A Systematic Analysis of Ethical and Governance Concerns Relating to Artificial Intelligence Adoption in the South African Public Sector

Wiston Mbhazima Baloyi*

Department of CoGHSTA, Limpopo Province, South Africa

***Corresponding author**

DOI: https://dx.doi.org/10.47772/IJRISS.2025.909000269

Received: 04 September 2025; Accepted: 11 September 2025; Published: 08 October 2025

ABSTRACT

Globally, artificial intelligence (AI)-related technologies are rapidly reforming public services in a modern way. Contemporary public administrations are experiencing increased operational efficiency, streamlined work processes and greater productivity through AI-driven technologies. However, emerging economies such as South Africa still face AI ethical and governance concerns that hamper their full potential. Therefore, this study proposes an integrated conceptual framework for ethical and governance concerns of AI in the South African public sector, which was developed based on the findings. Underpinned by the constructivist worldview, this study employs a qualitative approach using the rapid systematic analysis of literature to investigate ethical and governance concerns of AI in the South African public sector. The rapid systematic literature analysis focused on a comprehensive search of sources from the Web of Science, Science Direct and ProQuest databases for the period ranging from 2023 to 2025. A total number of 34 peer-reviewed journal articles, conference papers and academic book chapters were included in the systematic literature analysis. The key findings reveal that ethical and governance concerns, such as data privacy, security, bias, public trust, fairness, equity, accountability, transparency and regulatory framework, are fundamental in designing and implementing AI-enabled technologies in the public sector. Further, the study's findings suggest an integrated conceptual model that can be considered a valuable technique in adopting AI in the public sector. The model intends to provide a structural approach to improve AI adoption not only in the South African public sector. Decision-makers, policymakers, practitioners, and strategists of the public sector can find the implications of this study significant in AI adoption while taking into cognizance ethical and governance concerns.

Keywords: Artificial intelligence, AI ethical concerns, AI governance, Public administration, South Africa

INTRODUCTION AND BACKGROUND

The rapid growth of information and communication technology (ICT) impacts how people live and interact daily (Baloyi & Beyers, 2019). Many public sectors around the world are now reshaping their systems, practices and processes to anchor public administration and enhance people's lives through the assimilation of AI-driven technologies (e.g., machine learning, deep learning and natural language processing) in the digital realm (Rekunenko et al., 2025). Although these public sectors are besieged by the repercussions of ethical and governance (policy) issues brought about by AI-related technologies, the most impacted milieus are emerging economies like South Africa (Chilunjika, 2024). Most notably, while most advanced economies (e.g., the United States of America, Australia and the United Kingdom) are experiencing the rewards of AI technologies, such as economic growth (Attard-Frost et al., 2024; van Noordt et al., 2025), the countries in the African continent, including South Africa, are still lagging primarily due to ethical and governance challenges exacerbating the adoption of AI and its implementation progress, which are accompanied by the paucity of research on AI (Mtuze & Moraga, 2024; Diallo et al., 2025). These include, but are not limited to, the most prevalent ethical and governance concerns, such as data privacy, security, fairness, accountability, transparency, and bias (Barodi & Lalaoui, 2025).

AI governance and ethics remain a global challenge, interjecting into the efficient application of these cutting-edge digital technologies (Machado et al., 2025; Zidouemba, 2025). Most imperatively, AI prototypes are more obfuscated, primarily leading to opaque black boxes that diminish public servants' capability to elucidate and rationalize the algorithms (Alaran et al., 2025). Yet governing AI presents sophisticated complexities, necessitating a well-proportioned methodology entailing all-encompassing regional (e.g., Africa) and worldwide regulatory frameworks (Attard-Frost et al., 2024). Additionally, while regional stance is susceptible to the establishment and execution of tailored regulations that responds to the ever-changing business environmental factors (such as political, economic, social, technology, and ecology) in the digital epoch (Henk & Henk, 2025), the global governance perspective, on the other hand, expedites the international alliance with various countries, to ensure the standardization of the practices and thus, normalizing the threats and strengthening the integrated governance structure intending to handle universal implications and accelerate the development and execution of the AI projects (Marwala, 2024).

South Africa, like diverse countries in Africa, has made enormous strides to revolutionize and modernize (digital transformation) the public administration (i.e., at the national, provincial and local governments) with the view to respond to the proliferating demands for digitalized service delivery to citizens and disparate stakeholders using different emerging digital technologies such as AI, cloud computing, the Internet of Things (IoT), and big data (Shekgola & Modiba, 2025). Although South Africa is regarded as the most innovative country in the African continent concerning the development and implementation of AI-enabled technologies and other emerging digital technologies, particularly in the transportation and logistics sector, leading to remarkable distribution of products and ensuring efficient value chain activities (Nyugha, 2024), the assimilation of AI-driven algorithms in the public sector presents distinctive obstacles that impede its progress and realization of strategic objectives. For instance, as cited by Mtuze and Morige (2024), whereas the country has made endeavors to develop legislation (e.g., The Electronic Communications and Transactions Act – ECTA) to tackle the adoption of various technologies, these laws downplay and overstretch the emphasis of AI; therefore, they widen its ethical and governance issues. Thus, due to the lack of a regulatory framework governing the implementation of AI in the South African public sector, it is considered to be at a preliminary stage (Chitimira & Munedzi, 2024; Giwa & Ngepah, 2024; Mokwele, 2024). Against this background, this study examines AI ethical and governance concerns in the South African public sector, aiming to contribute towards the attainment of the National Development Plan's Sustainable Development Goals (SDGs) and Vision 2030. The research question is thus: What are the ethical and governance concerns impacting the adoption of AI in the South African public sector?

LITERATURE REVIEW

This section critically reviews literature relating to AI in the public sector, South African studies on AI, and an overview of ethical and governance concerns regarding artificial intelligence in the public sector. It commences with the discussion of the theoretical framework underpinning this study, such as the technology-organization-environment (TOE) framework.

Theoretical grounding

This study adopts Tornatzky and Fleischer's (1990) technology-organization-environment (TOE) framework to delve into ethical and governance concerns of AI in the South African public sector. The feasibility of embracing the TOE framework is based on the newness of AI in the South African public sector. As an extensively applied theory in social sciences, the TOE framework assists in appraising the state of public sector organizations by using three contextual determinants: technology, organization, and environment, when aspiring to adopt digital technologies (Baker, 2012). Despite this, the TOE framework has been cherished for its potential and capacity to assess the organizational conditions while embracing new technologies employing these contextual determinants (Bryan & Zuva, 2021). The three are, thus, clarified. The technological factor involves the consideration of digital technologies (e.g., AI) imposed by the ever-evolving business environment that are appropriate for espousal in the public sector (Bryan & Zuva, 2021). The organizational factor entails scanning internal capabilities (e.g., expertise, innovative culture, leadership drive, infrastructure, and rewards strategies) that are vital for adopting digital technologies in the digital domain. Further, it also takes into account the extent of the organization in terms of scope and magnitude (Baker, 2012). The

environmental factor involves scrutinizing the external circumstances to identify probable prospects and threats that can enable or constrain technological modernization (Bryan & Zuva, 2021). The environmental determinant revolves around compliance with applicable laws and regulations indispensable to directing digital technologies.

Artificial intelligence in the public sector

AI can be defined as "*systems that display intelligent behaviour by analyzing their environment and taking actions – with some degree of autonomy – to achieve specific goals*" (Tomazevic et al., 2024:113). AI-enabled technologies have the capacity to process vast amounts of computer data while detecting patterns to forecast results for related occasions (Jankovski et al., 2025). AI in the public sector is a multidimensional concept that includes systems such as face recognition, machine learning, natural language processing, virtual assistants, and chatbots, *inter alia*. The diffusion of AI into public sectors' internal systems has not only increased operational efficiency and enhanced automated decision-making processes in the digital economy but also streamlined problem-solving to render seamless public services, leading to the attainment of long-term goals (Henk & Henk, 2025; Zidouemba, 2025). Although advanced economies increasingly utilize AI technologies to render efficient and cost-effective public services, emerging economies are plagued with impediments such as digital infrastructural deficit, greater digital dividends and unreliable internet connectivity, hampering their full potential (Jankovski et al., 2025). More to this, the consideration of AI-enabled technologies in public sector dealings (rural development initiatives) has been keen on improving the country's economic growth (Naidoo, 2024). This is further affirmed by Vatamanu and Tofan (2025), who state that AI technologies are said to have a positive and significant effect on the country's economic growth.

Aside from the potential rewards brought about by AI technologies emphasized earlier, these disruptive digital technologies have been associated with potential challenges such as job displacement and skills gaps (Naidoo, 2024). This is consistent with the view of Anshari et al. (2025), who confirm that the assimilation of AI-related technologies into the internal processes and systems of the public sector can nurture prejudice in the decision-making procedures and strengthen possible job displacement of the public servants. On the other hand, the adoption of AI by the public sector has been susceptible to widening public servants' deskilling or contributing to skills shortages (Henk & Henk, 2025). It is corroborated that the provision of thorough training and development sessions (e.g., formal or informal capacitation) on AI to public servants can mitigate the risks of losing jobs and retrenchments, thereby circumventing job displacement (Giwa & Ngepah, 2024; Machado et al., 2025).

South African studies on artificial intelligence

Despite being in the early phase in the South African public sector, research on AI technologies is gradually gaining momentum in academic discourses. Several scholarly debates have devoted their research to AI in the South African public sector (Saal et al., 2025; Shekgola & Modiba, 2025). Briefly, scholars have focused their research on areas such as healthcare services (Nene & Hewitt, 2023; Janneker, 2025; Malope, 2025), public administration (Naidoo, 2024; Modiba, 2025) and municipal administration (Bester, 2024). The table below underscores the South African studies on AI.

Table 1. Research on Artificial Intelligence in South Africa

Author	Objective	Methodology	Findings
Nene and Hewitt (2023)	To develop a conceptual model intended to expedite the execution of AI in South African public hospitals.	A qualitative approach, phenomenological design.	The study discovered three major themes: skills pertaining to AI, management, and executive drive in expediting the execution of AI in the South African public hospitals.
Malope (2025)	To probe the effect of AI in enhancing	A quantitative approach,	The findings reveal that applying AI-based technologies in primary healthcare services

	e-health in South Africa.	systematic analysis of literature	can improve the administration of sicknesses like diabetes and other chronic diseases.
Modiba (2025)	To investigate the uptake of AI in improving record management practices.	A mixed-methods approach, in-depth interviews and surveys.	The AI-based technologies can drastically implement record management practices in the Gauteng Department of Education in South Africa.
Saal et al. (2025)	To examine the assimilation of AI at the Department of Basic Education in South Africa.	A qualitative, scoping review	The main findings indicate that AI has the possibility to ensure a customized learning strategy, improve inclusivity and promote education for pupils with distinct needs.
Shekgola and Modiba (2025)	To delve into the application of AI chatbots to institute a sophisticated digital government.	A qualitative approach, systematic literature review using content analysis.	The study finds that an AI chatbot can enable citizen accessibility to different services irrespective of their geographical area.

Overview of ethical and governance concerns of artificial intelligence in the public sector

While ethical concerns involve the lack of moral principles that obstruct the effective use of AI (Chilunjika, 2024), governance concerns revolve around the paucity of regulatory policies and the absence of transparent and accountable application of AI (Zidouemba, 2025). Despite the burgeoning recognition of digitalized service delivery by diverse governments across the globe, the constant evolution of emerging and disruptive digital technologies such as AI poses ethical and governance complications that are eroding not only the administrative tasks and internal operations of those governments but also public trust (Jankovski et al., 2025; Robles & Mallinson, 2025). Scholars have delved into ethical issues (such as data security, privacy and trust) and governance issues (transparency and accountability), among others, that have been identified as catastrophes interrupting the accomplishment of fully-fledged functioning of AI-based technologies, especially in emerging economies (Tomazevic et al., 2024; Barodi & Lalaoui, 2025; Rekunenko et al., 2025; Vatamanu & Tofan, 2025). While the integration of sound ethical dimensions (e.g., data security, privacy and trust) into AI models within the public sector is critical to promote the legitimate application of AI-enabled technologies, it is contended that citizens' well-being and social aspects can be improved immensely, leading to increased public trust (Erkkilä, 2023).

On the other hand, governance concerns pertaining to AI are increasingly devastating most public sectors throughout the world (Ulnicane & Aden, 2023). For example, notwithstanding the great strides made with the acceptance of emerging digital technologies (e.g., AI, the IoT, cloud computing and big data), research reveals that the public sectors in most countries in the African continent and other parts of the world are still finding ways on how to adopt AI-related technologies (de Almeida & dos Santos Júnior, 2025; de Fine Licht, 2025; Diallo et al., 2025). With that being said, the governance of AI technologies in terms of the development and implementation of policy and regulatory frameworks remains a hurdle and cannot be overemphasized. For instance, it is argued that to thrive in the digital epoch; the public sector ought to have effective AI governance structures and binding regulations that intend to improve governance operations and systems (Ulnicane & Aden, 2023). Similarly, the public sector's capacity to devise proper AI governance models seeking to address the governance issues, such as transparency, accountability and other ethical-related conundrums (e.g., data privacy concerns) is imperative to heighten internal operations and improve overall performance of those organisations (Vatamanu & Tofan, 2025).

RESEARCH DESIGN AND METHODOLOGY

Underpinned by the constructivist worldview, this study employs a qualitative approach using the rapid systematic literature analysis to gain insight into AI's ethical and governance concerns in the South African

public sector. This study is descriptive and theoretical, considering existing secondary data available online. According to Cochran (2020:1), a rapid systematic review is "*a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting specific methods to produce evidence for stakeholders in a resource-efficient manner*". This technique enables the researcher to source information that is readily available electronically rather than gather data directly from the participant. The rationale of espousing the rapid systematic review is based on the following advantages: (1) it helps save time, resources and costs related to the collection of primary data, (2) it shortens time associated with identifying and screening of data to ensure eligibility requirements, and (3) the method maximizes rigor of the data gathered (Garritty et al., 2024).

The data sources that were considered in the study include peer-reviewed journal articles, conference papers, and grey literature (i.e., government publications and reports). This study applied the rapid review as a technique and procedure to analyze the readily available information. The academic sources were accessed through the often utilized databases by scholars among different fields of study, such as the Web of Science, Science Direct and ProQuest search engines using keywords such as "*AI ethics*", "*digital ethics*", "*AI governance*", "*AI policy*", "*AI framework*", "*AI ethical concerns*", "*public sector*", "*public service*", "*service delivery*", "*public administration*", and "*South Africa*". More importantly, the preliminary pertinence was ascertained through the study's title. The study considered recent publications with dates ranging between 2023 and 2025 to ensure the currency of the sources included.

This study integrates Cochran's (2020) rapid systematic review and Xiao and Watson (2019) three phases of rigorous systematic literature analysis to examine AI's ethical and governance concerns in the South African public sector: planning, conducting, and reporting the review. The following research process of conducting a literature review was followed by a rapid systematic review of collecting and analyzing data, as illustrated in Figure 1.

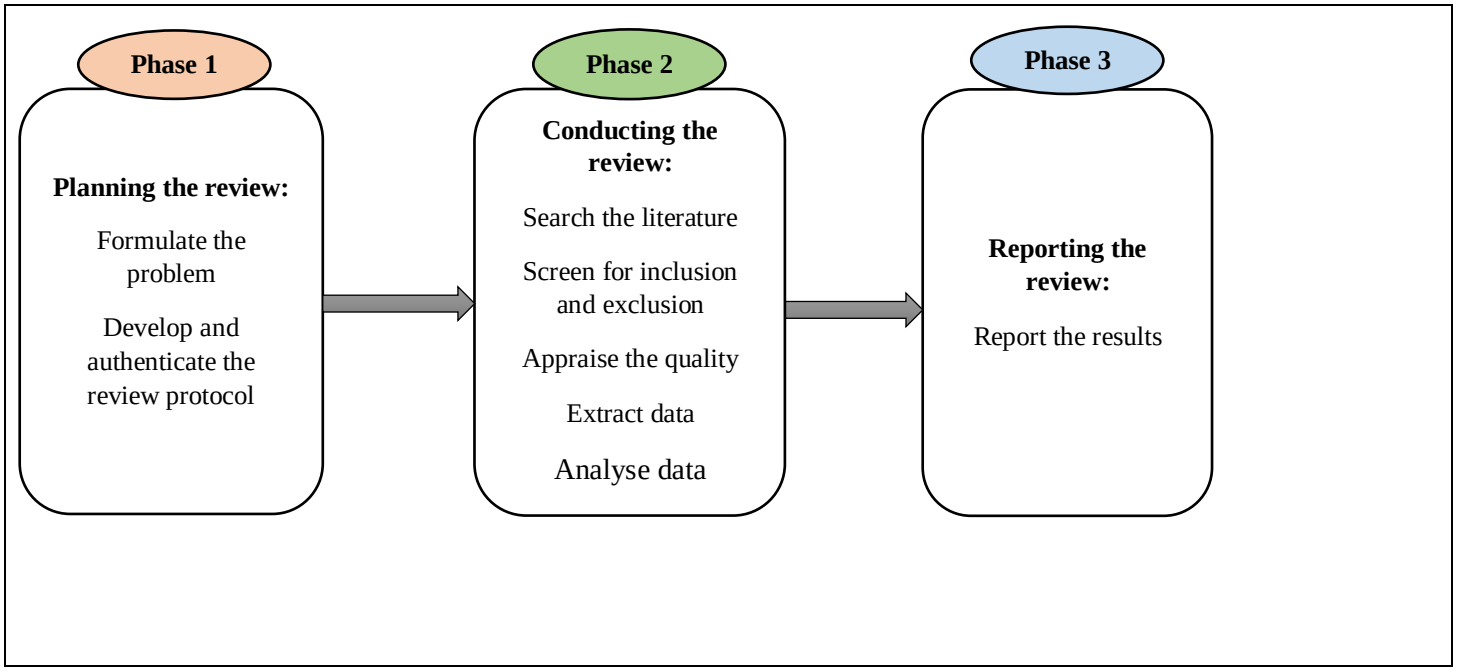

Figure 1: Process of conducting a literature review. Adapted from Xiao and Watson (2019)

Stage one: Planning the review.

The first stage in the process of conducting a rapid systematic review was the formulation of the research problem and the establishment and authentication of the review protocol. This has led to the formulation of the research question of this study, which is stated: "*What are the ethical and governance concerns impacting the application of AI in the South African public sector*?". The review protocol was also developed and authenticated, including "the purpose of the study, research questions, inclusion criteria, search strategies,

quality assessment criteria and screening procedures, strategies for data extraction, synthesis, and reporting" (Xiao & Watson, 2019:103).

Stage two: Conducting the review.

The second stage of a rapid review was conducting the literature review on the ethical and governance concerns impacting the application of AI.

- Literature search: A comprehensive search of relevant literature from the most commonly used databases, such as Web of Science, Science Direct and ProQuest, was conducted. The search strategy was limited to the keywords, such as "*AI ethics*", "*digital ethics*", "*AI governance*", "*AI policy*", "*AI framework*", "*AI ethical concerns*", "*public sector*", "*service delivery",* "*public service*", "*public administration*", and "*South Africa*".
- Screen for inclusion and exclusion criteria: For inclusion and exclusion criteria, the researchers read the abstracts of selected peer-reviewed journal articles and conference papers to determine their relevance. Further, only studies written in English were included to enhance reliability and comprehensibility while conducting a review. Only sources published between 2023 and 2025 were considered to capture current and most topical matters regarding ethical and governance issues in the public sector. On the other hand, the exclusion criteria were employed to eliminate unrelated and repetitive studies. Additionally, electronic grey literature, particularly technical reports, was excluded due to the absence of a peer-review procedure. Studies published before 2023 and non-English sources were also excluded. To improve the effectiveness and precision during the screening procedure, an EndNote Library Version 21.5 was utilized to arrange the references and eliminate duplicates after screening.
- Appraise the quality: The quality of the studies was appraised by reading and rereading the full-text sources (peer-reviewed journal articles, academic books and conference papers). Apart from that, the quality was established by contemplating peer-reviewed journal articles, conference papers, and academic book publications by the utmost and renowned publishers, leading to attributable research. The data quality was enhanced by assigning two independent reviewers who autonomously evaluated the extracted sources to determine their relevance to the ethical and governance issues impacting the application of AI in the public sector.
- Data extraction: The data extraction and analysis process was organized by reflecting on the study's background and problem statement, the literature review process and research methods applied in the studies. At the outset, researchers autonomously extracted data to perform cross-checking. A spreadsheet was utilized to extract data, including important information such as author, year of publication, study title, purpose, research approach, and main findings.
- Data analysis: A thematic analysis was employed to enhance rigor in interpreting and analyzing data collected from the rapid systematic review. As one of the most extensively used techniques in research, a thematic analysis refers to a qualitative method for detecting relevant patterns, meanings and themes of textual data using a systematic review (Ahmed et al, 2025). The thematic analysis is significant in improving the rigor and reliability of gathered data, thereby diminishing prejudice in data analysis and synthesis. The thematic classification has expedited the structured approach to ethical and governance issues impacting the application of AI in the public sector.

The following eligibility criteria were followed when selecting the sources, as indicated below.

Table 2: Eligibility criteria

Item	**Description**
Sector	Only sources that emphasize the public sector were included in the study.
Sources	This rapid systematic literature analysis embraced peer-reviewed journal articles, conference papers and academic book chapters.
Key concepts	The search for keywords was restricted to "*AI ethics*", "*digital ethics*", "*AI governance*", "*AI policy*", "*AI framework*", "*AI*

	ethical concerns", "*public sector*", "*service delivery",* "*public service*", "*public administration*", and "*South Africa*".
Time frame	The rapid systematic literature analysis contemplated studies available between 2023 and 2025 to enhance the currency of sources.
Research methodology	Both qualitative and quantitative research were appropriate for inclusion in the rapid systematic literature analysis.
Language	Only studies written in English were eligible for inclusion. Studies written in other languages were omitted based on the complexity of accessing translators.

Stage three: Reporting the review.

The final stage of conducting a rapid systematic review was elucidating and reporting the study's findings. The findings were based on the supporting views of the scholars and themes and concepts extracted from the literature.

RESULTS OF THE SYSTEMATIC ANALYSIS

The literature search for the rapid systematic review is underscored in Table 2. Primarily, 2,154 sources were identified through Web of Science, Science Direct and ProQuest search engines. After eliminating repeating and non-English sources, 1,235 distinctive sources remained. These sources were scrutinized based on their titles and abstracts, leading to 1,195 being excluded due to their irrelevance, including the basis that sources were not specific about ethical and governance issues relating to AI in the public sector. Only 34 sources were considered for inclusion in the final systematic analysis. These include 32 peer-reviewed journal articles, two conference papers and one book chapter. This technique enabled the researcher to stay focused and ensure that pertinent sources were selected to respond to the research enquiry.

The entire process (i.e., the inclusion and exclusion criteria) of the rapid systematic literature analysis has unfolded in the manner accentuated in Table 2. The studies conducted in both emerging and advanced economies were included.

Table 2. Eligibility criteria for rapid systematic review

Database	**Key words**	**Total number**	**Relevant sources**	**Kind of sources**	**Total sources used**
Web of Science	AI ethics, digital ethics, AI governance, AI policy, AI ethical issues, public sector, public service, public administration, and South Africa	481	10	Peer-reviewed journal articles	10
Science Direct	AI ethics, digital ethics, AI governance, AI policy, AI ethical issues, public sector, public service, public administration, and South Africa	755	13	Peer-reviewed journal articles, book chapters and conference papers	13
ProQuest	AI ethics, digital ethics, AI	918	11	Peer-	11

	governance, AI policy, AI ethical issues, public sector, public service, public administration, and South Africa			reviewed journal articles	
Total		**2,154**	**34**		**34**

DISCUSSION OF THE FINDINGS

The discussion infers the findings and delineates the main implications subsequent to the development of an integrated conceptual model for ethical and governance issues of AI in the South African public sector.

Ethical issues relating to artificial intelligence in the public sector

While AI technologies can present possible rewards (e.g., intensified efficiency and cost-effectiveness), they are also susceptible to potential hurdles such as ethical issues. This section elucidates those ethical issues concerning AI in the public sector as identified in the literature.

Data security and privacy

At the outset, while the decisions are being taken to incorporate AI into the internal value chain activities and systems to improve the administrative tasks of the public sector, it is critical to devise suitable business strategies and frameworks intended to manage vast amounts of data (Henk & Henk, 2025). Globally, even though the proliferating speed of AI in the public sector raises ethical challenges for its potential users and multitude of stakeholders, especially in emerging economies like South Africa (Bester, 2024; Malope, 2025), the need for the development of an integrated conceptual model cannot be overstated. Broadly, while the use of AI is considerably thriving, data security and privacy become an epicenter of its application and, as such, play an integral role in the uptake of AI in the public sector in the digital landscape (Shekgola & Modiba, 2025). In South Africa, for example, since AI in the public sector is in its nascent stage, the development and implementation of binding and robust AI policy and regulations for data security and privacy is fundamental to increase public trust in the usage of AI-powered technologies, including the healthcare services (Janneker, 2025). To this end, however, the absence of obligatory regulations governing AI systems and algorithms leads to non-compliance with data security and privacy of demographic (confidential) information and probable exploitation of data, thereby increasing cyberattacks and the potential data breaches (Agbarakwe & Chibueze, 2024).

Bias and public trust

According to Ulnicane & Aden (2023:669), bias refers to "*computer systems that systematically and unfairly discriminate against certain individuals or groups of individuals in favor of others*". One of the factors disrupting the effective functioning of AI in the public sector is the significant bias of algorithms, which present problems and uncertainties for possible obfuscations that erode public trust in citizens, innovativeness and the efficiency of service delivery to the citizens (Vatamanu & Tofan, 2025). Algorithmic bias has a detrimental effect and adverse consequences in the provision of quality services to the citizens, in that it obscures the algorithms that can also widen social disparities. Additionally, although many public sectors increasingly embed AI-enabled technologies into internal processes and systems, possible biases are regarded as a persistent barrier that discriminates against individuals from designated or marginalized groups (Ulnicane & Aden, 2023). To this end, nonetheless, the biases frequently originate from the training data that can be manipulative, leading to inequity and, on top of that, do not encompass the demographics representative of the entire population (e.g., cultural values, race, sex and customs) (Agbarakwe & Chibueze, 2024). For instance, possible biases in the public sector's healthcare services can occur when AI systems prejudice the results of the disease diagnosis after a patient's examination. Similarly, algorithmic bias not only disrupts the successful implementation of AI technologies but is also associated with ethical conundrums in their application, such as data privacy and security challenges, which impair public trust (Henk & Henk, 2025). Like several other

emerging economies globally, it is argued that South Africa faces AI bias as a significant obstacle, which emanates from the lack of robust data governance, particularly in primary healthcare services (Janneker, 2025).

Fairness and equity

AI-driven technologies (e.g., facial recognition) are prone to being fair and non-discriminatory in their application while nurturing equal access to technologies. Nevertheless, the opposite may be true in some emerging economies such as South Africa. For instance, inequality in South Africa is one of the socio-economic inconsistencies (unemployment and poverty form part of) that continue to happen, aggravating the digital dividends to society (Malope, 2025). To this end, the condition in which South Africa is positioned in terms of inequality can be regarded as a chronic disease that seems incurable and, as such, broadens the inequitable distribution of economic benefits and welfare. Further to that, even though strides are being made to embrace innovative methods of service delivery (e.g., basic education) through digital technologies (e.g., AI) in South Africa, most remote areas and designated groups are still faced with barriers associated with ICT, such as unreliable network connectivity, absence of digital competencies and limited access to the internet (Saal et al., 2025). Aside from that, it is argued that strategies should be in place to tackle the persisting bias concerns relating to AI by the public sectors in an endeavor to resolve fairness and equity challenges faced by the citizens, thereby shaping service delivery (Shekgola & Modiba, 2025). In the education milieu, for example, Agbarakwe and Chibueze (2024) corroborate that AI systems and algorithms should be prudently developed to circumvent any biases that may occur, leading to fairness and equal treatment of students.

Governance issues pertaining to artificial intelligence in the public sector

Based on the systematic literature analysis, this section delineates governance issues relating to AI in the public sector.

Accountability

Accountability is a fundamental concern in the uptake of AI in the public sector. As such, the ethical usage of AI-powered technologies necessitates the design and execution of strong legislative and regulatory models to enhance accountability (Barodi & Lalaoui, 2025). Despite AI being able to rationalize public services to numerous stakeholders (including the citizens), it is also predisposed to policy issues that are further interconnected to accountability (Erkkilä, 2023). Further, in most governments, public trust concerns in AI algorithms are interrupted by the non-responsiveness of governance programs, which creates accountability gaps for citizens (Attard-Frost et al., 2024). More significantly, governments are liable for prioritizing a well-defined and legitimate line of responsibilities to strengthen accountability in case of prejudice or unintentional repercussions (Shekgola & Modiba, 2025). Moreover, as cited by Chilunjika (2024), accountability of the AI application in the public sector can only be enhanced through automated methods by proposing clear-cut strategies and policy documents relevant to boost decision-making processes. Having said that, considering public engagement in AI policy development initiatives can profoundly stimulate accountability (Bester, 2024). Chitimira and Munedzi (2024:88) iterate that “there are no defined responsibility and accountability mechanisms for the harm caused by AI in South Africa”. Overall, an explicit structural mechanism (subject to appraisal by the oversight agencies) regarding accountability should be in place to nurture dynamic alterations resulting from AI.

Transparency

Research suggests the importance of transparency in diminishing the obscurity and the dark side of AI algorithms (Machado et al., 2025; Alaran et al., 2025; Rekunenko et al., 2025). According to Agbarakwe and Chibueze (2024), AI design transparency is imperative in enabling the organization to enhance algorithms, bolster accountability for internal processes (e.g., students’ educational assessments) and render quality services to diverse stakeholders. Similarly, it is contended that the government’s capability to embrace data-driven algorithms through AI technologies is paramount to improve the transparency of public services, while at the same time streamlining the public sector’s administrative processes to minimize bias and increase public trust (Henk & Henk, 2025). Additionally, while public involvement (through debates and conferences)

regarding AI policymaking issues is critical to strengthen collaboration with government, it demonstrates the significance of greater transparency and value creation to the citizens (Marwala, 2024). He further recapitulates that AI determinations involve transparent and accountable usage, which is key to promoting public trust and sound public administration. Mtuze and Morige (2024) stress that even though particular transparency criteria are devised for regulating AI algorithm threats in the digital arena, non-compliance remains a cardinal matter, presenting risks associated with infringing human rights. Accordingly, it is worth noting that transparency, as a primary principle of governance in a technological atmosphere, plays a central role in reshaping business models and the acceptance of AI-enabled technologies by the public sector, while curtailing the bias of the algorithms (Rekunenko et al., 2025).

Regulatory framework

Governance for AI technologies has always been an obstacle for most governments in emerging economies, not only for AI start-ups but also for organisations that have already been in the AI environment for some time (Mokwele, 2024). That being the case, the absence of sound and robust legal and regulatory frameworks governing ethical concerns (e.g., data privacy, security and public trust) impedes the efficient adoption of AI in the public sector. This is supported by Zidouemba (2025), who opines that the unregulated usage of AI-related challenges creates room for unequal access, thereby exacerbating the *status quo*. Generally, although the South African government is in the preliminary phase of AI adoption, it has tried its best to draft regulations (e.g., *South African National Artificial Intelligence Policy Framework*) governing AI design, seeking to curb ethical dilemmas, such as data privacy, security and public trust (Giwa & Ngepah, 2024). However, parliament has not yet approved this draft policy regulating AI. Van Noordt et al. (2025) claim that the public sector can only thrive from challenges linked to the design and implementation of AI by establishing a state-of-the-art legislative framework to address legal issues that may arise.

Towards The Development of An Integrated Conceptual Model

This study presents an integrated conceptual model designed to provide a blueprint for ethical and governance concerns of AI in the South African public sector. The fundamental goal of the framework is to illustrate the most significant ethical and governance problems that are imperative for adopting AI in the South African public sector. Most notably, the framework assimilates AI into organizational processes and systems to enhance the efficiency and effectiveness of service delivery in the South African public sector. Additionally, the envisaged framework intends to attain efficient value chain activities and provide cost-effective and modernized public services to the citizens and disparate stakeholders to realize the predetermined objectives of public services in South Africa. Figure 2 depicts a suggested framework.

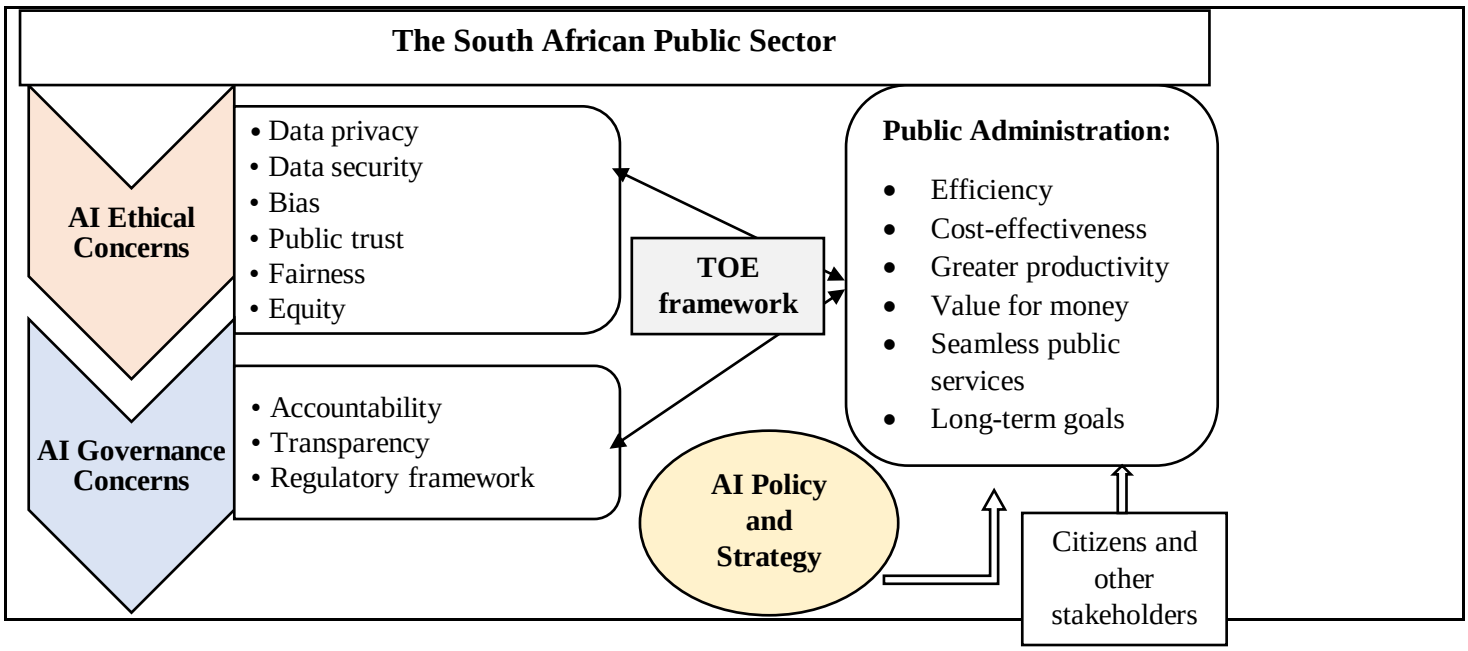

Figure 1: The integrated conceptual model for ethical AI

Managerial Implications and Contribution Of the Study

AI is swiftly revolutionizing public administration in a global phenomenon. The significance of AI in the public sector has rendered efficient, innovative and seamless public services to the citizens, specifically in developed countries (Robles & Mallinson, 2025). This study aimed to examine ethical and governance issues of AI in the South African public sector, an emerging economy country. While the adoption of AI in the South African public sector is still in embryonic phase, the findings of this study can help policy-makers, practitioners, strategists, and professionals in public institutions to take cognizance of ethical and governance issues (e.g., data privacy, security, transparency, trust, bias, and accountability) when developing and implementing 4IR policies, models and regulatory frameworks pertaining to AI.

On the other hand, the rapid systematic literature analysis makes a unique contribution to the body of knowledge by providing a crucial insight and a better understanding of the incorporation of ethical and governance issues pertaining to the AI adoption in the public sector, particularly for organisations in the nascent phase of AI, like South Africa. This study contributes to the literature by suggesting an integrated conceptual model that is imperative for AI adoption in the South African public sector. The model intends to improve AI adoption not only in the South African public sector. The model seeks to address AI ethical and governance issues, as well as current service delivery-related issues, by providing a structured and holistic approach to enhancing public administration and service delivery through the adoption of AI technologies.

CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS

The study examines ethical and governance issues of AI in the South African public sector, an emerging economy country. Although AI has been applauded for its capacity to ensure streamlined and efficient service delivery to the public, emerging economy countries like South Africa are still beleaguered with obstacles such as a lack of a legislative and regulatory framework governing AI and ethical issues. However, this study suggested an integrated conceptual model that could be valuable in guiding the development of AI policy and legislative framework in the South African public sector.

This study emphasized the rapid systematic literature analysis to probe ethical and governance concerns of AI in the South African public sector. Future studies can focus on other methodological choices, such as quantitative (e.g., surveys) and qualitative (e.g., interviews, focus groups and case studies) approaches to collect primary data from the respondents. This can enable the researcher to gain profound insight into AI's ethical and governance issues while validating the research results and ensuring the generalizability of research findings. As such, the suggested conceptual framework can be tested through these methods.

It is vital for the South African public sector to fast-track the adoption of AI-related technologies in order to thrive in the turbulent and changing digital economy. It is suggested that ethical issues such as data privacy, security, transparency, public trust, fairness, equity, bias, and accountability should all be integrated into AI systems and algorithms to enhance their functionality and public confidence while implementing and employing those AI technologies. Further to this, it is recommended that the South African government strengthen its AI regulatory framework by developing and enforcing a sound AI policy and strategy to accelerate the adoption of AI-enabled technologies in various spheres of government (national, provincial and local).

Ethical Considerations

The present research work contains no studies performed on animals/humans.

Conflict Of Interest

The author declares that he has no financial or personal relationships that may have inappropriately influenced him in writing this article.

Data Availability

The authors confirm that the data supporting the findings of this study are available within the article and in the bibliography.

REFERENCES

1. Agbarakwe, H. A., & Chibueze, O. O. (2024). Leveraging artificial intelligence for enhanced assessment and feedback mechanisms in Nigeria's higher education system. International Journal of Research and Innovation in Social Science, 8(9), 142-151. https://dx.doi.org/10.47772/IJRISS.2024.809012
2. Ahmed, S. K., Mohammed, R. A., Nashwan, A. J., Ibrahim, R. H., Abdalla, A. Q., Ameen, B. M. M., & Khdhir, R. M. (2025). Using thematic analysis in qualitative research. Journal of Medicine, Surgery and Public Health, 6, 1-6. https://doi.org/10.1016/j.glmedi.2025.100198
3. Alaran, M. A., Lawal, S. K., Jiya, M. H., Egya, S. A., Ahmed, M. M., Abdulsalam, A., ... & Lucero-Prisno III, D. E. (2025). Challenges and opportunities of artificial intelligence in the African health space. Digital Health, 11, 1-7. https://doi.org/10.1177/20552076241305915
4. Anshari, M., Hamdan, M., Ahmad, N., & Ali, E. (2025). Public service delivery, artificial intelligence and the sustainable development goals: Trends, evidence and complexities. Journal of Science and Technology Policy Management, 16(1), 163-181. https://doi.org/10.1108/JSTPM-07-2023-0123
5. Attard-Frost, B., Brandes, A., & Lyons, K. (2024). The governance of artificial intelligence in Canada: Findings and opportunities from a review of 84 AI governance initiatives. Government Information Quarterly, 41(2), 1-24. https://doi.org/10.1016/j.giq.2024.101929
6. Baker, J. 2012. The technology–organization–environment framework. Information Systems Theory: Explaining and Predicting Our Digital Society, 1: 231-245.
7. Baloyi, W.M., & Beyers, L.J.E. (2019). The impact of financial malpractice on service delivery: A conceptual examination of literature, in 4th Annual International Conference on Public Administration and Development Alternatives, 03–05 July, Johannesburg, viewed from http://ulspace.ul.ac.za/bitstream/handle/10386/ 2756/baloyi_impact_2019.pdf?sequence=1.
8. Barodi, M., & Lalaoui, S. (2025). Civil servants' readiness for AI adoption: The role of change management in Morocco's public sector. Problems and Perspectives in Management, 23(1), 63-75. http://dx.doi.org/10.21511/ppm.23(1).2025.05
9. Bester, J. (2024). Exploring the preparedness of South African rural municipalities in adopting and using artificial intelligence to improve service delivery. Journal of Public Administration and Development Alternatives (JPADA), 9(si1), 27-41. https://doi.org/10.55190/JPADA.2024.301
10. Bryan, J. D. & Zuva, T. 2021. A review on TAM and TOE framework progression and how these models integrate. Advances in Science, Technology and Engineering Systems Journal, 6(3): 137-145. https://dx.doi.org/10.25046/aj060316
11. Chilunjika, A. (2024). A review of the risks, challenges and benefits of using artificial intelligence (AI) technologies in public policy-making in South Africa. Social Sciences, Humanities and Education Journal (SHE Journal), 5(3), 393-411. http://e-journal.unipma.ac.id/index.php/SHE
12. Chitimira, H., & Munedzi, S. (2024). Exploring the role of artificial intelligence in eradicating poverty and enhancing global security in South Africa: A regulatory perspective. EuroEconomica, 43(1), 84-96. https://creativecommons.org/licenses/by-nc/4.0/
13. de Almeida, P. G. R., & dos Santos Júnior, C. D. (2025). Artificial intelligence governance: Understanding how public organisations implement it. Government Information Quarterly, 42(1), 1-18. https://doi.org/10.1016/j.giq.2024.102003
14. de Fine Licht, K. (2025). Resolving value conflicts in public AI governance: A procedural justice framework. Government Information Quarterly, 42(2), 1-11. https://doi.org/10.1016/j.giq.2025.102033
15. Diallo, K., Smith, J., Okolo, C. T., Nyamwaya, D., Kgomo, J., & Ngamita, R. (2025). Case studies of AI policy development in Africa. Data & Policy, 7, 1-12. https://doi.org/10.1017/dap.2024.71
16. Erkkilä, T. (2023). Global indicators and AI policy: Metrics, policy scripts, and narratives. Review of Policy Research, 40(5), 811-839. https://doi.org/10.1111/ropr.12556

17. Garritty, C., Hamel, C., Trivella, M., Gartlehner, G., Nussbaumer-Streit, B., Devane, D., Kamel, C., Griebler, U., & King, V. J. (2024). Updated recommendations for the Cochrane rapid review methods guidance for rapid reviews of effectiveness. Research Methods and Reporting, 384, 1-13. http://dx.doi.org/10.1136/bmj-2023-076335
18. Giwa, F., & Ngepah, N. (2024). Artificial intelligence and skilled employment in South Africa: Exploring key variables. Research in Globalization, 8, 1-11. https://doi.org/10.1016/j.resglo.2024.100231
19. Henk, A., & Henk, O. (2025). From antecedents to outcomes: A structured literature review on AI implementation in public sector organisations. [Paper presentation]. Proceedings of the 58th Hawaii International Conference on System Sciences. https://hdl.handle.net/10125/109068
20. Jankovski, B. S., Zola, F. C., Meneghel, G. V. G., & Zarelli, P. R. (2025). Implementation of artificial intelligence solutions in the Brazilian public sector: Challenges and opportunities. Revista Foco, 18(1), 1-16. https://doi.org/10.54751/revistafoco.v18n1-043
21. Janneker, W. (2025). The role of AI in transforming healthcare in South Africa. SAMJ: South African Medical Journal, 115(5B), 66-69. https://doi.org/10.7196/SAMJ.2025.v115i5b.3672
22. Machado, H., Silva, S., & Neiva, L. (2025). Publics' views on ethical challenges of artificial intelligence: A scoping review. AI and Ethics, 5(1), 139-167. https://doi.org/10.1007/s43681-023-00387-1
23. Malope, L. J. (2025). Unlocking the potential: Exploring the influence of artificial intelligence on e-health enhancement in South Africa. Journal of Economic and Social Development (JESD)–Resilient Society, 12(1), 146-156. https://www.researchgate.net/publication/392536423
24. Marwala, T. (2024). The balancing problem in the governance of artificial intelligence. Tokyo: Springer Nature.
25. Modiba, M. (2025). Adoption of artificial intelligence to enhance records management practices at the Gauteng Department of Education in South Africa. Collection and Curation, 44(1), 9-17. https://doi.org/10.1108/CC-12-2023-0044]
26. Mokwele, T. (2024). Exploring the potential of artificial intelligence for the South African Qualifications Authority. [Paper presentation]. The Focus Conference (TFC 2024. https://doi.org/10.2991/978-94-6463-630-7_15
27. Mtuze, S. S. K., & Morige, M. (2024). Towards drafting artificial intelligence (AI) legislation in South Africa. Obiter, 45(1), 161-179. https://hdl.handle.net/10520/ejc-obiter_v45_n1_a9
28. Naidoo, G. M. (2024). The potential of artificial intelligence in South African rural development. OIDA International Journal of Sustainable Development, 17(11), 207-218. https://www.ssrn.com/index.cfm/en/oida-intl-journal-sustainable-dev/
29. Nene, S. E., & Hewitt, L. M. (2023). Implementing artificial intelligence in South African public hospitals: A conceptual framework. Acta Commercii, 23(1), 1-6. https://doi.org/10.4102/ac.v23i1.1173
30. Nyugha, P. G. (2024). The importance of artificial intelligence to Africa's development process: Prospects and challenges. Primary Science and Engineering, 4(6), 20-30.
31. Rekunenko, I., Kobushko, I., Dzydzyguri, O., Balahurovska, I., Yurynets, O., & Zhuk, O. (2025). The use of artificial intelligence in public administration: Bibliometric analysis. Management, 23(1), 209-224. http://dx.doi.org/10.21511/ppm.23(1).2025.16
32. Robles, P., & Mallinson, D. J. (2025). Artificial intelligence technology, public trust, and effective governance. Review of Policy Research, 42(1), 11-28. https://doi.org/10.1111/ropr.12555
33. Saal, P. E., Chetty, K., Ntshayintshayi, N., Moosa, T., & Masuku, N. (2025). A scoping review of the integration of artificial intelligence in primary and secondary schools from 2020 to 2024: Policy implications for South Africa. Journal of Education (University of KwaZulu-Natal), 1(98), 62-85. http://journals.ukzn.ac.za/index.php/joe
34. 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
35. Tomazevic, N., Murko, E., & Aristovnik, A. (2024). Organisational enablers of artificial intelligence adoption in public institutions: A systematic literature review. Central European Public Administration Review, 22(1), 109-138. https://doi.org/10.17573/cepar.2024.1.05

36. Tornatzky, L. G. & Fleischer, M. 1990. Processes of technological innovation. Lexington, MA: Lexington Books.
37. Ulnicane, I., & Aden, A. (2023). Power and politics in framing bias in artificial intelligence policy. Review of Policy Research, 40(5), 665-687. https://doi.org/10.1111/ropr.12567
38. van Noordt, C., Medaglia, R., & Tangi, L. (2025). Policy initiatives for artificial intelligence-enabled government: An analysis of national strategies in Europe. Public Policy and Administration, 40(2), 215-253. https://doi.org/10.1016/j.giq.2023.101860
39. Vatamanu, A. F., & Tofan, M. (2025). Integrating artificial intelligence into public administration: Challenges and vulnerabilities. Administrative Sciences, 15(4), 1-23. https://doi.org/10.3390/admsci15040149
40. Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. Journal of Planning Education and Research, 39(1), 93-112. https://doi.org/10.1177/0739456X17723971
41. Zidouemba, M. T. (2025). Governance and artificial intelligence: the use of artificial intelligence in democracy and its impacts on the rights to participation. Discover Artificial Intelligence, 5(1), 1-11. https://doi.org/10.1007/s44163-025-00229-5