry in Sub-Saharan countries (Mezgebe et al., 2023)		X	X
P6	The impact of Industry 4.0 on South Africa's manufacturing sector (Ngepah et al., 2024)		X	X
P7	Industry 4.0 concepts within the sub–Saharan African SME manufacturing sector (Peter et al., 2023)		X	X
P8	Transformation of South Africa's energy landscape: Policy implications, opportunities, and technological innovations in the Fourth Industrial Revolution (Ukoba et al., 2025)		X	X
P9	A systematic review of fourth industrial revolution technologies in smart irrigation: Constraints, opportunities, and future prospects for sub-Saharan Africa (Wanyama et al., 2024)		X	X

Table 9.2 summarizes the excluded and included studies from the analysis.

Table 9.2
Exclusion and Inclusion Criteria

Exclusions Studies	Inclusions Studies
Studies not in the context of SA	Contextualise for SA or Sub-Saharan
Published before 2022	2022-2025
Not Review and Research articles	Book chapter, conferences, seminars, etc.
Not Open access	Open access
Other discipline	Domain of computer science, technology, social science, engineering, energy, humanities, market complexity, technological forecasting, and smart agriculture
Not written in English	

The criteria for inclusion and exclusion in this systematic review were strictly enforced to ensure that only relevant studies on SAI technologies in SA manufacturing were included. The identified studies were categorized and analyzed to synthesize approaches regarding the role of SAI in the diverse manufacturing industry.

Findings and Discussion

The systematic literature review highlights the key trends, advantages, and challenges of employing SAI in various sectors of SA manufacturing. P1 reviewed the enablers and constraints to AI adoption for improving public-sector management in South Africa, examining potential opportunities and risks, as well as ethical issues and relevant policies and initiatives. P2 addressed how AI is used in Industry 4.0 for smart manufacturing systems and Industry 5.0 for processes that are sustainable and focused on people. P4 explored how AI affects sustainable development, notably in smart farming, predictive analytics for health and poverty, and related fields. P3, P5, P6, and P7 discussed the significance and industrial relevance of intelligent manufacturing technologies, evaluating the impact of digital transformation on innovation and productivity within South African SMEs and other disciplines. P8 aimed to analyse South Africa's energy environment, focusing on policy implications, the adoption of renewable energy, and the impact of Industrial 4.0 technologies on the industry. P9 Smart Irrigation study examines the capacity of Industrial 4.0 technology to address agricultural issues in sub-Saharan Africa. These studies together underscore the emerging yet increasing interest in SAI integration within South Africa's economic sectors, despite considerable obstacles.

The benefits of SAI in the Manufacturing Public Sector Adoption

AI demonstrates substantial potential for enhancing operational efficiency, streamlining service delivery, and decreasing administrative burdens (Mahusin et al., 2024). P1 attests that SAI adoption has several advantages in service delivery, for example, the implementation of the Automated Biometric Identification System, an AI-based system intended to match individuals' fingerprints, facial features, and palm prints, as well as robotics, facial recognition, and virtual agents (Baloyi et al., 2025; Marakalala & Matlala, 2024).

Challenges of SAI in the Manufacturing Public Sector Adoption

Despite the recognised potential for efficiency gains and enhanced service delivery, ethical and governance concerns, alongside a lack of tailored frameworks, significantly impede its full integration (Baloyi et al., 2025). A primary factor hindering the widespread adoption of SAI in SA is the lack of suitable or inadequate policies, legislation, and regulations governing digital technology (Rekunenko et al., 2025). Additionally, there are frequent ethical dilemmas caused by regulatory concerns. Such as data protection, privacy, security, accountability, openness, and public confidence, as ethical considerations (Chilunjika, 2024). As a result, public sector management in Africa is particularly vulnerable to cyberattacks and data hacking (Pieterse, 2021). Nevertheless, AI-driven technologies have been widely used in various industries like healthcare, education, transportation, and municipal services, greatly improving human lives and making public services more accessible (Alaran et al., 2025). Thus, they are lauded as an effective instrument for addressing massive public sector issues.

The Benefits of SAI in the Manufacturing Energy Sector Adoption

The energy sector in SA is undergoing a significant transformation, driven by technological innovations, policy changes, and the global shift towards sustainable energy systems (Langerman et al., 2023). The country has historically relied on coal for electricity generation, with more than 80% of its power supply derived from coal-fired plants (Pegels, 2010). This exhibits moderate advancement in the integration of Industry 4.0, largely propelled by initiatives in renewable energy. The sector is currently undergoing a significant transformation, with Industrial 4.0 technologies, including smart grids and decentralized energy systems, providing essential pathways for achieving sustainable energy objectives (Hassan et al., 2023). P8 elaborated how the existing renewable energy capacity comprises 2500 MW of solar photovoltaic and 3670 MW of wind energy, with projections aiming for 40% renewable energy by 2030 (Ukoba et al., 2025).

The Challenges of SAI in the Manufacturing Energy Sector Adoption

Nonetheless, the sector encounters considerable challenges due to its substantial dependence on coal, which constitutes 77% of electricity generation, whereas renewable energy accounts for merely 12% (Dhansay et al., 2017). Disparities in rural electrification persist, underscoring the need for inclusive energy solutions. The findings indicate that although policy frameworks are in place, there is a need to enhance regulatory frameworks to expedite the adoption of renewable energy and tackle energy poverty through decentralized systems.

The Benefits of SAI in the Agricultural Sector Adoption

The implementation of Industrial 4.0 technologies has transformed agricultural approaches globally. The adoption of smart irrigation, utilising AI, provides significant advantages in optimising water usage and enhancing crop yields (Formanek et al., 2024). Smart irrigation extensively integrates Industry 4.0 technologies, including drones, AI, the IoT, Big Data technology, and Blockchain (Wanyama et al., 2024).

These innovations enable the tracking of soil moisture and weather in real-time, allowing for the planning of irrigation with pinpoint accuracy, optimizing water distribution, and providing insight into how crops utilize water in real-time (Odhiambo et al., 2021). However, smart irrigation in agriculture particularly shows promise, but is not yet widely used in SA.

New irrigation technologies worldwide demonstrate potential for enhancing agricultural output and mitigating the effects of climate change, despite persistent water shortages (Lebek & Krueger, 2023). Integration of new technologies like 3D printing, drones, robots, blockchain, and the IoT is defining the Industrial 4.0 that the world is seeing right now (Jacoby, 2023). Smart and precise irrigation, made possible by these developing intelligent technologies, is set to revolutionize farming in sub-Saharan Africa (Nigussie et al., 2020). However, challenges persist in the adoption in sub-Saharan Africa.

The Challenges of SAI in the Agricultural Sector Adoption

P9 is evident that the lack of technical knowledge, inadequate infrastructure, and limited access to technology are significant obstacles to adoption (Wanyama et al., 2024). The high starting prices of these technologies also make it difficult to be widely adopted. Although the technology promise is there, the findings show that specific structural restrictions are limiting current usage. Nonetheless, their implementation in sub-Saharan Africa presents a significant challenge, despite the pressing requirement for sustainable and data-driven irrigation systems to secure food and promote economic development in the region (Wanyama et al., 2024).

The Benefits of SAI in the SMEs Manufacturing Sector

Selected digital communication technologies, such as social media and mobile phones for internet access, have a positive influence on innovation, which in turn enhances labour productivity, contingent upon the use of these technologies (Gaglio et al., 2022). This indicates fundamental digitalisation rather than a thorough implementation of Industry 4.0. P6 indicates that AI significantly enhances productivity, sustainability, and decision-making through the use of AI-driven machine learning models as part of Industry 4.0 initiatives. AI systems were integrated with IoT sensors installed on industrial equipment, such as IoT, robotics, and cyber-physical systems, thereby increasing productivity and output quality (Ngepah et al., 2024).

The Challenges of SAI in the SMEs Manufacturing Sector

In developing nations like SA, SMEs within the manufacturing sector are either not adopting or are slowly integrating Industry 4.0 approaches, resulting in decreased competitiveness. P5 and P7 indicated how the sector faces significant challenges in implementing Industry 4.0 among manufacturing SMEs in emerging economies (Mezgebe et al., 2023; Peter et al., 2023).

Lack of investment in these technologies, dealing with weak intellectual property rights, data privacy restrictions, Industry 4.0 specialised skills shortages, and local skills shortages (Peter et al., 2023). The Coronavirus Disease of 2019 pandemic significantly affected the manufacturing sector in sub-Saharan countries, delaying the adoption (Mezgebe et al., 2023). This indicates that the sector necessitates immediate intervention to avert additional competitive disadvantage.

Recommendation

AI offers significant economic and social benefits, but successful adoption in SA depends on addressing infrastructure gaps, skills shortages, and policy alignment (Mapungwana & Chadyiwa, 2025). P2 emphasises that SA requires a balance between the efficiency gains of Industry 4.0 and the human-centric principles of Industry 5.0. This will ensure that AI promotes economic development, sustainability, and social inclusion simultaneously (Fosso-Wamba & Guthrie, 2024). Consequently, SA needs to respond to the global adoption of Industry 4.0 by embracing innovative technologies and fostering a culture of continuous learning and adaptation. A summary of the South African SAI adoption challenges and benefits elaborates on how the diverse manufacturing sector in SA leverages SAI for competitiveness (see Table 9.3).

Table 9.3
SAI Adoption Challenges and Benefits in SA

Paper ID	Findings	Benefits	Challenges	Recommendation
P1	Machine learning (ML), robotics, facial recognition, virtual agents, and biometric identification systems	SAI offers efficiency, cost savings, productivity gains, and improved service delivery	AI adoption remains nascent and fragmented inadequate policies	Invest in the training/ reskilling of public servants, foster an innovative digital culture, ring-fence funding for AI and related digital infrastructure, and pursue targeted pilots before scaling up
P2	Deep Learning, Predictive Maintenance, and Data Mining	Better asset management and Improved decision-making across sectors	High capital costs for SMEs	Deep Learning, Predictive Maintenance, and Data Mining
P3	Digital communication technologies, including the use of social media and a business mobile phone	Positive effect on labor productivity	Accessibility on advanced digital technologies	Public programs aimed at fostering inclusive digitalization must consider the types of digital technologies that are most accessible and beneficial to small firms
P4	Automation, smart factories, and real time analytics	Analytics use for forecasting, fault detection. Reduced operational costs	High implementation and maintenance costs. Skills shortages in AI and data science	Provide incentives for SMEs to adopt AI. Invest in national data and computing infrastructure
P5	Intelligent technologies	Enhanced global competitiveness, adapted as a post-COVID-19 recovery and growth opportunity to enhance production processes of the manufacturing industry	Technological lag and pandemic impacts	Proposition of a Triple Helix Collaboration Eco-system that delineates a recursive contribution of Government, academia, and industry

P6	Automated decision-making fault detection, and maintenance planning	Increased productivity and output quality. Reduced operational cost	High cost of adoption for SMEs	Provide financial incentives for AI adoption
P7	A network digitization, automation	Reduced downtime, better resource utilization, supply chain visibility	Skills deficit for I4.0 specialists, high capital needs amid electricity shortages, low organizational readiness, and lack of standards/ training	Establish pan-African commissions for guidance/digital assessment; create pilot labs; run digital literacy programs; build foreign expert networks; develop leadership frameworks and global benchmarks
P8	Smart grids and decentralized energy systems	Sustainable energy in SA	Dependence on coal	Energy solutions and regulatory frameworks to expedite the adoption of renewable energy
P9	Data-driven irrigation systems, drones, the IoT, Big Data technology, and Blockchain	Secure food and promote economic development in the country. optimising water usage and enhancing crop yields	Regional infrastructural and economic challenges	Leverage existing mobile phone penetration for IoT data collection, collaborative partnerships, and innovative financing models

Conclusion

Regardless of diverse sectors, the selected studies show that inadequate infrastructure, lack of funding, and the need to increase capacity are the biggest obstacles to implementing SAI in SA. Manufacturing studies reveal challenges that necessitate systemic remedies, whereas energy and agriculture studies offer more optimistic projections, accompanied by specific investment and timelines. This indicates that the adoption of SAI in South Africa's industrial sector is crucial.

Limitations

The chapter focused solely on manufacturing, drawing insights from the selected literature papers, whereas a multi-sectoral analysis may provide more comprehensive insights.

Recommendation

In response to the global adoption of Industry 4.0, SA needs to respond by embracing innovative technologies and fostering a culture of continuous learning and adaptation. The global community, including China, has responded by developing initiatives that support the manufacturing industry in line with Industry

4.0 (Kang et al., 2016). This has led to increased efficiency, productivity, and competitiveness in the manufacturing sector. The outcome of this chapter may provide industry leaders with knowledge on adopting SAI to enhance operational efficiency, thereby contributing to improved economic viability and environmental sustainability.

References

Adams, D. (2023). Smart factory concept for an agri-processing plant in the Western Cape. *South African Journal of Industrial Engineering, 34*(3), 198–214. https://doi.org/10.7166/34-3-2950

Akoh, E. I. (2024). Adoption of artificial intelligence for manufacturing SMEs' growth and survival in South Africa: A systematic literature review. *International Journal of Research in Business and Social Science, 13*(6), 23–37. https://doi.org/10.20525/ijrbs.v13i6.3561

Alaran, M. A., Lawal, S. K., Jiya, M. H., Egya, S. A., Ahmed, M. M., Abdulsalam, A., Haruna, U. A., Musa, M. K., & Lucero-Prisno III, D. E. (2025). Challenges and opportunities of artificial intelligence in African health space. *Digital Health, 11*. https://doi.org/10.1177/20552076241305915

Baloyi, W. M., Meyer, N., & Rossouw, D. (2025). A critical review of the enablers and constraints of artificial intelligence in the South African public sector. *Journal of Contemporary Management, 22*(1), 380–403. https://doi.org/10.35683/jcman1185.302

Chilunjika, A. (2024). A review of the risks, challenges and benefits of using artificial intelligence (AI) technologies in public policy-making in South Africa. *Artificial Intelligence Social Sciences, Humanities and Education Journal, 53*, 393–411. https://hdl.handle.net/10210/512650

Dhansay, T., Musekiwa, C., Ntholi, T., Chevallier, L., Cole, D., & De Wit, M. J. (2017). South Africa's geothermal energy hotspots inferred from subsurface temperature and geology. *South African Journal of Science, 113*(11-12), 1–7. https://doi.org/10.17159/sajs.2017/20170092

Espina-Romero, L., Gutiérrez Hurtado, H., Ríos Parra, D., Vilchez Pirela, R. A., Talavera-Aguirre, R., & Ochoa-Díaz, A. (2024). Challenges and opportunities in the implementation of AI in manufacturing: A bibliometric analysis. *Sci, 6*(4), Article 60. https://doi.org/10.3390/sci6040060

Formanek, C., Tilbury, C. R., & Shock, J. P. (2024). *Opportunities of reinforcement learning in South Africa's just transition.* arXiv. https://arxiv.org/abs/2411.15145

Fosso-Wamba, S., & Guthrie, C. (2024). Artificial intelligence and industry 4.0 and 5.0: A bibliometric study and research agenda. *Procedia Computer Science, 239,* 718–725. https://doi.org/10.1016/j.procs.2024.06.228

Gaglio, C., Kraemer-Mbula, E., & Lorenz, E. (2022). The effects of digital transformation on innovation and productivity: Firm-level evidence of

South African manufacturing micro and small enterprises. *Technological Forecasting and Social Change, 182,* Article 121785. https://doi.org/10.1016/j.techfore.2022.121785

Hassan, Q., Sameen, A. Z., Salman, H. M., Al-Jiboory, A. K., & Jaszczur, M. (2023). *The role of renewable energy and artificial intelligence towards environmental sustainability and net zero.* Research Square. https://doi.org/10.21203/rs.3.rs-2970234/v1

Langerman, K. E., Garland, R. M., Feig, G., Mpanza, M., & Wernecke, B. (2023). South Africa's electricity disaster is an air quality disaster, too. *Clean Air Journal, 33*(1). https://doi.org/10.17159/caj/2023/33/1.15799

Lebek, K., & Krueger, T. (2023). Conventional and makeshift rainwater harvesting in rural South Africa: Exploring determinants for rainwater harvesting mode. *International Journal of Water Resources Development, 39*(1), 113–132. https://doi.org/10.1080/07900627.2021.1983778

Mahusin, N., Sallehudin, H., & Satar, N. S. M. (2024). Malaysia public sector challenges of implementation of artificial intelligence (AI). *IEEE Access, 12*, 121035–121051. https://doi.org/10.1109/ACCESS.2024.3448311

Maisiri, W., & Van Dyk, L. (2021). Industry 4.0 skills: A perspective of the South African manufacturing industry. *SA Journal of Human Resource Management, 19,* Article 1416. https://doi.org/10.4102/sajhrm.v19i0.1416

Maphisa, X., Nkadimeng, M., & Telukdarie, A. (2024). Contextual intelligence: An AI approach to manufacturing skills' forecasting. *Big Data and Cognitive Computing, 8*(9), Article 101. https://www.mdpi.com/2504-2289/8/9/101

Mapungwana, P., & Chadyiwa, M. (2025). Synthesising the potential of artificial intelligence in the fulfilment of sustainable development goals in South Africa: An ethical perspective. *Social Sciences & Humanities Open, 12,* Article 101883. https://doi.org/10.1016/j.ssaho.2025.101883

Marakalala, M. C., & Matlala, M. M. (2024). Border management identification: The biometric technology to detect criminals and terrorists often travel using falsified identity documents. *OIDA International Journal of Sustainable Development, 17*(12), 59–70. https://ssrn.com/abstract=5006155

Mezgebe, T. T., Gebreslassie, M. G., Sibhato, H., & Bahta, S. T. (2023). Intelligent manufacturing eco-system: A post COVID-19 recovery and growth opportunity for manufacturing industry in Sub-Saharan countries. *Scientific African, 19,* Article e01547. https://doi.org/10.1016/j.sciaf.2023.e01547

Ngepah, N., Saba, C. S., & Kajewole, D. O. (2024). The impact of industry 4.0 on South Africa's manufacturing sector. *Journal of Open Innovation: Technology, Market, and Complexity, 10*(1), Article 100226. https://doi.org/10.1016/j.joitmc.2024.100226

Nigussie, E., Olwal, T., Musumba, G., Tegegne, T., Lemma, A., & Mekuria, F. (2020). IoT-based irrigation management for smallholder farmers in rural sub-Saharan Africa. *Procedia Computer Science, 177*, 86–93. https://doi.org/10.1016/j.procs.2020.10.015

Odhiambo, K. O., Iro Ong'or, B. T., & Kanda, E. K. (2021). Optimization of rainwater harvesting system design for smallholder irrigation farmers in Kenya: A review. *AQUA – Water Infrastructure, Ecosystems and Society, 70*(4), 483–492. https://doi.org/10.2166/aqua.2021.087

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., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372,* Article 71. https://doi.org/10.1136/bmj.n71

Papadimitriou, I., Gialampoukidis, I., Vrochidis, S., & Kompatsiaris, I. (2024). AI methods in materials design, discovery and manufacturing: A review. *Computational Materials Science, 235,* Article 112793. https://doi.org/10.1016/j.commatsci.2024.112793

Pegels, A. (2010). Renewable energy in South Africa: Potentials, barriers and options for support. *Energy policy, 38*(9), 4945–4954. https://doi.org/10.1016/j.enpol.2010.03.077

Peter, O., Pradhan, A., & Mbohwa, C. (2023). Industry 4.0 concepts within the sub–Saharan African SME manufacturing sector. *Procedia Computer Science, 217,* 846–855. https://doi.org/10.1016/j.procs.2022.12.281

Philbeck, T., & Davis, N. (2018). The fourth industrial revolution: Shaping a new era. *Journal of International Affairs, 72*(1), 17–22. https://www.jstor.org/stable/26588339

Pieterse, H. (2021). The cyber threat landscape in South Africa: A 10-year review. *The African Journal of Information and Communication, 28*. https://doi.org/10.23962/10539/32213

Pypenko, I. S., & Melnyk, Yu. B. (2021). Principles of digitalisation of the state economy. *International Journal of Education and Science, 4*(1), 42–50. https://doi.org/10.26697/ijes.2021.1.5

Rekunenko, I., Kobushko, I., Dzydzyguri, O., Balahurovska, I., Yurynets, O., & Zhuk, O. (2025). The use of artificial intelligence in public administration: Bibliometric analysis. *Problems and Perspectives in Management, 23*(1), 209–224. https://doi.org/10.21511/ppm.23(1).2025.16

Ukoba, K., Jen, T.-C., & Yusuf, A. A. (2025). Transformation of South Africa's energy landscape: Policy implications, opportunities, and technological innovations in the fourth industrial revolution. *Energy Strategy Reviews, 59,* Article 101752. https://doi.org/10.1016/j.esr.2025.101752

Wanyama, J., Bwambale, E., Kiraga, S., Katimbo, A., Nakawuka, P., Kabenge, I., & Oluk, I. (2024). A systematic review of fourth industrial revolution technologies in smart irrigation: Constraints, opportunities, and future prospects for sub-Saharan Africa. *Smart Agricultural Technology, 7,* Article 100412. https://doi.org/10.1016/j.atech.2024.100412

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2). http://www.jstor.org/stable/4132319

Information about the authors:

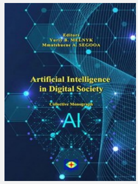
Mogoale Phumzile Mseteka – https://orcid.org/0000-0003-1770-5739; PhD, Dr, Postdoctoral Research Fellow, Tshwane University of Technology, Pretoria, South Africa.

Pretorius Agnieta Beatrijs – https://orcid.org/0000-0002-6510-2468; Doctor of Technologiae, Dr, Senior Lecturer and Assistant Dean, Tshwane University of Technology, Pretoria, South Africa.

Mogase Refilwe Constance – https://orcid.org/0000-0001-7337-8547; Doctor of Computing, Dr, Senior Lecturer and Head of Department, Tshwane University of Technology, Pretoria, South Africa.

Segooa Mmatshuene Anna – https://orcid.org/0000-0002-4190-8256; Doctor of Computing, Dr, Senior Lecturer, Tshwane University of Technology, Pretoria, South Africa.

Artificial Intelligence in Digital Society,
Volume 1, 2026
DOI: 10.26697/9786177089192.2026
ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)



Chapter 10. Human-Machine Collaboration in Sub-Saharan Africa: Bridging the Skills Gaps and Infrastructure Challenges

Bisha Z. [1], **Modiba F. S.** [1]

[1] *Nelson Mandela University, South Africa*

***Received:** 07.12.2025; **Accepted:** 10.02.2026; **Published:** 10.03.2026*

Abstract

Africa's large population faces high unemployment, underscoring the critical importance of livelihood strategies. Artificial intelligence's (AI) potential to transform work offers opportunities but also threatens the under-skilled workforce. Therefore, this study investigated Sub-Saharan Africa's efforts to ensure human-machine collaboration to address the challenges of unemployment and related issues of poverty and high inequality. A systematic literature review was conducted to examine studies published between 2020 and 2025. Using the Sustainable Development Goals (SDGs) framework and the capability approach, it analysed how human-machine collaboration influences progress toward achieving the United Nations Agenda 2030. Findings reveal that the region is unable to benefit from advanced technologies. Additionally, challenges related to infrastructure, digital and AI literacy, telecommunications, and transportation affect business success. With a willing youth population, the region offers universities opportunities to introduce AI curricula and forge private-sector partnerships that equip students with practical AI skills. These findings contribute to the digitalisation literature and highlight potential avenues for skilling and reskilling the SSA's workforce to coexist with AI systems. Policymakers should prioritise digital transformation to prevent inadequate infrastructure from hindering the region's development. Addressing these challenges creates opportunities for the region and accelerates progress toward the SDGs.

Keywords: artificial intelligence, human-machine collaboration, skills, employment, digital transformation, Sub-Saharan Africa.

Cite this chapter as:

Bisha, Z., & Modiba, F. S. (2026). Human-machine collaboration in Sub-Saharan Africa: Bridging the skills gaps and infrastructure challenges. In Y. B. Melnyk & M. A. Segooa (Eds.), *Artificial Intelligence in Digital Society, Vol. 1.* (pp. 146–159). KRPOCH. https://doi.org/10.26697/aids.2026.10

The electronic version of this chapter is complete. It can be found online in the AIDS Archive https://doi.org/10.26697/aids

Introduction

Human-Machine collaboration is a vital component of Work 5.0, highlighting the development of human capabilities and decision-making beyond their replacement by automation (Mourtzis et al., 2023). Human-machine collaborations are social-technical systems whereby humans and technology work together to complete tasks. This type of partnership is characterised by automation and technical autonomy (Jarrahi et al., 2023). Automation refers to instances in which technological systems replace tasks formerly performed by humans, generally to increase productivity by reducing direct human engagement in activities (Simmler & Frischkencht, 2021). However, substituting for human input, human-machine collaboration improves performance by integrating human and machine strengths, whereby each contributes to tasks for which they are best suited (Kolbeinsson et al., 2019). Others argue that these technologies threaten job security, especially in labour- intensive sectors (Mvile & Bishoge, 2024). Therefore, human-machine collaboration requires evaluating the tasks performed by humans and machines, as well as the degree of autonomy afforded to technological systems (Simmler & Frischkencht, 2021; Kolbeinsson et al., 2019).

This chapter investigates how human-machine collaboration can address socio- economic issues and identify gaps to be addressed in preparing Sub-Saharan Africa's (SSA) workforce to be artificial intelligence (AI) literate, not to threaten job security. The noted technical needs include infrastructure challenges, an unstable electricity supply, and limited digital connectivity, which hinder effective technological implementation in SSA (Bakibinga-Gaswaga et al., 2020). Human-machine collaboration is essential because automation creates structural job changes rather than displacement, as machines require human support (Vermeulen et al., 2018).

The impact of human-machine interaction has been widely researched (Kolbeinsson et al., 2019; Morutzis et al., 2023; Simmler & Frischkencht, 2021; Wang & Li, 2025), but research in the African context remains limited. This study, therefore, fills this gap by synthesising findings from SSA on the topic. It provides recommendations for enhancing human-machine collaboration using the SSA case. Policymakers can thus align digital transformation efforts in ways that will not entrench existing educational and digital inequalities. Hence, elevating human-machine collaboration is crucial to foster inclusive growth and digital transformation (Das, 2024; Modiba et al., 2024) in SSA, harnessing its demographic potential and ensuring that technology is developed with and for the people. Therefore, the following research questions are posed to guide the review:

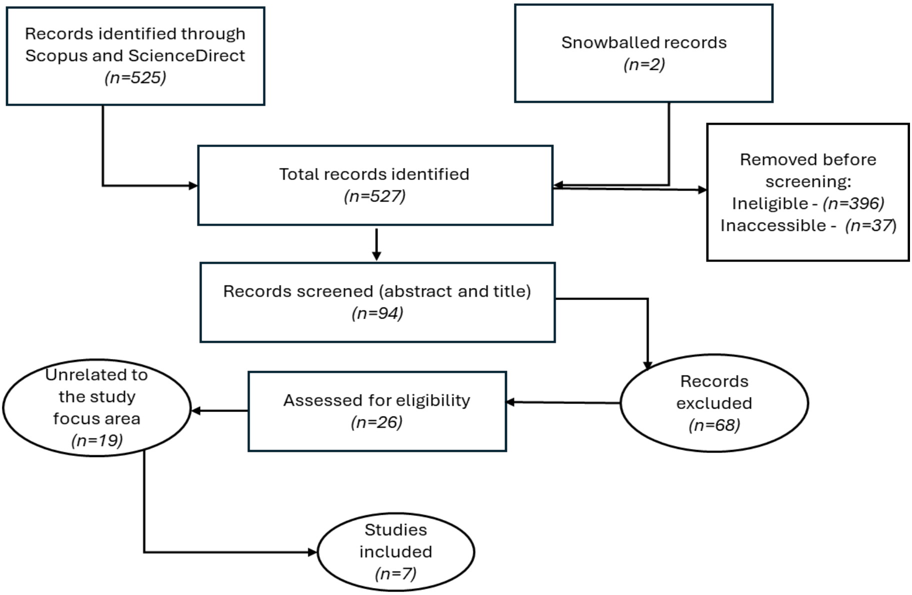
- What are the impediments to human-machine collaboration in SSA?
- How is human-machine collaboration embraced in SSA?
- How can human-machine collaboration address SSA's socio-economic issues?

Methods and Materials

Data collection for analysis entailed systematically retrieving peer-reviewed and grey literature. Materials were sourced from the ScienceDirect and Scopus databases and from Google for grey literature. Guided by the Preferred Items for Systematic and Meta-Analysis – PRISMA (Page et al., 2021), identified records were screened, appraised (Segooa et al., 2025), and seven were included in the study (see Figure 10.1).

Figure 10.1

Adapted PRISMA Flow Diagram



Note. Adapted from "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews" by Page et al., 2021, *BMJ, 372,* Article 71 (https://doi.org/10.1136/bmj.n71). Copyright 2021 BMJ Publishing Group Ltd.

Using the search string "human machine collaboration AND digital transformation AND Sub-Saharan Africa AND employment AND infrastructure," the databases were searched and filtered according to the inclusion and exclusion criteria in Table 10.1.

Table 10.1

Inclusion and Exclusion Criteria

Inclusion	Exclusion
- Sub-Saharan Africa - Studies on human-machine collaboration, AI integration, skills gap, workforce displacement, and digital transformation - English studies - Research articles - Period between 2020 and 2025 - Full-text available through institutional access	- Regions outside (SSA) - Studies not addressing human-machine collaboration, AI, workforce issues, digital transformation - Non-English studies - Conference proceedings, reviews, book chapters, encyclopaedias - Before 2020 and after 2026 - Inaccessible texts

Theoretical Framework

The capability approach suggests that people require specific skills to support their capacity to change their life situations (Sen, 2005). It also argued that this approach can be used to test how specific technologies impact people's lives (Modiba & Kaye, 2023). In the case of technologies such as AI, the capability approach can help identify skill deficiencies and how to address them to enable human-machine collaboration (Bobitan et al., 2024). The SDG framework presents 17 goals with corresponding targets (UN, 2025) that countries can use to track their progress in addressing challenges to sustainable development. For this study, SDGs 1 (no poverty), 2 (zero hunger), 4 (quality education), 8 (decent work and economic growth), 9 (industry, innovation, and infrastructure), and 10 (reduced inequalities) will be used to evaluate how the use of technology impacts sustainable development.

While other scholars use theories such as institutional enactment, systems theory, and the sustainable livelihood framework, sustainable development, diffusion of innovation, and resilient theory (Nahar, 2024; Wang & Li, 2025), these were not deemed suitable for this study, given that the human aspect is the key focus. Therefore, the SDG framework is used in conjunction with the capability approach to analyse the data using content and thematic analysis.

Socioeconomic Context of Human-Machine Collaboration

Human-machine collaboration uses technologies such as AI, Augmented Reality (AR), and Virtual Reality (VR) to enhance human skills and decision-making (Isaza & Cepa, 2024). These advanced technologies reconfigure tasks, automating, shifting, or eliminating roles, requiring workers to adapt (Agreli et al., 2021). Thus, AI can work alongside people rather than replace them (Hudson, 2025; Resh et al., 2025).

The Sub-Saharan African (SSA) labour market is characterised by informality, with 86% of jobs (excluding agriculture) under this sector (Hanine et al., 2024). Thus, technology adoption is constrained by cost considerations.

However, it is believed that labour-intensive industries such as manufacturing may adopt advanced technologies, thereby threatening technological and inclusive development (Arruda & Pimenta, 2024).

Sub-Saharan Africa's education systems face severe inequality and limited digital access; only 6% of schools have internet access, which is the lowest proportion globally (Langthaler & Bazafkan, 2020). This lack of internet access could exacerbate socio-technical disparities rooted in limited electricity access and expensive hardware and data costs, thereby impeding skill acquisition related to digital platforms and AI (Chakroun et al., 2019). Thus, improving educational equity is essential for technology adoption.

Sub-Saharan Africa's large youth population offers a demographic dividend if employment challenges are addressed (Mamphiswana & Bekele, 2020). The projected population growth of 2.5 billion by 2050 requires 1.1 billion new jobs (Hanine et al., 2024). However, there are concerns about whether youth are acquiring the skills required for the Fourth and Fifth Industrial Revolutions (Masilo, 2025). Realising this potential requires significant investment in human capital and in institutions capable of absorbing the workforce.

Skills Gaps and Intelligent Machines

An extensive skills gap separates the demands of the 4IR economy from the SSA workforce. To compete, they need to develop foundational and intermediate digital skills, including AI literacy (Banga & te Velde, 2019; Bobitan et al., 2024).

Technical skills, such as problem-solving and data analysis (Bashir & Daniels, 2022), and soft skills, such as judgment, communication, and adaptability (Chigbu & Makapela, 2025), are also critical. The skills challenge stems from the misalignment between educational and industry needs. Nevertheless, traditional education systems are often of low quality and fail to teach digital and problem-solving skills (Okoye et al., 2024). Therefore, there is a need to equip graduates with complementary digital, technical, and soft skills to alleviate the region's high unemployment.

Infrastructure Constraints and Digital Readiness

Energy access affects digital readiness. Only 70% of communities in SSA have access to electricity, which also affects broadband access, the rollout of digital technologies, and ICT use (Tryphone et al., 2023). Connectivity is uneven: while mobile broadband covers 81% of the population, only 30% are online, particularly in rural areas, underscoring the need for targeted infrastructure policies (Alper & Miktus, 2019).

Sub-Saharan Africa lags other regions in digital technology adoption, resulting in a digital divide (Astuti & Ayinde, 2025; Wang & Li, 2025). Human capital, infrastructure, and political stability all influence these disparities.

According to Das (2024), when organisations show a seamless integration of ICT, IoT, and AI, it can be assumed that digital transformation has been achieved. Therefore, where disparities exist, equitable access requires reconsideration.

Sectoral Experiences of Human-Machine Collaboration

According to Chigbu and Makapela (2025), human-machine collaboration underlies Industry 5.0 (I5.0), Education 5.0, and Work 5.0, emphasising capability augmentation rather than replacement. This collaboration leverages human strengths, such as decision-making and creativity, while AI automates routine tasks. However, risks such as deskilling and surveillance can threaten autonomy, requiring trust and transparency in design.

Automation is likely to emerge first in capital-rich sectors such as mining and high-wage manufacturing, where global firms already use robotic loaders and trucks (Gaus & Hoxtell, 2019). As noted by Anosike et al. (2024), Intelligent Agriculture (IA) is used for food security, leveraging technologies such as IoT. However, they face financial, technological, and political barriers. Moreover, small businesses in the region's manufacturing sector are adopting Industry 4.0 technologies, but still lag international competitors (Peter et al., 2023). Therefore, financial support is crucial for the adoption of AI and IA.

The adoption of AI in SSA public administration faces challenges related to accountability, inclusion, and integrity. However, e-governance is improving transparency, but there are concerns about marginalising public personnel (Plantinga, 2024). Artificial intelligence is transforming healthcare in SSA, increasing diagnostic accuracy (e.g., 92% for tuberculosis) and enabling predictive analysis to reduce outbreaks by up to 85% (Serge Andigema et al., 2025). AI-powered telemedicine also improves resource allocation and access to healthcare in low-resource areas.

Results and Discussions

The findings are presented in accordance with the research questions set out in the chapter. The use of advanced technologies remains generally low in the reviewed records, with the discussion centering on how AI technologies might be used to support various business processes, summarised in Table 10.2.

Impediments to Human-Machine Collaboration in SSA

Results from Kenyan small businesses indicate an interest in using AI tools, though they have not yet been adopted (SSA-2). The study highlighted limitations of current human-AI interactions, which are predominantly two-way and misaligned with the relational, decentralised structure of Kenyan businesses. This underscores the need for context-responsive, customised technologies. The latter is emphasised by prompt engineering, which may fail to collaborate with users if prompts are not carefully developed, thereby confirming the AI skills cited by Banga and te Velde (2019). It also signals a need for locally designed AI technologies. SSA-1, SSA-3 and SSA-4 cite a lack of skills as an impediment to adopting AI tools, thus affecting human-machine collaboration.

Table 10.2

Summary of Findings

Identifier	Source	Country	Key Findings
SSA-1	Dlamini & Ndzinisa (2025)	SSA	AI could increase existing social disparities. AI-supported education is required for digital literacy
SSA-2	Ankrah et al. (2025)	Kenya	The individualised interaction between humans and AI in the business sector is community-driven. Prompt engineering is limiting collaboration
SSA-3	Takawira & Pooe (2025)	South Africa	SME equipped for I5.0. Financial investment, skilled workers, and digital infrastructure are key enablers
SSA-4	Okoruwa et al. (2022)	SSA	Low-skilled workers are affecting technology adoption. Lack of infrastructure and the digital divide hinder the adoption of emerging technologies
SSA-5	Klenam et al. (2025)	SSA	AI is reducing the cost of additive manufacturing. Machine learning handling large datasets
SSA-6	David-Olawade et al. (2025)	Nigeria	Lack of AI foundational knowledge
SSA-7	Armar et al. (2025)	SSA	The adoption of advanced technologies in SSA faces significant hurdles. Improving delivery systems can address infrastructural limitations

SSA-6 highlights limited AI knowledge and awareness of AI applications as another challenge affecting this human-machine engagement. Infrastructural challenges constitute a significant hurdle to accessing these technologies (SSA-1; SSA-4; SSA-5; SSA-7). SSSA-1and SSA-7 also mention low internet usage in the region, aligning with Alper and Miktus (2019) and Tryphone et al. (2023). Financial resources and affordability were cited as another hindrance (SSA-3; SSA-4; SSA-7; Anosike et al., 2023). SSA-4 and SSA-6 argue that policy formulation processes and governance issues also affect the adoption of advanced technologies.

Embracing Human-Machine Collaboration

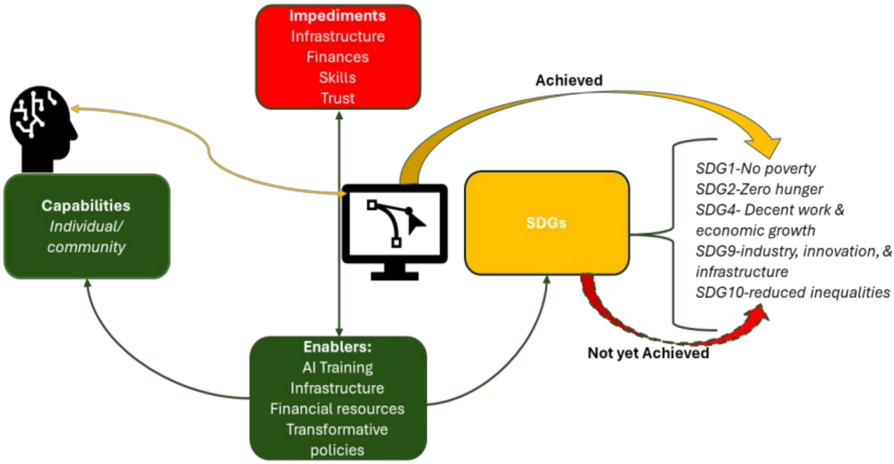
The Kenyan study highlights small businesses' interest in adopting AI technologies. SSA-1 report on AI-powered tools used in the region, such as chatbots to triage resources (Rwanda), drought forecasting drones (SA), blood delivery systems (Ghana), and satellite imagery and vulnerable group identification (Togo). They, however, emphasise the need for tools to be contextually relevant and designed to meet local needs, particularly the social capital valued by these businesses. As noted by SSA-1, the proliferation of AI technologies in the region can potentially exacerbate existing disparities. Like the digital divide that continues to hinder digital transformation in the least connected areas (Modiba et al., 2024), there is a need to manage the adoption of digital technologies. The proposed conceptualisation by SSA-2 presents a radical technology adoption. The collaboration between people and AI should therefore shift from individual-AI interaction to community-AI, thereby strengthening collaboration between humans and machines. In additive manufacturing (AM), AI is used to manage production and design processes and to handle large datasets (SSA-5). However, the collaborative aspect is still lacking. Healthcare students in Nigeria expressed interest in AI training and believed that its integration into healthcare would improve patient outcomes (SSA-6). The need for training supports Peter's (2023) findings.

Human-Machine Collaboration Potentially Addressing Socio-Economic Issues in SSA

It can be used to identify vulnerable people in need of aid, support the provision of quality education and health services, and help tackle the SDGs by generating inclusive solutions to address existing inequalities, provided AI models are well-trained (SSA-1). It can further assist with optimising work, addressing upskilling and data management challenges in resource-constrained business sectors (SSA-2; SSA-7). Work optimisation could be viewed as a threat to those who need to be absorbed in the labour market. Job creation and the formalisation of sectors are other benefits of these advanced technologies (SSA-4; SSA-5). SSA-3 argues that with I5.0 collaborative robots (cobots) present the ultimate human-machine collaboration that will improve operation of small businesses through enhanced employability with cobots working alongside human, improved productivity (SSA-7), and job satisfaction through the combination of human creativity and problem solving and machine precision, supporting literature findings of Chigbu and Makapela (2025); Hudson (2025) and Resh et al. (2025). Within the AM space, it can foster regional inclusive industrialisation and reduce dependence on imports for critical systems (SS-5). The above findings indicate that skills and infrastructure are major factors affecting human-machine collaboration. While the CA argues that with skills, people are able to create and adopt opportunities before them, the infrastructural issue is a significant impediment because it takes away possibilities for those without adequate infrastructures, limiting their abilities to acquire and sharpen their digital and AI skills to be prepared for the 5.0 to collaborate and co-create with AI tools. This means that this limitation affects some communities in SSA in addressing issues of poverty and zero through participation in the digital economy and access to quality education that equips them with such skills (SSA-1). Figure 10.2 illustrates the noted gaps that the digital transformation agenda could address and the SDGs that could be achieved through the proposed collaborations.

Figure 10.2

Human-Machine Collaboration for SDG Attainment



The framework shows that when capabilities are nurtured, humans can learn the requisite AI skills to work together with the tools to help solve socio-economic problems that humans understand well. Forging their creativity and machine competencies, the region can also develop technologies that resonate with communities' needs.

Conclusion

Human-machine collaboration is a key feature of Work 5.0 and is crucial for SSA, as it promotes augmentation and complementarity, enabling job creation rather than job displacement. Nevertheless, successful implementation faces substantial barriers: a significant skills gap and inadequate infrastructure, such as unreliable electricity and limited connectivity. These challenges may exacerbate the digital divide and hinder the region's capacity to develop contextually and culturally relevant AI tools.

Achieving inclusive development requires a multifaceted approach, including reforms to the education system, the development of digital and socio-behavioural skills, and investments in smart, integrated infrastructure, typically fostered through public-private partnerships. The future of work in SSA depends on ensuring that technology is developed for and with the people of the region.

This study's scope was limited to two academic databases. Future research should pursue comparative studies across SSA countries and use empirical methods to investigate the practical implementation of human-machine systems in various work contexts.

References

Agreli, H., Huising, R., & Peduzzi, M. (2021). Role reconfiguration: What ethnographic studies tell us about the implications of technological change for work and collaboration in healthcare. *BMJ Leader, 5,* 245–249. https://doi.org/10.1136/leader-2020-000344

Alper, M. E., & Miktus, M. (2019). Digital connectivity in sub-Saharan Africa: A comparative perspective. *International Monetary Fund.* https://www.imf.org/en/publications/wp/issues/2019/09/27/digital-connectivity-in-sub-saharan-africa-a-comparative-perspective-48692

Ankrah, E. A., Awori, K., Nyairo, S., Muchai, M., Ochieng, M., Kariuki, M., Geissler, J., & O'Neill, J. (2025, April). Social by nature: How socio-tecture shapes the work of SMBs and considerations for reimagining collaborative human-AI systems. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1–19). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642264

Anosike, A., Liravi, P., & Silas, U. (2024). A roadmap for intelligent agriculture in Africa: A case study of sub-Saharan Africa. *IEOM Society International*. https://index.ieomsociety.org/index.cfm/article/view/ID/14066

Arruda, E. P., & Pimenta, D. (2024). Challenges and implications of microwork in the age of artificial intelligence: A global socioeconomic analysis. *Human Resources Management and Services*, *6*(2), Article 3452. https://doi.org/10.54254/2753-8818/6/20243452

Astuti, H. M., & Ayinde, L. A. (2025). Uneven progress: Analyzing the factors behind digital technology adoption rates in Sub-Saharan Africa (SSA). *Data & Policy*, *7*, Article e23. https://doi.org/10.1017/dap.2024.47

Bakibinga-Gaswaga, E., Bakibinga, S., Bakibinga, D. B. M., & Bakibinga, P. (2020). Digital technologies in COVID-19 responses in sub-Saharan Africa: Policies, problems and promises. *The Pan African Medical Journal*, *35*(Suppl 2), Article 38. https://doi.org/10.11604/pamj.supp.2020.35.2.23456

Banga, K., & te Velde, D. W. (2019). *Preparing developing countries for the future of work: Understanding the skills ecosystem in a digital era*. Pathways Commission. https://pathwayscommission.bsg.ox.ac.uk/sites/default/files/2019-11/preparing_developing_countries.pdf

Bashir, S., & Daniels, C. (2022). Digital skills in Africa: Prospects for AU-EU collaboration. In *Africa-Europe cooperation and digital transformation* (pp. 184–198). Routledge. https://doi.org/10.4324/9781003274322

Bobitan, N., Dumitrescu, D., Popa, A. F., Sahlian, D. N., & Turlea, I. C. (2024). Shaping tomorrow: Anticipating skills requirements based on the integration of artificial intelligence in business organizations – A foresight analysis using the scenario method. *Electronics, 13*(11), Article 2198. https://doi.org/10.3390/electronics13112198

Chakroun, B., Miao, F., Mendes, V., Domiter, A., Fan, H., Kharkova, I., Avramova, E., & Rodriguez, S. (2019). *Artificial intelligence for sustainable development: Synthesis report, mobile learning week 2019*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000370308

Chigbu, B. I., & Makapela, S. L. (2025). AI in education, sustainability and the future of work: An integrative review of Industry 5.0, Education 5.0 and Work 5.0. *Journal of Open Innovation: Technology, Market and Complexity*, Article 100645. https://doi.org/10.1016/j.joitmc.2025.100645

Das, D. K. (2024). Exploring the symbiotic relationship between digital transformation, infrastructure, service delivery, and governance for smart sustainable cities. *Smart Cities*, *7*(2), 806–835. https://doi.org/10.3390/smartcities7020034

David-Olawade, A. C., Wada, O. Z., Adeniji, Y. J., Aderupoko, I. V., & Olawade, D. B. (2025). Artificial intelligence readiness among healthcare students in Nigeria: A cross-sectional study assessing knowledge gaps, exposure, and adoption willingness. *International Journal of Medical Informatics*, Article 106085. https://doi.org/10.1016/j.ijmedinf.2025.106085

Dlamini, R., & Ndzinisa, N. (2025). Towards a critical discourse on artificial intelligence and its misalignment in sub-Saharan Africa: Through an equality, equity, and decoloniality lens. *Journal of Education (University of KwaZulu-Natal)*, *98*, 42–61. https://doi.org/10.17159/2520-9868/i98a03

Gaus, A., & Hoxtell, W. (2019). *Automation in Sub-Saharan Africa: Is the Future of Work at Risk?* Konrad Adenauer Stiftung. https://www.kas.de/de/kurzum/detail/-/content/automation-in-sub-saharan-africa-is-the-future-of-work-at-risk

Hanine, S., Dinar, B., & Meftah, S. (2024). From tripalium to otium: What future for work in the era of disruptive technologies? *International Journal of Economic and Management Decisions*, *2*(4), 43–58. https://doi.org/10.62241/ijemd.24.4358.2345

Isaza, L., & Cepa, K. (2024). Automation and augmentation: A process study of how robotization shapes tasks of operational employees. *European Management Journal*, 1-14. https://doi.org/10.1016/j.emj.2024.01.003

Jarrahi, M. H., Lutz, C., Boyd, K., Oesterlund, C., & Willis, M. (2023). Artificial intelligence in the work context. *Journal of the Association for Information Science and Technology*, *74*(3), 303–310. https://doi.org/10.1002/asi.24730

Klenam, D. E. P., McBagonluri, F., Asumadu, T. K., Osafo, S. A., Bodunrin, M. O., Agyepong, L., Ojo, S. O., & Soboyejo, W. O. (2025). Additive manufacturing: Shaping the future of the manufacturing industry—Overview of trends, challenges and opportunities. *Applications in*

Engineering Science, Article 100224. https://doi.org/10.1016/j.apples.2025.100224

Kolbeinsson, A., Lagerstedt, E., & Lindblom, J. (2019). Foundation for a classification of collaboration levels for human-robot cooperation in manufacturing. *Production & Manufacturing Research*, *7*(1), 448–471. https://doi.org/10.1080/21693277.2019.1645628

Langthaler, M., & Bazafkan, H. (2020). *Digitalization, education and skills development in the Global South: An assessment of the debate with a focus on Sub-Saharan Africa* (ÖFSE Briefing Paper No. 28). Austrian Foundation for Development Research. https://ideas.repec.org/p/zbw/oefseb/28.html

Mamphiswana, R., & Bekele, M. (2020). *The fourth industrial revolution: Prospects and challenges for Africa*. Ethiopian Academy of Sciences. https://saiia.org.za/wp-content/uploads/2020/10/IAMOT-2020-Full-Paper-Submission-47.pdf

Masilo, M. (2025). Mathematics teaching for sustainable development: Challenges and successes. *Interdisciplinary Journal of Education Research*, *7*(2), Article a02. https://doi.org/10.38140/ijer-2025.vol7.2.02

Modiba, F. S., & Kaye, S. (2023). Evaluation of information and communication technologies (ICTs) tools contributing to rural development. *Russian Journal of Agricultural and Socio-Economic Sciences*, *142*(10), 19–29. https://doi.org/10.18551/rjoas.2023-10.03

Modiba, F. S., Musasa, G., Matindike, S., Kwanhi, T., Damiyano, D., & Mago, S. (2024). Can the digital economy transform financial inclusion in rural communities? A gendered lens. *Journal of Infrastructure, Policy and Development*, *8*(8), Article 3756. https://doi.org/10.24294/jipd.v8i8.3756

Mourtzis, D., Angelopoulos, J., & Panopoulos, N. (2023). The future of the human-machine interface (HMI) in society 5.0. *Future Internet*, *15*(5), Article 162. https://doi.org/10.3390/fi15050162

Mvile, B. N., & Bishoge, O. K. (2024). Mining and sustainable development goals in Africa. *Resources Policy*, *90*, Article 104710. https://doi.org/10.1016/j.resourpol.2024.104710

Nahar, S. (2024). Modeling the effects of artificial intelligence (AI)-based innovation on sustainable development goals (SDGs): Applying a system dynamics perspective in a cross-country setting. *Technological Forecasting and Social Change*, *201*, Article 123203. https://doi.org/10.1016/j.techfore.2024.123203

Okoruwa, V. O., Ogwang, T., & Ndung'u, N. S. (2022). *Regional views on the future of work: Sub-Saharan Africa*. African Economic Research Consortium. https://policycommons.net/artifacts/2243086/regional-views-on-the-future- ofwork/3001211/

Okoye, M. C., Hui, X., & David, A. M. (2025). Comparative analysis of technical and vocational education and training systems in China and Sub- Saharan

Africa for sustainable development. *Discover Education*, *4*(1), Article 535. https://doi.org/10.1007/s44217-025-00535-9

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., McGuinness, L. A., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*, Article 71. https://doi.org/10.1136/bmj.n71

Peter, O., Pradhan, A., & Mbohwa, C. (2023). Industry 4.0 concepts within the sub-Saharan African SME manufacturing sector. *Procedia Computer Science*, *217*, 846–855. https://doi.org/10.1016/j.procs.2022.12.281

Plantinga, P. (2024). Digital discretion and public administration in Africa: Implications for the use of artificial intelligence. *Information Development*, *40*(2), 332–352. https://doi.org/10.1177/02666669221150161

Resh, W. G., Ming, Y., Xia, X., Overton, M., Gürbüz, G. N., & De Bruhl, B. (2025). *Complementarity, augmentation, or substitutivity? The impact of generative artificial intelligence on the U.S. Federal Workforce.* arXiv. https://arxiv.org/abs/2503.09637

Sadik-Zada, E. R., & Jalabi, S. (2025). Powering agricultural revival: How solar-based irrigation is transforming Northeast Syria's war-torn fields. *The Electricity Journal*, *38*(2), Article 107471. https://doi.org/10.1016/j.tej.2025.107471

Segooa, M. A., Modiba, F. S., & Motjolopane, I. (2025). Generative artificial intelligence tools to augment teaching scientific research in postgraduate studies. *South African Journal of Higher Education*, *39*(1), 294–314. https://doi.org/10.20853/39-1-6345

Sen, A. (2005). Development as a capability expansion. In S. Fukuda-Parr & A. K. Shiva Kumar (Eds.), *Readings in human development: Concepts, measures and policies for a development paradigm* (2nd ed., pp. 3–16). Oxford University Press. https://archive.org/details/readingsinhumand0000unse/mode/2up

Serge Andigema, A., Tania Cyrielle, N. N., & Ekwelle, E. (2025). *Artificial intelligence in African healthcare: Catalysing innovation while confronting structural challenges* [Preprints] https://doi.org/10.20944/preprints202506.1824.v1

Simmler, M., & Frischknecht, R. (2021). A taxonomy of human-machine collaboration: Capturing automation and technical autonomy. *AI & Society*, *36*(1), 239–250. https://doi.org/10.1007/s00146-020-01004-z

Takawira, B., & Pooe, D. (2025). SME readiness for Industry 5.0: A systematic literature review. *The Southern African Journal of Entrepreneurship and*

Small Business Management, *17*(1), Article 946. https://doi.org/10.4102/sajesbm.v17i1.946

Tryphone, K., Joseph, C., & Ndanshau, M. O. (2023). Determinants of digital transformation in Sub-Saharan Africa: Some fiscal policy implications. *African Journal of Economic Review*, *11*(4), 34–48. https://ideas.repec.org/a/ags/afjecr/339655.html

United Nations. (2025). *The Sustainable Development Goals Report 2025*. https://unstats.un.org/sdgs/report/2025/

Vermeulen, B., Kesselhut, J., Pyka, A., & Saviotti, P. P. (2018). The impact of automation on employment: Just the usual structural change? *Sustainability*, *10*(5), Article 1661. https://doi.org/10.3390/su10051661

Wang, W., & Li, Q. (2025). Smart farming revolution: Leveraging machine learning for sustainable agriculture. *Journal of Cleaner Production*, *527*, Article 146434. https://doi.org/10.1016/j.jclepro.2025.146434

Information about the authors:

Bisha Zamagoba – https://orcid.org/0000-0001-7918-8197; BA Honours, MA Development Studies Student, Nelson Mandela University, Gqeberha, South Africa.

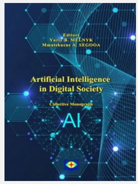
Modiba Florah Sewela – https://orcid.org/0000-0001-6905-067X; Doctor of Literature and Philosophy in Development Studies, Senior Lecturer, Department of Development Studies, Nelson Mandela University, Gqeberha, South Africa.

PART V

STRATEGIES FOR TRAINING SPECIALISTS IN THE DIGITAL SOCIETY USING ARTIFICIAL INTELLIGENCE

Check for updates

Artificial Intelligence in Digital Society,
Volume 1, 2026
DOI: 10.26697/9786177089192.2026
ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)



Chapter 11. Creating a Higher Education Ecosystem Based on Artificial Intelligence Implementation

Melnyk Y. B. [1,2], **Pypenko I. S.** [1,2]

[1] *Kharkiv Regional Public Organization "Culture of Health", Ukraine*
[2] *Scientific Research Institute KRPOCH, Ukraine*

Received: *03.12.2025;* ***Accepted:*** *10.02.2026;* ***Published:*** *10.03.2026*

Abstract
The artificial intelligence (AI) implementing in higher education started as a spontaneous process among all stakeholders. The study aims to explore the benefits and challenges of using AI in academic university teaching, and to develop and justify a model for the optimal implementation of AI for the development of the higher education ecosystem. The prospects of AI implementation for developing the higher education ecosystem are considered. The advantages and problems of using AI in academic university teaching are characterised based on the classification of directions of using AI in higher education. The model of optimal implementation of AI in the educational ecosystem of higher education, based on the systems approach, has been developed and substantiated. This model include structural (universities, faculties, departments, institutes, etc.) and functional (internal – content of education, forms and methods of teaching, diagnosing of learning outcomes, administering of educational service, and eternal – include academic achievement: levels of knowledge, skills, and competences) components. The results are essential for developing university strategies for developing educational ecosystem The curriculum should be relevant, meeting the interests of students and the current needs of employers. Education stakeholders are encouraged to use the available benefits of AI responsibly to address the challenges of student learning and teacher organisation in universities.

Keywords: artificial intelligence, higher education, Human-AI System, educational ecosystem, benefits and challenges of artificial intelligence, stakeholders in higher education

Cite this chapter as:
Melnyk, Y. B., & Pypenko, I. S. (2026). Creating a higher education ecosystem based on artificial intelligence implementation. In Y. B. Melnyk & M. A. Segooa (Eds.), *Artificial Intelligence in Digital Society, Vol. 1.* (pp. 161–173). KRPOCH. https://doi.org/10.26697/aids.2026.11
The electronic version of this chapter is complete. It can be found online in the AIDS Archive https://doi.org/10.26697/aids

Artificial intelligence technologies are becoming increasingly embedded in people's daily lives. A new system of relationships is emerging, the "Human-AI System" (Melnyk & Pypenko, 2023), which opens up prospects for further study and use of artificial intelligence (AI) in almost all areas of human activity. Education plays an important role in building sustainable development.

The interest of higher education in AI has increased significantly after the emergence of ChatGPT based on artificial intelligence. The accessibility and simplicity of this chatbot has made it extremely popular with all stakeholders in higher education (Baidoo-Anu & Ansah, 2023; Bonsu & Baffour-Koduah, 2023; Melnyk & Pypenko, 2024). However, this chapter is not limited to examining the use of ChatGPT in higher education. It focuses on exploring the issue more broadly.

When we use the term AI in our study, we mean computer systems, various AI technologies and applications, intelligent learning systems, chatbots, robotic and automated assessment systems that support and enhance education.

The chapter focuses on the benefits and challenges of using AI stakeholders in education. Special attention is paid to the model of optimal implementation of AI for building an educational ecosystem of higher education.

The study aims to explore the issues of benefits and challenges of using AI in academic university teaching, and to develop and justify a model of optimal implementation of AI for the development of the educational ecosystem of higher education.

A number of theoretical methods were used in the present study: analysis, synthesis, comparison, generalisation, systematization, classification to define the benefits and challenges of AI use by stakeholders; systems approach, modelling and optimisation methods to develop a model for the optimal implementation of AI in a higher educational ecosystem.

In the present study, we used internet resources to search for information based on the main concepts of AI in education, and analysed previous studies and reviews of periodicals. Studies published in scientific journals in a given field covered the following scientometric bases: Google Scholar, Education Resources Information Center (ERIC), Social Science Citation Index (SSCI), MDPI.

For the review, we selected English-language research studies on the use of AI in higher education that were published within the last 5 years in reputable scientific peer-reviewed journals from Web of Science and Scopus.

We used a search string that specified such selection criteria: "artificial intelligence", "higher education", "students and teachers", "diagnostic purposes", "assessing students", "providing feedback", "learning analytics", "special educational needs", "legitimacy of using AI-based chatbots".

Higher education is an open social system closely linked to advanced scientific research. Over the past five years, higher education has been greatly enriched by new AI technologies.

AI can have a number of applications in education: for assessing students (González-Calatayud et al., 2021; Hooda et al., 2022; Smerdon, 2024), for diagnostic purposes (Gupta et al., 2021), for providing feedback to students and teachers (Nazaretsky et al., 2024; Guo et al., 2024; Banihashem et al., 2024), thus ensuring continuous formative evaluation (Darvishi et al., 2022; Escalante et al., 2023).

Numerous studies show that AI can be used for personalised learning (Pratama et al., 2023; Kolchenko, 2018; Sajja et al., 2024), for gaming and active learning (Alam, 2022; Fachada et al., 2023; Kanja & Paschal, 2023).

Researchers believe that AI could also be used for students with special educational needs (Hopcan et al, 2023; Sharma et al, 2023; Chalkiadakis et al, 2024).

Studies have been conducted to investigate the use of AI for language learning by students (Divekar et al., 2022; Li, 2024; Han, 2024). This opens up the possibility of using AI to help international students overcome difficulties and facilitate their integration into different educational and cultural environments (Ma et al., 2024; Bannister et al., 2024; Wang, T., et al., 2023).

The use of AI for learning analytics (Ouyang et al., 2024; Ouyang et al., 2023; Salas-Pilco et al., 2022) and learning management (Ahmad et al., 2022; Dai et al., 2024; Chen et al., 2020) was also explored.

Thus, the education system is constantly enriched with new advanced technologies and methodological approaches, and innovative forms and methods of teaching are regularly introduced. This contributes both to the improvement (professional development) of teachers and to the involvement of students in the learning process, activating their cognitive processes and motivating their development. In addition, it provides employers with an influx of young, information technology-savvy professionals.

Among the new information technologies, it is worth mentioning those that open up fundamentally new possibilities: blockchain technology and artificial intelligence technologies.

Studies have described the benefits of implementing blockchain technology in various sectors, including higher education. According to the authors (Bhaskar et al., 2021; Pypenko & Melnyk, 2020; Raimundo et al., 2021), blockchain technology can be implemented in various areas of education to improve efficiency, effectiveness, privacy controls and technological enhancements. This is in line with today's requirements for the training of young professionals in universities.

Furthermore, according to Melnyk and Pypenko (2020), blockchain technology will facilitate the transition of education to a new, higher quality level. In the process of education, digital identifications are used. The whole education chain of those who study is systematised (school – university – production). All acts are realised in the consecutive order and agreed upon. The freedom of choice as for the goal, content, forms and methods of studying is considered. There is a possibility to choose a teacher/lecturer and the appropriate time for studying. The authors (Melnyk & Pypenko, 2020) believe that this modern technology will help people's nature. It will make the educational process easy, useful and interesting.

However, some researchers (Loukil, et al., 2021) note that despite the positive aspects of blockchain, several concerns continue to undermine its adoption in education, such as legal, immutability and scalability issues.

Next, let us take a look at artificial intelligence technology. Like blockchain technology, it has advantages and some disadvantages.

One of the most obvious and recognised by many researchers, problems with introducing AI into higher education is the issue of the ethics and legality of AI.

A number of studies have highlighted the need for ethical considerations and guidelines for the implementation of AI. A meta-review by Bond et al. (2024) identified research gaps that point to the need for greater ethical, methodological and contextual considerations in future research, as well as interdisciplinary approaches to the application of AI in higher education. Pisica et al. (2023) point to the need to control AI technologies in terms of careful monitoring, regulation and legislation to avoid ethical violations, privacy dilemmas and bias, and to adapt higher education stakeholders to new technologies and methods.

Next in importance, in our view, is the question of the right and legitimacy of using various AI technologies and applications, chatbots, in higher education.

The studies describe the challenges and benefits of implementing chatbots in higher education. Researchers (Abulibdeh et al., 2024) believe that in addition to ethical issues, AI-based chatbots such as ChatGPT will need to address curriculum revisions, continuous learning strategies and compliance with industry standards.

A study by Baidoo-Anu and Ansah (2023) note that among other benefits of ChatGPT, the chatbot promotes personalised and interactive learning, creates prompts for formative assessment activities that provide continuous feedback to inform teaching and learning, etc. This study highlights some inherent limitations of ChatGPT: creation of false information, data training bias, privacy issues, etc. This has been confirmed by other studies (Rasul et al., 2023) which investigated the benefits of the generative AI model, ChatGPT, in higher education and highlighted the following: the potential to facilitate adaptive learning, provide personalised feedback, support research and data analysis, provide automated administrative services, and help develop innovative assessments. Among the problems cited are concerns about academic integrity, reliability issues, inability to assess and reinforce graduate skills, limitations in assessing learning outcomes, and potential biases and distortions in information processing.

A solution to the problem of AI usability that stakeholders in higher education may face is seen by some researchers through the use of AI licensing, which is an important legal tool (Malgier & Pasquale, 2024). Licensing should be used in many high-risk areas of AI. They believe that ex-ante licensing of large-scale use of AI should become commonplace in jurisdictions committed to enabling democratic governance of AI.

Exploring the legitimacy of using AI-based chatbots in scientific research, Melnyk and Pypenko (2023) proposed a new method for indicating the involvement of AI and the role of chatbots in a scientific publication. Melnyk and Pypenko (2023) have developed a basic logo that can be used to indicate chatbot participation and contribution to publications. The authors have designed and implemented an information technology platform, AIC AI Chatbots, for practical applications. (https://doi.org/10.26697/ai.chatbots). It provides technological solutions related to using AI-based chatbots (text, image, and video) in scientific research and publishing.

When considering the issue of law, legitimacy and the use of attribution for AI, it is also useful to consider the protection of the rights of the individual who creates or performs work without AI. A study by Pypenko (2023) proposed attribution of a product created by humans without AI involvement. The author (Pypenko, 2023) believes that this helps to protect the human right to work and to increase the value of natural human labour.

Perhaps one of the most significant challenges slowing down the effective integration of AI in higher education is the profit orientation of app developers (Luckin & Cukurova, 2019). Developers rarely have the pedagogical background and didactic knowledge required to create a quality educational product.

As mentioned above, there have been many studies in recent years that have examined the use of AI in higher education. In many of them, the authors pointed to both benefits and problems for stakeholders.

The impact of distance learning and trends in using AI-based chatbots in higher education among stakeholders were explored (Aleedy et al., 2022; Al-Sharafi et al., 2023; Pypenko et al., 2020). These studies suggest that blended learning and the use of AI chatbots in higher education can be effectively used to assist students with their academic matters, progress monitoring, academic advice and administrative matters during their studies.

Others, such as Wang S. et al. (2023), argue that AI can enhance learning and provide personalised educational support. However, there are risks and limitations: confidentiality issues, cultural differences, linguistic competence and ethical implications.

Among other challenges to the use of AI in higher education, researchers highlight the following: privacy concerns, security and bias (Al-Zahrani & Alasmari, 2024); reliance on technology, lack of human touch, risk of cheating, displacement of teacher jobs (Clugston, 2024); lack of technology skills among students and teachers, and lack of applicability in different contexts, limited reliability (Celik et al., 2022; Crompton et al., 2022).

Among the advantages of using AI in higher education, researchers highlight the following:

- improving planning, implementation of immediate feedback and evaluation (Celik et al., 2022);

- minimising the administrative tasks of the educator, assisting with different types of tasks in the form of learning analytics, virtual reality and minimising the workload of the teacher, effective and easy assessment of students (Ahmad et al., 2022);

- facilitation of learning, personalised approach and feedback; effectiveness of AI tools and applications such as virtual and augmented reality, voice assistants, translation tools, chatbots, gamification, learning and tutoring programmes, instant assessment, etc. (Pisica et al., 2023);

- personalised learning, immersive learning experiences, improved student engagement and motivation, cost-effective learning, integrated learning and intelligent tutoring system, continuous evaluation and improvement over time, raising academic standards and quality of education (Clugston, 2024).

Pypenko (2024) proposed classifying the directions of implementing AI in higher education:

1. Content of education (e.g. development of training programmes, courses, topics).

2. Forms and methods of teaching (e.g. personalisation of learning and tutoring; a wide range of verbal, visual, gaming and other learning methods; innovative technologies such as virtual reality and augmented reality; translation tools; chatbots).

3. Diagnosing of learning outcomes (e.g. use of testing, quizzes, ease of student assessment, provision of continuous feedback).

4. Administering of educational services (e.g. developing competitive education strategies, optimising learning planning, data analysis, planning, record keeping, course selection, credit counting, using chatbots for marketing).

Undoubtedly, the described classification allows researchers studying the possibilities of AI implementation in higher education to systematise the advantages and problems of using AI in educational environment.

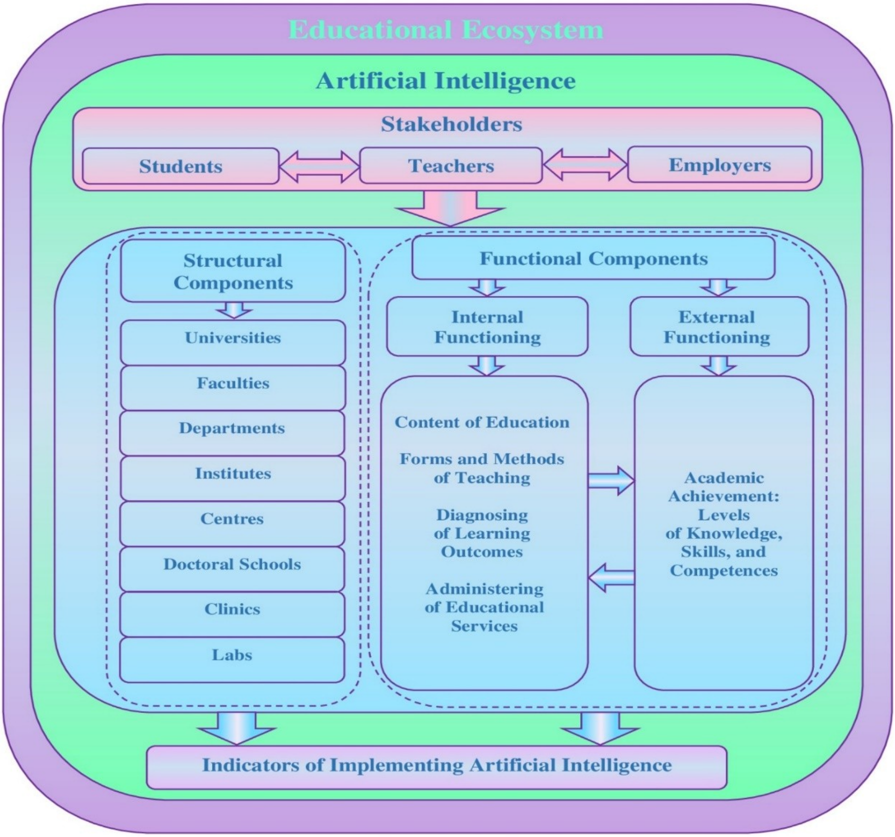
In our opinion, the methodology of the systems approach to the implementation of the above-mentioned AI directions in higher education will be the most optimal solution.

This allows each component of the system to operate both at a sub-system level and in conjunction with others to achieve maximum efficiency.

This concept required us to develop a model for the optimal implementation of AI in the higher education ecosystem (Melnyk & Pypenko, 2025). Figure 11.1 shows this model.

Figure 11.1

Model for the Optimal Implementation of Artificial Intelligence in a Higher Educational Ecosystem



The systemic approach to substantiate the model of optimal implementation of AI in the educational ecosystem of a higher school allowed us to identify the following three parameters: stakeholders, components of the educational ecosystem of a higher school, indicators of implementing artificial intelligence.

We have identified the following stakeholders of higher education: students, teachers, employers.

Using the systems approach methodology to substantiate this model allowed us to identify structural and functional components. Structural components include universities, faculties, departments, institutes, centres, doctoral schools, clinics, labs. Functional components are divided into two groups: internal functioning components and external functioning components. Internal functioning components include content of education; forms and methods of learning; diagnosing of learning outcomes; administering of educational services. Eternal functioning components include academic achievement: levels of knowledge, skills, and competences.

Indicators of the implementation of artificial intelligence make it possible to determine the level of effectiveness of the implementation of this model in practice.

Conclusions

There is growing concern about the ethical and legal implications of using AI in higher education systems. Educational stakeholders are encouraged to use the available benefits of AI responsibly and effectively to meet the challenges of student learning in higher education, taking into account the ethical and legal implications of its use. Addressing these challenges and regularly improving digital literacy in higher education will contribute to the development of advanced educational ecosystems.

University administrators should consider both the social demand from students and their own capacity to implement AI to deliver innovative study programmes. These programmes should be relevant and meet the current needs of employers. It is also important to pay attention to building the capacity of higher education stakeholders for the intensive AI development process in the near future.

References

Abulibdeh, A., Zaidan, E., & Abulibdeh, R. (2024). Navigating the confluence of artificial intelligence and education for sustainable development in the era of industry 4.0: Challenges, opportunities, and ethical dimensions. *Journal of Cleaner Production, 437*, Article 140527. https://doi.org/10.1016/j.jclepro.2023.140527

Ahmad, S. F., Alam, M. M., Rahmat, M. K., Mubarik, M. S., & Hyder, S. I. (2022). Academic and administrative role of artificial intelligence in education. *Sustainability, 14*(3), Article 1101. https://doi.org/10.3390/su14031101

Alam, A. (2022, April). A digital game based learning approach for effective curriculum transaction for teaching-learning of artificial intelligence and machine learning. *In 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)* (pp. 69–74). IEEE. https://doi.org/10.1109/ICSCDS53736.2022.9760932

Aleedy, M., Atwell, E., & Meshoul, S. (2022). Using AI chatbots in education: Recent advances challenges and use case. In M. Pandit, M. K. Gaur, P. S. Rana, & A. Tiwari. (Eds.), *Artificial Intelligence and Sustainable Computing. Algorithms for Intelligent Systems* (pp. 661–675). Springer. https://doi.org/10.1007/978-981-19-1653-3_50

Al-Sharafi, M. A., Al-Emran, M., Iranmanesh, M., Al-Qaysi, N., Iahad, N. A., & Arpaci, I. (2023). Understanding the impact of knowledge management factors on the sustainable use of AI-based chatbots for educational purposes using a hybrid SEM-ANN approach. *Interactive Learning Environments, 31*(10), 7491–7510. https://doi.org/10.1080/10494820.2022.2075014

Al-Zahrani, A.M., & Alasmari, T.M. (2024). Exploring the impact of artificial intelligence on higher education: The dynamics of ethical, social, and educational implications. *Humanities and Social Sciences Communications, 11*, Article 912. https://doi.org/10.1057/s41599-024-03432-4

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI, 7*(1), 52–62. https://doi.org/10.61969/jai.1337500

Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education, 21*(1), Article 23. https://doi.org/10.1186/s41239-024-00455-4

Bannister, P., Alcalde Peñalver, E., & Santamaría Urbieta, A. (2024). International students and generative artificial intelligence: A cross-cultural exploration of HE academic integrity policy. *Journal of International Students, 14*(3), 149–170. https://doi.org/10.32674/jis.v14i3.6277

Bhaskar, P., Tiwari, C. K. & Joshi, A. (2021). Blockchain in education management: Present and future applications. *Interactive Technology and Smart Education, 18*(1), 1-17. https://doi.org/10.1108/ITSE-07-2020-0102

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education, 21*, Article 4. https://doi.org/10.1186/s41239-023-00436-z

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: a systematic review of research. *TechTrends, 66*, 616–630. https://doi.org/10.1007/s11528-022-00715-y

Chalkiadakis, A., Seremetaki, A., Kanellou, A., Kallishi, M., Morfopoulou, A., Moraitaki, M., & Mastrokoukou, S. (2024). Impact of artificial intelligence and virtual reality on educational inclusion: A systematic review of technologies supporting students with disabilities. *Education Sciences, 14*(11), Article 1223. https://doi.org/10.3390/educsci14111223

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Clugston, B. (2024). *Advantages and disadvantages of AI in education*. https://www.ucanwest.ca/blog/education-careers-tips/advantages-and-disadvantages-of-ai-in-education/

Crompton, H., Jones, M. V., & Burke, D. (2022). Affordances and challenges of artificial intelligence in K-12 education: A systematic review. *Journal of Research on Technology in Education, 56*(3), 248–268. https://doi.org/10.1080/15391523.2022.2121344

Dai, R., Thomas, M. K. E., & Rawolle, S. (2024). The roles of AI and educational leaders in AI-assisted administrative decision-making: a proposed framework for symbiotic collaboration. *The Australian Educational Researcher*, 1-17. https://doi.org/10.1007/s13384-024-00771-8

Darvishi, A., Khosravi, H., Sadiq, S., & Gašević, D. (2022). Incorporating AI and learning analytics to build trustworthy peer assessment systems. *British Journal of Educational Technology, 53*(4), 844–875. https://doi.org/10.1111/bjet.13233

Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen, Y., Zhu, H., Hendler, J. A., & Braasch, J. (2022). Foreign language acquisition via artificial intelligence and extended reality: Design and evaluation. *Computer Assisted Language Learning, 35*(9), 2332–2360. https://doi.org/10.1080/09588221.2021.1879162

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education, 20*(1), Article 57. https://doi.org/10.1186/s41239-023-00425-2

Fachada, N., Barreiros, F. F., Lopes, P., & Fonseca, M. (2023, August). Active learning prototypes for teaching game AI. *In 2023 IEEE Conference on Games (CoG)* (pp. 1–4). IEEE. https://doi.org/10.1109/CoG57401.2023.10333229

González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial Intelligence for Student Assessment: A Systematic Review. *Applied Sciences, 11*(12), Article 5467. https://doi.org/10.3390/app11125467

Guo, K., Zhang, E. D., Li, D., & Yu, S. (2024). Using AI-supported peer review to enhance feedback literacy: An investigation of students' revision of feedback on peers' essays. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13540

Gupta, P., Yadav, D., & Dey, R. (2021). AI diagnosis: Rise of AI-powered assessments in modern education systems. *Transnational Marketing Journal, 9*(3), 625–633. https://www.ceeol.com/search/article-detail?id=998415

Han, Z. (2024). ChatGPT in and for second language acquisition: a call for systematic research. *Studies in Second Language Acquisition, 46*(2), 301–306. https://doi.org/10.1017/S0272263124000111

Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering, 2022*(1), Article 5215722. https://doi.org/10.1155/2022/5215722

Hopcan, S., Polat, E., Ozturk, M. E., & Ozturk, L. (2023). Artificial intelligence in special education: A systematic review. *Interactive Learning Environments, 31*(10), 7335–7353. https://doi.org/10.1080/10494820.2022.2067186

Kanja, M. W., & Paschal, M. J. (2023). AI game activities for teaching and learning. *In Creative AI Tools and Ethical Implications in Teaching and Learning* (pp. 153–167). IGI Global. https://doi.org/10.4018/979-8-3693-0205-7.ch008

Kolchenko, V. (2018). Can modern AI replace teachers? Not so fast! Artificial intelligence and adaptive learning: personalized education in the AI age. *HAPS Educator, 22*(3), 249–252. https://doi.org/10.21692/haps.2018.032

Li, Y. (2024). Usability of ChatGPT in second language acquisition: Capabilities, effectiveness, applications, challenges, and solutions. *Studies in Applied Linguistics and TESOL, 24*(1). https://doi.org/10.52214/salt.v24i1.12864

Loukil, F., Abed, M. & Boukadi, K. (2021). Blockchain adoption in education: a systematic literature review. *Education and Information Technologies, 26*, 5779–5797. https://doi.org/10.1007/s10639-021-10481-8

Luckin, R., & Cukurova, M. (2019). Designing educational technologies in the age of AI: A learning sciences-driven approach. *British Journal of Educational Technology, 50*(6), 2824–2838. https://doi.org/10.1111/bjet.12861

Ma, D., Akram, H., & Chen, I. H. (2024). Artificial intelligence in higher education: a cross-cultural examination of students' behavioral intentions and attitudes. *International Review of Research in Open and Distributed Learning, 25*(3), 134–157. https://doi.org/10.19173/irrodl.v25i3.7703

Malgieri, G., & Pasquale, F. (2024). Licensing high-risk artificial intelligence: toward ex ante justification for a disruptive technology. *Computer Law & Security Review, 52*, Article 105899. https://doi.org/10.1016/j.clsr.2023.105899

Melnyk, Yu. B., & Pypenko, I. S. (2024). Artificial intelligence as a factor revolutionizing higher education. *International Journal of Science Annals, 7*(1), 8–16. https://doi.org/10.26697/ijsa.2024.1.2

Melnyk, Yu. B., & Pypenko, I. S. (2020). How will blockchain technology change education future?! *International Journal of Science Annals, 3*(1), 5–6. https://doi.org/10.26697/ijsa.2020.1.1

Melnyk, Y. B., & Pypenko, I. S. (2025). Implementing of artificial intelligence in a higher educational ecosystem. *International Journal of Science Annals, 8*(1), 13–20. https://doi.org/10.26697/ijsa.2025.1.1

Melnyk, Yu. B., & Pypenko, I. S. (2023). The legitimacy of artificial intelligence and the role of ChatBots in scientific publications. *International Journal of Science Annals, 6*(1), 5–10. https://doi.org/10.26697/ijsa.2023.1.1

Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2024). AI or Human? Evaluating student feedback perceptions in higher education. In Ferreira Mello, R., Rummel, N., Jivet, I., Pishtari, G., & Ruipérez Valiente, J.A. (Eds.), *Lecture Notes in Computer Science: Vol. 15159. Technology Enhanced Learning for Inclusive and Equitable Quality Education* (pp. 284–298). Springer. https://doi.org/10.1007/978-3-031-72315-5_20

Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education, 20*(1), Article 4. https://doi.org/10.1186/s41239-022-00372-4

Ouyang, F., & Zhang, L. (2024). AI-driven learning analytics applications and tools in computer-supported collaborative learning: A systematic review. *Educational Research Review, 44*, Article 100616. https://doi.org/10.1016/j.edurev.2024.100616

Pisica, A. I., Edu, T., Zaharia, R. M., & Zaharia, R. (2023). Implementing artificial intelligence in higher education: Pros and cons from the perspectives of academics. *Societies, 13*(5), Article 118. https://doi.org/10.3390/soc13050118

Pratama, M. P., Sampelolo, R., & Lura, H. (2023). Revolutionizing education: harnessing the power of artificial intelligence for personalized learning. *Klasikal: Journal of Education, Language Teaching and Science, 5*(2), 350–357. https://doi.org/10.52208/klasikal.v5i2.877

Pypenko, I. S. (2024). Benefits and challenges of using artificial intelligence by stakeholders in higher education. *International Journal of Science Annals,* 7(2), 1–9. https://doi.org/10.26697/ijsa.2024.2.7

Pypenko, I. S. (2023). Human and artificial intelligence interaction. *International Journal of Science Annals, 6*(2), 54–56. https://doi.org/10.26697/ijsa.2023.2.7

Pypenko, I. S., Maslov, Yu. V., & Melnyk, Yu. B. (2020). The impact of social distancing measures on higher education stakeholders. *International Journal of Science Annals, 3*(2), 9–14. https://doi.org/10.26697/ijsa.2020.2.2

Pypenko, I. S., & Melnyk, Yu. B. (2020). Creating a business ecosystem based on blockchain technology. *International Journal of Education and Science, 3*(4), 53. https://doi.org/10.26697/ijes.2020.4.26

Raimundo, R., & Rosário, A. (2021). Blockchain system in the higher education. European Journal of Investigation in Health, *Psychology and Education, 11*(1), 276–293. https://doi.org/10.3390/ejihpe11010021

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., Sun, M., Day, I., Rather, R. A., & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1), 41–56. https://doi.org/10.37074/jalt.2023.6.1.29

Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., & Demir, I. (2024). Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education. *Information, 15*(10), Article 596. https://doi.org/10.3390/info15100596

Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial intelligence and learning analytics in teacher education: A systematic review. *Education Sciences, 12*(8), Article 569. https://doi.org/10.3390/educsci12080569

Sharma, S., Tomar, V., Yadav, N., & Aggarwal, M. (2023). Impact of AI-based special education on educators and students. *In AI-Assisted Special Education for Students with Exceptional Needs* (pp. 47–66). IGI Global. https://doi.org/10.4018/979-8-3693-0378-8.ch00

Smerdon, D. (2024). AI in essay-based assessment: Student adoption, usage, and performance. *Computers and Education: Artificial Intelligence, 7*, Article 100288. https://doi.org/10.1016/j.caeai.2024.100288

Wang, S., Wang, H., Jiang, Y., Li, P., & Yang, W. (2023). Understanding students' participation of intelligent teaching: An empirical study considering artificial intelligence usefulness, interactive reward, satisfaction, university support and enjoyment. *Interactive Learning Environments, 31*(9), 5633–5649. https://doi.org/10.1080/10494820.2021.2012813

Wang, T., Lund, B. D., Marengo, A., Pagano, A., Mannuru, N. R., Teel, Z. A., & 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), Article 6716. https://doi.org/10.3390/app13116716

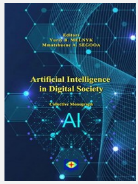
Information about the authors:

Melnyk Yuriy Borysovych – https://orcid.org/0000-0002-8527-4638; Doctor of Philosophy in Pedagogy, Affiliated Associate Professor; Chairman of Board, Kharkiv Regional Public Organization "Culture of Health" (KRPOCH); Director, Scientific Research Institute KRPOCH, Ukraine.

Pypenko Iryna Sergiivna – https://orcid.org/0000-0001-5083-540X; Doctor of Philosophy in Economics, Affiliated Associate Professor, Secretary of Board, Kharkiv Regional Public Organization "Culture of Health"; Scientific Research Institute KRPOCH, Ukraine.

Check for updates

Artificial Intelligence in Digital Society,
Volume 1, 2026
DOI: 10.26697/9786177089192.2026
ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)



Chapter 12. Generative Artificial Intelligence in South Africa's Higher Education: Assessing Readiness and Responsible Adoption

Modiba F. S. [1], **Segooa M. A.** [2], **Motjolopane I.** [1]

[1] *Nelson Mandela University, South Africa*
[2] *Tshwane University of Technology, South Africa*

Received: *03.12.2025;* ***Accepted:*** *10.02.2026;* ***Published:*** *10.03.2026*

Abstract

The use of Generative Artificial Intelligence (AI) presents both opportunities and challenges in the South African higher education sector, particularly when guidelines for its use are lacking. The absence of comprehensive policies and frameworks is problematic, as it enables the unethical deployment of these tools and fosters inappropriate institutional responses to their use. This chapter aims to explore the readiness and levels of Generative AI adoption in South African universities. The study used a systematic literature review to research the phenomenon. Data were sourced from databases such as ScienceDirect and Scopus, as well as institutional reports and policies, to ensure comprehensive coverage of the topic under investigation, and analysed using content analysis guided by the Generative AI maturity framework. The results highlight varying levels of adoption, from exploration to implementation. Therefore, this study presents a framework for institutions to assess their Generative AI readiness and to identify gaps, thereby informing the formulation of policies and guidelines for the use of these tools. The study contributes to the limited literature on universities' readiness to foster a supportive environment for Generative AI tools in higher education. Additionally, it offers practical guidelines for policymakers to address potential readiness and adoption gaps.

Keywords: generative artificial intelligence, responsible artificial intelligence, higher education, readiness, South Africa.

Cite this chapter as:

Modiba, F. S., Segooa, M. A., & Motjolopane, I. (2026). Generative artificial intelligence in South Africa's higher education: Assessing readiness and responsible adoption. In Y. B. Melnyk & M. A. Segooa (Eds.), *Artificial Intelligence in Digital Society, Vol. 1.* (pp. 174–187). KRPOCH. https://doi.org/10.26697/aids.2026.12

The electronic version of this chapter is complete. It can be found online in the AIDS Archive https://doi.org/10.26697/aids

Introduction

Generative Artificial Intelligence (AI) has transformed the world of work and learning, offering tools that can help users to be more efficient. They are used in industry and higher education institutions (HEIs), and the latter are expected to be partners in developing Generative tools that will respond effectively to various business and academic contexts (Crumbly et al., 2025). The use of these tools can enhance students' digital skills and prepare them for the future of work. Research has shown that students are interested in learning to use Generative AI tools so they can apply these skills in their future employment (Rispler et al, 2025). Cardon et al. (2023) have however indicated that the policy stance in the organisation is the one that encourages employees to use Generative AI, meaning that clear policies on the use of these tools will influence uptake in any environment. Similarly, the adoption and use of Generative AI depend on HEIs' policy frameworks.

In HEIs, the use of Generative AI is closely linked to its responsible use (Rasul et al., 2025), as students may still use the technology even when explicitly instructed not to. Therefore, policies are imperative because they guide users of these technologies in using them responsibly. According to Alba et al. (2025), HEI policies aim to address academic integrity and ethical considerations to prevent plagiarism and unauthorised assistance, particularly in universities that permit the use of Generative AI. Policy and guideline formulation are important for producing graduates skilled in Generative AI who also understand that ethical aptitude in using these tools is the cornerstone of responsible use. Nevertheless, at some universities, students are unaware of their institution's guidelines on Generative AI (Al Zaidy, 2024). It also argued that these policy documents must be continually adapted as tools advance rapidly; therefore, regular review is necessary (Alba et al., 2025).

Recent studies have explored Generative AI's potential to, improve student learning experiences (Megbowon, 2025), and enhance the research process for postgraduate research (Segooa et al., 2025). The new technologies of generative artificial intelligence have been the factors that have revolutionised the industry of higher education (Melnyk & Pypenko, 2024). Despite increasing interest in the topic and rising expectations for institutions to develop policy guidelines to leverage Generative AI, there is a notable gap in studies documenting the presence of institution-wide policy guidelines in South African Higher Education (Chaka et al., 2024; Sadiq et al., 2021). Therefore, the current study builds on studies on Generative AI policy and academic frameworks that could guide institutional adoption of these tools. This study simplified the South African National AI Policy Framework (SANAIF) by assigning categories, thereby enabling the expansion of pillars aligned with institutional policy objectives (Department of Communications and Digital Technologies, DCDT, 2024). This contribution helps align the principles required for AI policy and guideline formulation across various institutions of higher learning. To achieve the purpose of this study, the following research questions are set to guide this study:

- What institutional readiness measures are required to support the use of Generative AI in South African universities?
- What are the levels of readiness in South African HEIs in the adoption of Generative AI?
- What are the factors that influence the adoption of Generative AI in South African institutions of higher learning?

Related Studies

While theoretical framing, such as technology organisation environment, the technology acceptance model, and diffusion of innovation theory, are usually used for technology-related studies (Depietro et al, 1990; Davis, 1989; Rogers, 2003). This study uses the Generative AI maturity framework to guide its analysis.

Generative AI frameworks are more relevant because they are better aligned with specific AI innovations than generic technological ones (Chukhlomin, 2024; Sadiq et al., 2021). The five phases that informed the assessment of the maturity level for the South African public university are demonstrated in Table 12.1.

Table 12.1

Adapted Readiness Levels for Assessment of the South African Public Institution

Phases	Description
Awareness	No Generative AI policy or guideline has been published.
Experimentation	Generative AI policy or guidelines exist.
Implementation	Students and academics are encouraged to use Generative AI tools at the individual level.
Integration	Generative AI tools are integrated with some of the university's workflow systems.
Transformation	The university integrates Generative AI tools into most of its workflows and processes.

Note. From "Gen AI maturity framework report: A comprehensive roadmap for organisations to evaluate and elevate their Generative AI capabilities" by AIM Research, 2024 (https://aimresearch.co/generative-ai-maturity-framework). Copyright 2024 AIM Media House LLC.

From "Generative AI capability maturity model for online and adult learning: Introducing the EMERALD-GenAI-CMM-OAL framework" by Chukhlomin V., 2024 (https://doi.org/10.2139/ssrn.4769557). Copyright 2024 Elsevier Inc.

From "Artificial intelligence maturity model: A systematic literature review" by Sadiq et al., 2021, *PeerJ Computer Science, 7,* Article e661 (https://doi.org/10.7717/peerj-cs.661). Copyright 2021 PeerJ.

The adoption of Generative AI in some areas is determined by institutional factors, which, in this study, are linked to the levels of readiness shown in Table 1. For example, Aldreabi et al. (2025) found that institutional support, ease of use, and access to digital technology were significant determinants of students' adoption of Generative AI.

Moreover, students' perceptions of these learning tools influence their adoption; for example, viewing Generative AI as an assistant has been associated with greater acceptance of the technology (Kanont et al., 2024).

However, several concerns may hinder the adoption of Generative AI, including rapid technological change, risks to academic integrity from student overreliance, and inadequate regulation, which can lead to bias and inaccuracy (Hughes et al., 2025). Educators' lack of confidence is a significant barrier, highlighting the need for targeted support (Kohnke et al., 2023).

Other adoption issues, as noted by Malacaria et al. (2023), include operational challenges related to infrastructure, maintenance and monitoring that affect tool use, as well as workforce competencies that limit optimal use of the tools. Complex interfaces and resource constraints are also cited as contributing factors (Weinberg, 2025).

Cordero et al. (2024) suggest the need for clear ethical guidelines, the development of effective prompts, ongoing development, and staff training to ensure that Generative AI is ethically incorporated into teaching practices. This process should be accompanied by constant monitoring and evaluation. Moreso, when policies are formulated, issues of copyright, data protection and ethical implications of Generative AI use within the institution should be considered.

Methods and Materials

The study employed a qualitative review using a three-phase approach to identify grey literature and academic sources, as depicted in the PRISMA flow diagram in Figure 12.1.

The first phase involved using Google Search and higher education institutions to identify policies and guidelines on Generative AI.

The second phase involved searching for peer-reviewed journal articles in databases such as Scopus and ScienceDirect published between 2021 and 2025.

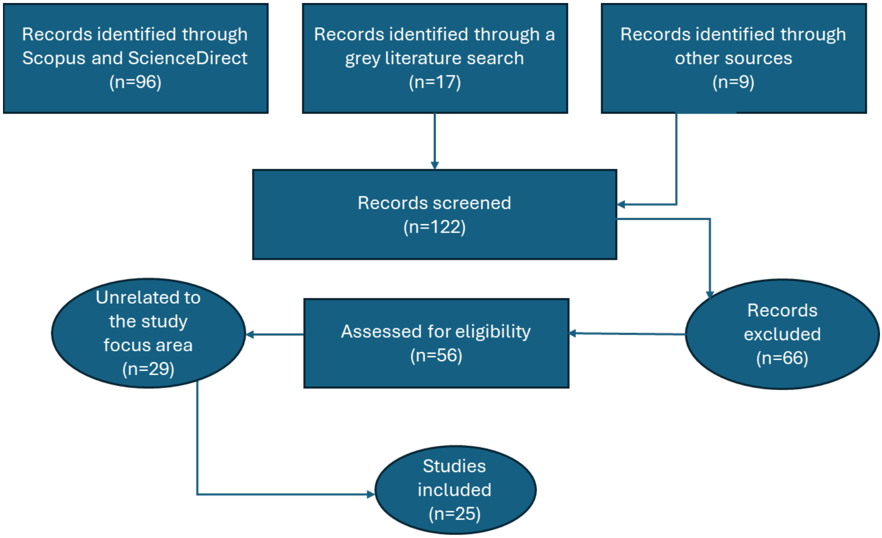
The last phase employed a snowball sampling approach to identify additional documents and empirical studies on Generative AI.

The policies were sourced using the search string: "Generative Artificial Intelligence" AND "institutions of higher learning" OR "Education" AND "South Africa" AND "AI Policy" OR "Generative AI guidelines".

Seventeen legal frameworks, in the form of policies and guidelines, were sourced from university websites and analysed for their readiness for Generative AI, including the Digital and Communication AI framework, to understand national Generative AI priorities.

Figure 12.1

Adapted PRISMA Flow Diagram Source



Note. Adapted from "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews" by Page et al., 2021, *BMJ, 372,* Article 71 (https://doi.org/10.1136/bmj.n71). Copyright 2021 BMJ Publishing Group Ltd.

A similar search string was applied to Scopus, in accordance with the inclusion and exclusion criteria in Table 12.2; the search returned 28 entries. When applying the country filter to South Africa, two records were returned. However, upon screening, the two records were excluded because they focused on Sub-Saharan Africa and Zimbabwe.

Table 12.2

Inclusion and Exclusion Criteria

Inclusion	Exclusion
Journal articles	Conference proceedings and journal preprints
Period between 2021 and 2024	Before 2021
Must focus on South African institutions of higher learning	Not focusing on South African Higher education
Must include Generative AI usage and policy frameworks	Not focusing on Generative AI

ScienceDirect yielded 68 records; after applying the period and article type filters, 57 were screened. After conducting the quality appraisal, 10 articles met the inclusion criteria, of which eight were accessible. Of the eight reviewed articles, only one met the inclusion criteria. All phases led to the inclusion of 25 sources for review. The generated data were analysed using content analysis guided by the Generative AI framework.

Results and Discussions

This section presents findings from grey literature and peer-reviewed sources (Table 12.3).

Table 12.3

Summary of Analysed Papers

Source	Focus	Methods	Theory	Summary of findings
Otto, 2024	Generative AI in pedagogy	Qualitative	Constructive theoretical approach	Prompts training for students. Polices clarifying AI usage. Clarification on AI plagiarism.
Mogoale et al. (2025)	Education 5.0	SLR	None	Collaborative AI policy formulation. Training needs for Generative AI skills. Ethical competencies for using these tools.
Chaka et al., (2024)	Institution-wide policies on AI	Literature review	None	Ethical concerns, academic Integrity, transparency, accountability and ownership, privacy, security, and safety.
Megbowon (2025)	Students' perceptions of Gener ative AI	Qualitative	Technology Acceptance Model	Concern about its ethical use. Lack of trust. A need for guidelines, awareness and training.
Mbangeleli and Funda (2024)	Generative AI in South African HEIs	SLR	None	Infrastructure and digital divide concerns. Instilling principles of integrity. Facilitating responsible use of AI.
Patel and Ragolane (2024)	Opportunities and challenges of Generative AI	Quantitative	None	Relevant infrastructure HEIs' and stakeholders' collaboration are essential for responsible usage.
Xulu et al. (2024).	ChatGPT adoption in a university	Quantitative	Technology organisation environment	Academic integrity Lack of policy frameworks. Improper adoption of these tools.
Mithi (2024)	The impact of Generative AI on formative assessments	Qualitative	None	Lack of training on responsible usage and ethical concerns. Educators' knowledge on Generative AI Guiding students on how to use these tools.

All 26 public universities were assessed and categorised as traditional, comprehensive, or universities of technology. The results indicate that most public HEIs have policies and guidelines to govern the ethical use of Generative AI. Of the 26 reviewed universities, 17 had guidelines. However, guidelines could not be identified on public platforms for the remaining nine universities: two universities of technology (UoTs), three traditional universities (TUs), and four comprehensive universities (CUs). One of the institutions has regulations, but they are not publicly available to external users. However, those without published guidelines demonstrated awareness by hosting academic events on Generative AI (Monono, 2024). Additionally, research papers on the use of Generative AI in such institutions were identified (Xulu et al., 2024; Mithi et al., 2024). Moreover, training in the use of Grammarly was offered in one of the comprehensive universities.

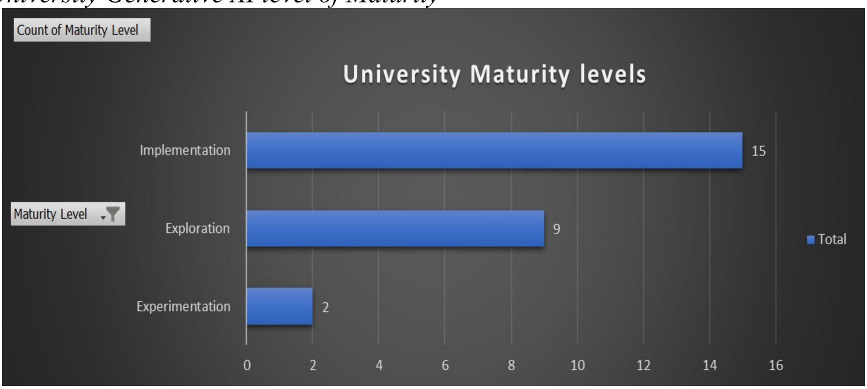
The available guidelines vary. Two traditional universities had extended guidelines to include educators and researchers, thereby providing a holistic policy. Other universities provide guidelines for students on the use of Generative AI in academic work, such as assignments and tests. Students are guided in using the tools and in what is unacceptable (three TUs; three CUs and one UoT), thereby promoting the ethical use of these technologies. Additionally, students are required to submit a declaration that the work submitted aligns with the Generative AI guidelines; this was observed in two TUs, one CU and one UoT.

Those at an advanced stage of exploring Generative AI were encouraging lecturers to include AI in the syllabus (three TUs, two CUs and one UoT). Others emphasised the importance of equipping students with AI skills and maintained dedicated Generative AI sites that outlined guidelines, available tools, and how they could support various tasks. Intellectual property (IP) guidelines on Generative AI in research, which help staff manage IP-related issues, were also noted (TU). Similarly, one TU encouraged academics to comply with data privacy legislation and leverage guidelines from Harvard and the University of Cape Town. One CU also had copyright guidelines.

The results also suggested a need to train staff and students in Generative AI skills, enabling them to understand the benefits and risks involved (Kohnke et al., 2023; Mbangeleli & Funda, 2024; Mithi et al., 2024; Mogoale et al., 2025). The need for students to be trained to use prompts was further emphasised (Otto, 2024), consistent with the findings of Cordero et al. (2024) and Kohnke et al. (2023). Such knowledge also helps users recognise the misuse of these tools and limit overdependence on them for academic and research purposes (Mithi, 2024), instead of using them as collaborators and co-creators of content.

Figure 12.2

University Generative AI level of Maturity



The findings in Figure 12.2 depict universities' readiness, as defined by the Generative AI maturity framework, comprising three phases. Level 1, Exploration with nine frequencies; they remain at the awareness stage, with no guidelines for generative AI. Nevertheless, there is evidence of academic events on Generative AI hosted by these universities. Level 2, Experimentation with two universities testing these tools. Level 3, Implementation with 15 universities engaging with Generative AI tools, enabling students and staff to use them.

Figure 12.3

Generative AI Maturity Classification by Type

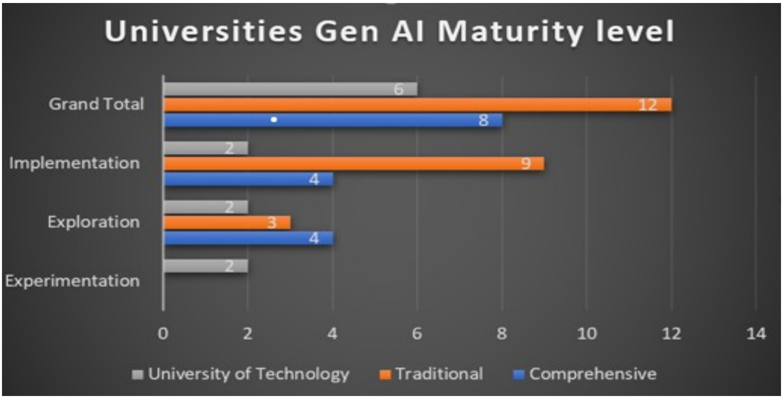


Figure 12.3 highlights institutions at the implementation level (TUs with eight frequencies), whereas UoTs appear to cross-cut Exploration, Experimentation, and Implementation, with two frequencies each. Comprehensive universities also display a frequency of four institutions on Exploration and Implementation, respectively.

Issues of the digital divide and inadequate infrastructure highlighted gaps in physical capital affecting some universities (Mbangeleli & Funda, 2024; Patel & Ragolane, 2024). This finding suggests a need for infrastructure support for universities to prevent the digital divide from becoming entrenched at the university level and to help address the country's structural challenges. It also confirms the findings of Malacaria et al. (2023). Some HEIs adopt AI haphazardly, resulting in fragmented and inconsistent use of AI tools (Patel & Ragolane, 2024).

Ethical principles were addressed in all reviewed studies, with a focus on the ethical use of tools, academic integrity, data privacy, transparency, and accountability as factors affecting adoption (Chaka et al., 2024; Otto, 2024; Mogoale et al., 2025). Additionally, privacy, security and safety can be compromised when users are not educated about tools. Moreover, accountability, transparency, and ownership can be compromised when users are poorly informed. Thus, showing the interconnectedness between ethical principles and human capital. Similarly, Cordero et al. (2024) and Mithi et al. (2024) suggest that clear guidelines are necessary to avoid irresponsible adoption or use of Generative AI (Xulu et al., 2024; Megbowon, 2025).

The human approach to AI was framed as requiring training to address ethical concerns. Additionally, the argument that educators must implement countermeasures to address the lack of academic integrity (Mithi, 2024) reflects a soft approach to addressing AI-dependent students. Advocating and facilitating adherence to the principles of integrity and the ethical use of these tools could help address challenges related to cheating and academic integrity (Mbangeleli & Funda, 2024), aligning with Chaka et al. (2024) and Hughes et al. (2025).

Proposed Framework

Some universities in South Africa are interested in Generative AI and are already engaging with these tools. Some have developed guidelines to help staff and students understand the tools, particularly their benefits and risks, and to educate them on how to use them ethically and responsibly. However, some universities lacked published guidelines, particularly those classified as UoTs, despite being actively engaged in technology and planning to strengthen their Generative AI activities. The guidelines of the two CUs could also not be found. However, according to Alba et al. (2025), instances like this do not reflect a lack of Generative AI use, as educators may have course-level guidelines. This finding is also evident at CU and UoT, where empirical evidence suggests the use of tools (Mithi et al., 2024; Xulu et al., 2024). Some traditional universities also lacked guidelines, showing that even such institutions can lag. It was also noted that institutions such as Thensa have been instrumental in helping UoTs and CUs close gaps in Generative AI.

Based on these results, a South African Generative AI Readiness Framework (SA-GAIRF) is proposed in Figure 12.4.

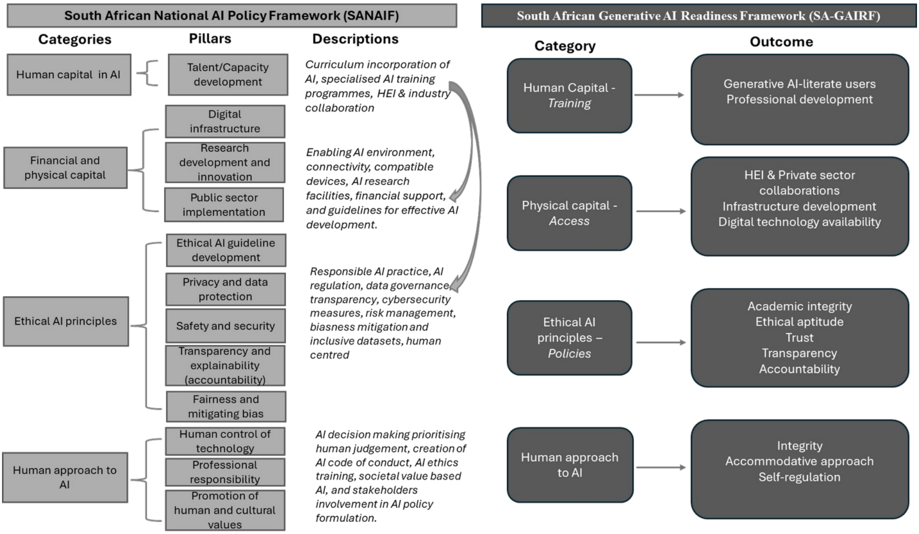
It illustrates alignment with the categories of human capital, physical capital, ethical principles, and the human approach to AI, as discussed in the introduction. This mapping to SANAIF showed the financial capital as the only element not emphasised as a factor in the results.

This study suggests that universities should formulate policies, guidelines, or statements on the ethical use of Generative AI to leverage the opportunities these tools offer. Challenges related to human capital can be addressed through capacity development for both staff and students, thereby strengthening integrity and self-regulation when the tools are used. The framework can be used to formulate a university-wide guideline on the ethical use of Generative AI.

Moreover, educators and students who adopt Generative AI in the absence of their university's guidelines may tailor their course-level user declarations, informed by the SA-GAIRF categories three and four, and aligned with the respective SANAIF pillars (DCDT, 2024).

Figure 12.4

Adapted Generative AI Framework for HEIs



Note. Adapted from "South Africa national AI policy framework" by DCDT, 2024 (https://www.dcdt.gov.za/sa-national-ai-policy-framework/file/338-sa-national-ai-policy-framework.html). Copyright 2024 Department of Communications & Digital Technologies.

Conclusion

The Generative AI policy landscape in SA shows positive progress. The key readiness factor for integrating and adopting Generative AI is the availability of guidelines in HEIs. Additionally, including a declaration or statement acknowledging the use of Generative AI tools contributes to academic integrity and the ethical use of these tools. When guidelines are unavailable, educators can develop course-specific guidelines to help students develop generative AI skills. The development of such guidelines can also assist educators in participating institutionally in the policymaking process. The involvement of educators at this level would ensure that classroom experiences are articulated and accommodated in the policy process. While human capital is important, financial and physical access are essential to ensure supportive infrastructure and AI-powered systems that promote inclusive access.

This study advances limited research on universities' readiness to ensure a supportive environment for Generative AI tools in higher education. It further provides policymakers with practical guidelines for addressing potential readiness and adoption gaps. Additionally, it offers higher education institutions in the global

South specific guidelines to support the adoption of Generative AI, taking into account infrastructural, human capacity, and ethical considerations that must be factored in for the inclusive and responsible use of these tools. However, its limitations include the use of only two databases; access to additional databases could have enabled more studies. Future studies could draw on additional databases and conduct empirical research to develop a deeper understanding of the readiness factors influencing guideline development and the effective use of Generative AI tools in South African Universities.

References

AIM Research. (2024). *Gen AI maturity framework report: A comprehensive roadmap for organisations to evaluate and elevate their Generative AI capabilities*. https://aimresearch.co/generative-ai-maturity-framework

Alba, C., Xi, W., Wang, C., & An, R. (2025, February 26–March 1). *ChatGPT comes to campus: Unveiling core themes in AI policies across US universities with large language models.* In Proceedings of the 56th ACM Technical Symposium on Computer Science Education (Vol. 2, pp. 1359–1360). ACM Conference, Pittsburgh, PA, United States. https://doi.org/10.1145/3641555.3705141

Aldreabi, H., Dahdoul, N.K.S., Alhur, M., Alzboun, N. and Alsalhi, N.R., 2025. Determinants of Student Adoption of Generative AI in Higher Education. *Electronic Journal of e-Learning, 23*(1), 15–33. https://doi.org/10.34190/ejel.23.1.3599

Al Zaidy, A. (2024). The impact of generative AI on student engagement and ethics in higher education. *Journal of Information Technology, Cybersecurity, and Artificial Intelligence, 1*(1), 30–38. https://doi.org/10.70715/jitcai.2024.v1.i1.004

Chaka, C., Shange, T., Nkhobo, T., & Hlatshwayo, V. (2024). An environmental review of the generative artificial intelligence policies and guidelines of South African higher education institutions: A content analysis. *International Journal of Learning, Teaching and Educational Research, 23*(12), 487–511. https://doi.org/10.26803/ijlter.23.12.25

Crumbly, J., Pal, R., & Altay, N. (2025). A classification framework for generative artificial intelligence for social good. *Technovation*, *139*, Article 103129. https://doi.org/10.1016/j.technovation.2024.103129

Chukhlomin, V. (2024). *Generative AI capability maturity model for online and adult learning: Introducing the EMERALD-GenAI-CMM-OAL framework.* https://doi.org/10.2139/ssrn.4769557

Cordero, J., Torres-Zambrano, J., & Cordero-Castillo, A. (2024). Integration of generative artificial intelligence in higher education: Best practices. *Education Sciences, 15*(1), Article 32. https://doi.org/10.3390/educsci15010032

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–339. https://doi.org/10.2307/249008

DePietro, R., Wiarda, E., & Fleischer, M. (1990). The context for change: Organization, technology and environment. In L. G. Tornatzky & M. Fleischer (Eds.), *The Processes of Technological Innovation* (pp. 151–175). Lexington Books. https://archive.org/details/processesoftechn0000torn

Department of Communications and Digital Technologies. (2024). *South Africa national AI policy framework.* DCDT. https://www.dcdt.gov.za/sa-national-ai-policy-framework/file/338-sa-national-ai-policy-framework.html

Hughes, L., Malik, T., Dettmer, S., Al-Busaidi, A. S., & Dwivedi, Y. K. (2025). Reimagining higher education: Navigating the challenges of generative AI adoption. *Information Systems Frontiers,* 1–23. https://doi.org/10.1007/s10796-025-10582-6

Kanont, K., Pingmuang, P., Simasathien, T., Wisnuwong, S., Wiwatsiripong, B., Poonpirome, K., Songkram, N., & Khlaisang, J. (2024). Generative-AI, a learning assistant? Factors influencing higher-ed students' technology acceptance. *Electronic Journal of e-Learning, 22*(6), 18–33. https://doi.org/10.34190/ejel.22.6.3196

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Exploring generative artificial intelligence preparedness among university language instructors: A case study. *Computers and Education: Artificial Intelligence, 5*, Article 100156. https://doi.org/10.1016/j.caeai.2023.100156

Malacaria, S., Grimaldi, M., Greco, M., & De Mauro, A. (2023). Business talk: Harnessing generative AI with data analytics maturity. *International Journal on Cybernetics & Informatics, 12*(7), 1–10. https://doi.org/10.5121/ijci.2023.120701

Mbangeleli, N., & Funda, V. (2024*).* Mapping the evidence around the use of AI-powered tools in South African universities: A systematic review. *Proceedings of the 1st International Conference on Education Research, 1*(1), 158–167. https://doi.org/10.34190/icer.1.1.3180

Melnyk, Yu. B., & Pypenko, I. S. (2024). Artificial intelligence as a factor revolutionizing higher education. *International Journal of Science Annals, 7*(1), 8–16. https://doi.org/10.26697/ijsa.2024.1.2

Mithi, J., Madzvamuse, S., Mbanje, S., & Lomahoza, S. (2024, November 4–6). Generative artificial intelligence and formative assessment: Perspectives from higher education in South Africa. *Proceedings of the 1st International Conference on Education Research, 1*(1), 449–458. https://doi.org/10.34190/icer.1.1.3231

Mogoale, P. D., Pretorius, A., Mogase, R. C., & Segooa, M. A. (2025). Integrating artificial intelligence within South African higher learning institutions.

South African Journal of Information Management, 27(1), Article 1939. https://doi.org/10.4102/sajim.v27i1.1939

Monono, K. (2024, November 18). *AI Expo Africa 2024: Navigating the intersection of AI, cybersecurity and innovation*. Tshwane University of Technology. https://aiexpoafrica.com/ai-expo-africa-2024-welcomes-international-vendors-ngos/

Otto, L. (2024). Assessing the use of ChatGPT as a pedagogical tool: A small study. *Africa Education Review, 20*(6), 81–96. https://doi.org/10.1080/18146627.2025.2471272

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., McGuinness, L. A., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*, Article 71. https://doi.org/10.1136/bmj.n71

Patel, S., & Ragolane, M. (2024). The implementation of artificial intelligence in South African higher education institutions: Opportunities and challenges. *Technium Education and Humanities*, 9, 51–65. https://doi.org/10.47577/teh.v9i.11452

Rasul, T., Nair, S., Kalendra, D., Balaji, M. S., de Oliveira Santini, F., Ladeira, W. J., Islam, J. U., Hammami, S., & Hossain, M. U. (2024). Enhancing academic integrity among students in GenAI era: A holistic framework. *The International Journal of Management Education, 22*(3), Article 101041. https://doi.org/10.1016/j.ijme.2024.101041

Rispler, C., Eizenberg, M. M., & Yakov, G. (2025). Understanding students' perceptions of generative AI: Implications for pedagogy and graduate employability. *Journal of Teaching and Learning for Graduate Employability, 16*(1), 145–170. https://doi.org/10.21153/jtlge2025vol16no1art2084

Rogers, E. M. (2003). Diffusion of innovations (5th ed.). Free Press. https://teddykw2.wordpress.com/wp-content/uploads/2012/07/everett-m-rogers-diffusion-of-innovations.pdf

Sadiq, R. B., Safie, N., Abd Rahman, A. H., & Goudarzi, S. (2021). Artificial intelligence maturity model: A systematic literature review. *PeerJ Computer Science, 7*, Article e661. https://doi.org/10.7717/peerj-cs.661

Segooa, M. A., Modiba, F. S., & Motjolopane, I. (2025). Generative artificial intelligence tools to augment teaching scientific research in postgraduate studies. *South African Journal of Higher Education, 39*(1), 294–314. https://dx.doi.org/10.20853/39-1-6275

Weinberg, A. I. (2025). *A framework for the adoption and integration of generative AI in midsize organizations and enterprises (FAIGMOE)*. ArXiv. https://doi.org/10.48550/arXiv.2510.19997

Xulu, H. H., Hlongwa, N. S., & Maguraushe, K. (2024, December). Unlocking the potential of AI in higher education: A multi-dimensional study of ChatGPT adoption at a South African university. *Proceedings of the Focus Conference (TFC 2024),* 516–532. https://doi.org/10.2991/978-94-6463-630-7_28

Information about the authors:

Modiba Florah Sewela – https://orcid.org/0000-0001-6905-067X; D LITT ET PHIL in Development Studies, Senior Lecturer, Department of Development Studies, Nelson Mandela University, Gqeberha, South Africa.

Segooa Mmatshuene Anna – https://orcid.org/0000-0002-4190-8256; Doctor of Computing, Senior Lecturer, Department of Informatics, Tshwane University of Technology, Pretoria, South Africa.

Motjolopane Ignitia – https://orcid.org/0000-0001-9047-6720; PhD in Information Systems, Professor, Nelson Mandela University, Business School, Gqeberha, South Africa.

CONTRIBUTORS

BADUZA Gugulethu – https://orcid.org/0000-0003-4092-6521; PhD, Dr, Senior Lecturer, Rhodes University, Makhanda, South Africa.

BISHA Zamagoba – https://orcid.org/0000-0001-7918-8197; BA Honours, MA Development Studies Student, Nelson Mandela University, Gqeberha, South Africa.

BVUMA Stella – https://orcid.org/0000-0001-8351-5269; PhD in Information Technology Management; Professor, Director, University of Johannesburg, Johannesburg, South Africa.

KGOPA Alfred Thaga – https://orcid.org/0000-0001-5455-7064; PhD (Informatics), Dr, Senior Lecturer, University of South Africa, Roodepoort, South Africa.

MAKELANA Phenuel – https://orcid.org/0000-0003-0986-1117; Doctor of Computing, Lecturer, Department of Computer Sciences, Vaal University of Technology, Vanderbiljpark, South Africa.

MELNYK Yuriy Borysovych – https://orcid.org/0000-0002-8527-4638; Doctor of Philosophy in Pedagogy, Affiliated Associate Professor; Chairman of Board, Kharkiv Regional Public Organization "Culture of Health" (KRPOCH); Director, Scientific Research Institute KRPOCH, Ukraine.

MODIBA Florah Sewela – https://orcid.org/0000-0001-6905-067X; Doctor of Literature and Philosophy in Development Studies, Senior Lecturer, Department of Development Studies, Nelson Mandela University, Gqeberha, South Africa.

MOGASE Refilwe Constance – https://orcid.org/0000-0001-7337-8547; Doctor of Computing, Dr, Senior Lecturer and Head of Department, Tshwane University of Technology, Pretoria, South Africa.

MOGOALE Phumzile Mseteka – https://orcid.org/0000-0003-1770-5739; PhD, Dr, Postdoctoral Research Fellow, Tshwane University of Technology, Pretoria, South Africa.

MOTJOLOPANE Ignitia – https://orcid.org/0000-0001-9047-6720; PhD in Information Systems, Professor, Nelson Mandela University, Business School, Gqeberha, South Africa.

MSWELI Nkosikhona Theoren – https://orcid.org/0000-0003-4709-0763; PhD (Information Systems), Dr, Senior Lecturer, University of South Africa, Roodepoort, South Africa.

MYKHAYLYSHYN Ulyana Bohdanivna – https://orcid.org/0000-0002-0225-8115; Doctor of Psychological Sciences, Full Professor; Head of the Department of Psychology, Uzhhorod National University, Uzhhorod, Ukraine.

PENXA Lungile – https://orcid.org/0009-0006-2576-5474; PhD, Dr, Lecturer, University of the Western Cape, Cape Town, South Africa.

PRETORIUS Agnieta Beatrijs – https://orcid.org/0000-0002-6510-2468; Doctor of Technologiae, Dr, Senior Lecturer and Assistant Dean, Tshwane University of Technology, Pretoria, South Africa.

PYPENKO Iryna Sergiivna – https://orcid.org/0000-0001-5083-540X; Doctor of Philosophy in Economics, Affiliated Associate Professor, Secretary of Board, Kharkiv Regional Public Organization "Culture of Health"; Scientific Research Institute KRPOCH, Ukraine.

RAMAFI Pelonomi – https://orcid.org/0000-0003-2477-7060; Mcom, Lecturer, University of Witwatersrand, Johannesburg, South Africa.

SATHEKGE Machiniba Sylvia – https://orcid.org/0009-0001-9410-3267; Doctor of Business Administration, Doctor, Professor of Practice, University of Johannesburg, Johannesburg, South Africa.

SEGOOA Mmatshuene Anna – https://orcid.org/0000-0002-4190-8256; Doctor of Computing, Senior Lecturer, Department of Informatics, Tshwane University of Technology, Pretoria, South Africa.

STADNIK Anatoliy Volodymyrovych – https://orcid.org/0000-0002-1472-4224; Doctor of Philosophy in Medicine, MD, Affiliated Associate Professor, Kharkiv Regional Public Organization "Culture of Health", Kharkiv, Ukraine; Uzhhorod National University, Uzhhorod, Ukraine.

About the Editors

Yuriy Borysovych MELNYK is a scientist, a lecturer, a psychologist and a civic leader.

He holds a PhD in Pedagogy, MPSI, MIM, MPES and is an Affiliated Associate Professor.

He worked in general education institutions, vocational and technical institutions, higher education institutions and public organisations for over twenty-five years as a teacher, psychologist, Head of the Psychology and Healthy Lifestyle Department, Associate Professor in the Psychology Department and Professor in the Psychology and Pedagogy Department.

He developed and implemented training programmes for schoolchildren and undergraduate and postgraduate students. He also developed pedagogical and psychological research methodologies and technologies for various levels of implementation, including organisational and managerial, socio-pedagogical and educational levels. He has many years of experience in organising international scientific conferences, training courses, and masterclasses. He has also organised international competitions for young scientists.

Author of over 200 scientific works, including dissertations, monographs, articles in international peer-reviewed journals, conference proceedings, congress and symposium collections, as well as textbooks and programmes. These publications cover a range of topics, including artificial intelligence, blockchain technology, virtual assets, educational management, pedagogical logistics, psychology, psychotechnology and health culture. He introduced the concept of the "Human-AI System" into scientific discourse and developed the AIC "AI Chatbots Attribution".

He currently holds the positions of Professor of the Laboratory of Psychological Research and Director of the Scientific Research Institute. He is also Chairman of the Board of Public Organisations. He participates in research aimed at studying the legal and ethical aspects of the use of artificial intelligence, as well as the practical possibilities of implementing AI in the field of education.

Mmatshuene Anna SEGOOA holds a Doctor of Computing (Informatics) and is a recognised award-winning academic and researcher.

She is a Senior Lecturer in the Department of Informatics, Faculty of ICT, Tshwane University of Technology (South Africa), leading international collaborations and partnerships with experience in securing funding for research and mobility projects.

She has expertise in teaching undergraduate and postgraduate diploma courses, developing short learning programmes, reviewing qualifications, and supervising master's and doctoral students to graduation.

She has authored and co-authored 30 publications in Information Systems, focusing on Artificial Intelligence, Big Data Analytics, Information System Security, Cloud Computing, Design Science Research, and Systematic Literature Reviews, particularly in ICT4D.

Dr. Segooa has served as organiser and session chair of international conferences, reviewer for local and international journals, Master of Ceremony, moderator, faculty Research Day adjudicator, keynote speaker, external examiner, and graduation name reader. She continues to mentor emerging researchers, foster global partnerships, and advance research impact locally and internationally.

SCIENTIFIC EDITION (Monograph)

Artificial Intelligence in Digital Society
Volume 1

Collective Monograph

Editors:
Yuriy Borysovych MELNYK
Mmatshuene Anna SEGOOA

ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)
DOI: https://doi.org/10.26697/9786177089192.2026

Managing editor, proofreading: Melnyk Y. B.
Computer page positioning and layout: Pypenko I. S.
Administrator of site: Stadnik A. V.
Designer: Sviachena Ya. Yu.

Format 60x84/16
Print on coated paper. Full colour digital printing.
Conv. printing sheet 11.16. Order № 1-35
100 copies.

CONTACT INFORMATION
KRPOCH Scientific Research Institute
https://institute.culturehealth.org

Publisher KRPOCH
(Kharkiv Regional Public Organization "Culture of Health")
Shchedryka lane, 6, of. 6, Kharkiv, Ukraine, 61105
Emails: CultureHealth@ukr.net; KhRPOCH@gmail.com
URL: https://doi.org/10.26697/publisher; T/F: +38 057 775 75 23
Certificate to registration ДК № 4387, 10.08.2012

Artificial Intelligence in Digital Society

Collective Monograph

The series of monographs "Artificial Intelligence in Digital Society" aims to bring together the existing theories and practices of artificial intelligence in the Western and Eastern worlds. The idea of synergy plays a central role in the monographs. The monographs in this series represent an attempt to integrate the experience accumulated in recent years and propose new solutions to the problems of using artificial intelligence. This series of monographs is an important resource for researchers, professors and graduate students, as well as practitioners and specialists in the field of artificial intelligence.

ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)
DOI: 10.26697/9786177089192.2026
BISAC: TEC052000