

Ethical AI Governance Frameworks for Public Sector Decision-Making

Blessing, Okon Tom

Lincoln International Business School
University of Lincoln, United Kingdom

Email: blessingtom282@gmail.com

Phone: +447473734308

Received: 10.02.2026 | Revised: 24.02.2026 | Accepted: 26.04.2026

Abstract

Purpose: The purpose of this paper is to critically examine the frameworks of ethical governance for artificial intelligence (AI) in public sector decision-making. It aims to explore the implications, challenges, and necessary ethical considerations that arise when AI technologies are deployed within government processes.

Design/Methodology: The paper adopts a purely doctrinal qualitative methodology, focusing on a critical review of existing literature on AI governance frameworks. Through a doctrinal analysis of ethical guidelines, case studies, and regulatory practices, the paper evaluates the alignment of current frameworks with the ethical principles needed for responsible AI implementation in public decision-making processes. A systematic review of academic articles, regulatory documents, and white papers on AI governance is conducted to examine the extent to which these frameworks address ethical challenges.

Findings: The findings indicate that while several AI governance frameworks exist, they often lack uniformity and fail to comprehensively address the unique ethical challenges posed by AI in public sector decision-making. Critical issues such as algorithmic bias, lack of transparency, and inadequate accountability mechanisms are highlighted as major barriers to effective AI governance. The paper argues for a robust, adaptable framework that incorporates ethical guidelines, regulatory oversight, and public engagement to ensure AI deployment is in the best interest of society.

Originality/Value: This paper offers a critical contribution to the discourse on AI governance by highlighting the ethical gaps in current frameworks. It proposes a comprehensive, dynamic model for AI governance in public decision-making that prioritizes transparency, inclusivity, and public trust. The study's value lies in its focus on applying ethical principles to the specific challenges faced by public institutions in integrating AI technologies.

Keywords: Artificial Intelligence, Ethical Governance, Public Sector Decision-Making, Algorithmic Accountability, Transparency, Fairness, Public Trust.

1. Introduction

Artificial intelligence (AI) has rapidly become an integral component of public sector decision-making, offering the potential to enhance efficiency, reduce human error, and improve services to the public. From automated policy analysis to predictive models for resource allocation, the use of AI is transforming how governments operate. However, this transformation brings with it significant ethical challenges. As AI systems are increasingly utilized to make decisions that affect public welfare, there is a growing need for robust ethical governance frameworks that ensure these technologies are deployed responsibly. AI governance refers to the policies, guidelines, and practices that regulate the development, deployment, and use of AI systems. Within the public sector, this governance must prioritize transparency, accountability, fairness, and inclusivity. However, many existing AI governance frameworks, while recognizing these values, fall short in providing a comprehensive solution to the ethical dilemmas posed by AI's capacity for bias, opacity, and decision-making that is difficult to audit or understand. This paper seeks to explore the ethical dimensions of AI governance frameworks within the context of public sector decision-making. It critically examines the existing governance models, identifies gaps in their ability to address the ethical implications of AI deployment, and proposes guidelines for developing more effective frameworks. The ethical principles of accountability, transparency, fairness, and privacy are fundamental to this exploration, as they underpin any legitimate AI governance model in the public sector.

2. Literature Review

Ethical Challenges in AI Governance for Public Sector Decision-Making

The integration of artificial intelligence (AI) into public sector decision-making has raised numerous ethical challenges, particularly concerning transparency, accountability, fairness, and privacy. AI systems, when deployed in governmental contexts, are expected to make decisions that affect public welfare, such as determining eligibility for social benefits or influencing judicial outcomes (Batool et al., 2025). The ethical dimensions of AI governance are crucial because these systems often operate in a "black box" mode, where decision-making processes are not fully

transparent or easily auditable (Camilleri, 2024). This opacity undermines public trust and poses risks of perpetuating biases that could disproportionately affect marginalized groups (Mišić et al., 2025). In public sector applications, where decisions have a direct impact on the well-being of citizens, these ethical issues are especially pronounced. For instance, the implementation of AI in areas such as law enforcement and healthcare has raised concerns about fairness and discrimination. Studies indicate that AI algorithms may inadvertently reinforce societal biases embedded in historical data, leading to discriminatory outcomes (Gunasekara, 2025). This phenomenon, often referred to as algorithmic bias, can result in unequal treatment of individuals based on race, gender, or socioeconomic status, which is unacceptable in the public sector context where equity and justice are paramount (Ribeiro et al., 2025).

Current AI Governance Frameworks

Various frameworks have been proposed to address these ethical challenges, but their application and effectiveness remain inconsistent. These frameworks typically emphasize principles such as transparency, accountability, fairness, and inclusivity, but often fail to provide practical guidance on how these principles should be implemented in public sector decision-making (Zhu, 2025). Some of the most widely recognized frameworks are those put forward by international bodies such as the European Union and the Organisation for Economic Co-operation and Development (OECD), which advocate for a human-centric approach to AI governance (Vatamanu, 2025). These frameworks suggest that AI systems should be designed and deployed with a focus on safeguarding human rights and promoting societal well-being. However, their broad nature and lack of specificity in terms of implementation often leave gaps in addressing real-world challenges. The EU's AI Act, for example, categorizes AI systems based on their risk levels, but its focus on high-risk applications does not adequately address the full spectrum of ethical concerns, especially in areas where AI systems interact with vulnerable populations (Kim et al., 2025). Moreover, while these frameworks stress the importance of transparency, they provide little guidance on how to achieve true transparency in the decision-making processes of complex AI systems. Without robust mechanisms for accountability, the risk of "black-box" decision-making remains a critical issue (Mišić et al., 2025).

The Role of Accountability and Transparency in AI Governance

Accountability and transparency are foundational principles in AI governance, particularly in the public sector, where trust and legitimacy are essential. AI systems must not only be transparent in their functioning but also accountable for their outcomes. This means that governments must have mechanisms in place to audit AI systems and ensure that decisions made by these systems can be traced and explained (Batool et al., 2025). Currently, many AI governance frameworks do not offer clear guidelines on how public sector agencies can establish these mechanisms. The ethical principle of accountability involves holding AI developers, deployers, and users responsible for the decisions made by AI systems. Without accountability, there is a risk of "ethical outsourcing," where governments rely on AI systems without fully understanding or overseeing their decision-making processes (Corrêa et al., 2022). This has been particularly problematic in areas like predictive policing, where AI systems are used to determine where police should patrol or whom they should investigate. The lack of transparency and accountability in these systems can lead to decisions that disproportionately affect marginalized communities, raising serious concerns about the fairness and legitimacy of AI in public sector decision-making (Goh & Griva, 2018).

The Need for Inclusive and Participatory AI Governance Models

In addition to transparency and accountability, the principle of inclusivity is critical in AI governance. AI systems should be designed in a way that includes diverse perspectives, particularly from marginalized communities who are often the most affected by the decisions made by these systems. Public sector decision-making through AI must take into account the social, political, and economic implications of technological choices. Inclusive AI governance requires the participation of various stakeholders, including policymakers, ethicists, technologists, and the general public, in shaping the development and deployment of AI systems (Vatamanu, 2025).

Incorporating diverse perspectives is especially important in the design of AI systems that affect public welfare. For example, the development of AI systems for healthcare delivery must consider the needs and values of diverse populations, including those from different socioeconomic backgrounds and cultural contexts. A participatory approach to AI governance helps ensure that AI systems are not only technically

efficient but also socially responsible and aligned with the values of the society they serve (Madanchian et al., 2025).

Regulatory and Ethical Gaps in AI Governance

Despite the numerous frameworks available, gaps in regulation and ethical guidelines remain a significant concern. Many of the existing frameworks focus primarily on risk assessment and mitigation, without addressing the underlying ethical implications of AI deployment (de Almeida et al., 2025). Additionally, the rapid pace of AI development often outstrips the ability of regulatory bodies to establish meaningful guidelines that keep up with emerging technological capabilities. This creates a situation where ethical considerations are often reactive rather than proactive, resulting in insufficient safeguards for the public. AI governance frameworks also face challenges in dealing with the dynamic and evolving nature of AI technologies. As AI systems become more autonomous and capable of learning from new data, the ethical implications of their actions can change over time, creating difficulties for regulators who are tasked with ensuring that these systems continue to meet ethical standards (Mišić et al., 2025). The lack of adaptability in current frameworks makes it challenging to address these evolving ethical dilemmas in real-time.

Toward a Comprehensive Ethical AI Governance Framework

To address these ethical gaps, scholars have called for the development of a more comprehensive and adaptable ethical AI governance framework that not only emphasizes transparency, accountability, and fairness but also integrates mechanisms for continuous evaluation and public participation. A truly ethical AI governance framework in the public sector must prioritize the public interest and ensure that AI systems are used to promote societal well-being, rather than exacerbate inequality or infringe upon fundamental rights (Corrêa et al., 2022). Such a framework would include clear guidelines for the design, deployment, and monitoring of AI systems, with mechanisms for ongoing assessment to ensure that these systems remain ethical over time. Moreover, it would emphasize the importance of inclusivity, ensuring that AI systems are designed with input from diverse stakeholders and are accessible to all segments of society. In doing so, AI governance can move beyond mere compliance with regulatory standards and towards a more holistic approach that prioritizes ethical decision-making and the safeguarding of public trust (Madanchian et al., 2025).

3. Methodology

This study adopts a purely doctrinal qualitative research methodology to critically examine existing ethical AI governance frameworks for public sector decision-making. A doctrinal approach, commonly used in legal and ethical research, involves the critical analysis and interpretation of primary and secondary legal texts, policies, regulatory documents, and academic literature. The purpose of this methodological approach is to explore and analyze the theoretical and practical dimensions of AI governance in public administration and to identify gaps in current frameworks with a view to proposing solutions for improvement. The data for this study were primarily gathered through a comprehensive review of the literature on AI governance, focusing on ethical frameworks that have been developed for use in the public sector. Key documents reviewed include academic articles, reports from regulatory bodies, and guidelines from international organizations such as the European Union (EU) and the Organisation for Economic Co-operation and Development (OECD). Additionally, reports from think tanks and non-governmental organizations (NGOs) focused on AI ethics and governance were included to provide a holistic understanding of the current state of AI governance. The sources were selected based on their relevance to the topic of AI governance in the public sector and their focus on ethical principles, transparency, accountability, and inclusivity in AI deployment. The review process involved identifying and synthesizing key themes and arguments related to the ethical challenges of AI in public decision-making, as well as evaluating the effectiveness of current governance models in addressing these challenges.

Data Collection

The primary data sources for this study were obtained through electronic databases such as Google Scholar, JSTOR, SpringerLink, and ScienceDirect. These platforms were used to locate peer-reviewed journal articles, regulatory frameworks, and reports that provide insight into the ethical implications of AI in governance. A snowball sampling technique was also employed, where references within selected articles and documents were further explored to ensure comprehensive coverage of the subject.

Data Analysis

The data analysis process involved the thematic analysis of the literature. Thematic analysis is a widely used qualitative research method that identifies and analyzes patterns or themes within the data. In this study, themes related to key ethical principles such as fairness, transparency, accountability, and privacy were extracted from the selected texts. These themes were then critically examined to identify how well current governance frameworks address these issues in the context of AI deployment in the public sector. Special attention was given to evaluating whether the frameworks incorporate sufficient safeguards to prevent algorithmic bias, promote public trust, and ensure equitable treatment of all citizens.

4. Results

Thematic Findings from the Literature

The literature revealed several recurring themes and issues related to ethical AI governance frameworks in the public sector. These can be categorized into four main areas: algorithmic bias, transparency and accountability, public trust, and inclusivity.

Algorithmic Bias

A significant concern identified in the literature is the risk of algorithmic bias. Numerous studies highlighted that AI systems often inherit biases from historical data, which can lead to discriminatory outcomes. These biases may disproportionately affect marginalized groups, including racial minorities and low-income populations. Gunasekara (2025) and Camilleri (2024) argue that public sector applications of AI, such as in law enforcement or social welfare programs, are particularly vulnerable to the perpetuation of existing societal inequalities. These findings suggest a need for AI governance frameworks to prioritize mechanisms for detecting and mitigating bias in algorithmic decision-making.

Transparency and Accountability

Transparency and accountability emerged as crucial elements in ensuring the ethical deployment of AI in public sector decision-making. Several scholars, including Mišić et al. (2025) and Zhu (2025), emphasized that the opacity of AI decision-making processes undermines public trust in government use of these technologies. The

literature also revealed that while some frameworks, such as the EU's AI Act, advocate for transparency, there remains a significant gap in how to make complex algorithms explainable to the public. This lack of clarity in decision-making processes leads to concerns over accountability when AI systems make erroneous or biased decisions.

Public Trust

Public trust was frequently mentioned as a critical factor in the successful implementation of AI governance frameworks. Ribeiro et al. (2025) found that trust in AI systems is closely tied to the perceived fairness and transparency of these systems. In the public sector, where decisions made by AI systems affect individuals' lives and well-being, ensuring public trust is essential. However, the literature points out that the lack of clear accountability mechanisms and the potential for discrimination erode public confidence in AI-driven decisions. The importance of building and maintaining public trust through participatory governance and open communication about the use of AI in public policy decisions was highlighted by authors such as Madanchian et al. (2025).

Inclusivity

Inclusivity in AI governance emerged as another significant theme. Several sources (e.g., Vatamanu, 2025; Madanchian et al., 2025) underscored the need for inclusive approaches to AI design and implementation, ensuring that diverse perspectives are considered in the development of AI systems. This is particularly important in the public sector, where decisions must reflect the needs and values of a broad range of citizens. However, the literature suggests that many AI governance frameworks do not adequately address the need for diversity in AI development teams or in the decision-making processes that determine how AI systems are deployed.

Evaluation of Current AI Governance Frameworks

The analysis of existing AI governance frameworks revealed significant gaps in addressing the ethical challenges identified above. Although frameworks such as the EU's AI Act and the OECD's principles for trustworthy AI provide broad guidelines

for ethical AI use, they often lack specificity in addressing the nuanced challenges that arise in public sector applications. For instance, while the EU AI Act emphasizes transparency, it does not offer concrete solutions for ensuring that AI systems are fully explainable or auditable (Kim et al., 2025). Furthermore, the issue of algorithmic bias is only briefly addressed in many frameworks, with few detailed provisions for detecting and mitigating such biases in practice. Moreover, many frameworks focus on high-risk AI applications but fail to provide comprehensive guidelines for lower-risk public sector uses, such as administrative tasks or decision-making in healthcare and education. This suggests the need for a more comprehensive and adaptable AI governance framework that can be applied to a wide range of public sector activities, with particular attention to ensuring that AI systems are deployed in a way that is both ethical and inclusive.

5. Discussion

The findings from the literature reveal that while ethical AI governance frameworks in the public sector have made strides in addressing transparency, accountability, fairness, and inclusivity, significant gaps remain. These gaps can lead to unintended ethical dilemmas, especially in areas like algorithmic bias, trust erosion, and unequal representation in AI design. This section critically discusses these findings, explores the implications for future AI governance frameworks, and offers recommendations for addressing the identified challenges.

Ethical Implications of Algorithmic Bias

Algorithmic bias remains one of the most pressing ethical issues in AI governance, particularly within the public sector, where the consequences of biased decision-making can be far-reaching. The literature emphasizes that AI systems are often trained on historical data that reflects societal biases, which can lead to discriminatory outcomes in areas such as criminal justice, healthcare, and social welfare (Gunasekara, 2025). These biases not only undermine the fairness of AI-driven decisions but also perpetuate existing inequalities. The challenge, however, lies in developing effective methods for detecting and mitigating bias in AI algorithms. As identified in the literature, current governance frameworks offer limited solutions for addressing these

issues, leaving public sector agencies with insufficient tools to safeguard against bias in AI systems (Mišić et al., 2025). This gap in ethical governance calls for frameworks that go beyond theoretical principles and provide practical, actionable guidelines for mitigating algorithmic bias. Public sector AI governance should include mandatory auditing processes, the use of diverse and representative data sets, and the incorporation of fairness metrics into AI system design. Moreover, policymakers must ensure that AI developers and deployers are held accountable for the ethical implications of their systems' outputs, particularly in public-facing applications that impact vulnerable populations.

Challenges to Transparency and Accountability

Transparency and accountability are foundational principles in AI governance, yet the review revealed that current frameworks often fail to provide concrete measures for achieving these goals. The complexity and opacity of AI systems make it difficult for citizens and regulatory bodies to understand how decisions are made, which erodes trust in the technology (Ribeiro et al., 2025). Without clear accountability mechanisms, the public sector risks adopting AI systems that operate without adequate oversight, making it difficult to rectify harmful or biased decisions once they occur. The need for a more transparent approach to AI governance is crucial. Current frameworks, including the EU AI Act, make strides in promoting transparency but fall short of ensuring true accountability for the outcomes of AI decision-making (Kim et al., 2025). One key challenge is the "black-box" nature of many AI algorithms, which can make even experts struggle to explain the reasoning behind AI decisions. Public sector institutions must take proactive steps to ensure that AI systems are explainable, traceable, and subject to regular audits. This could involve the development of standardized frameworks for AI accountability that include clear mechanisms for addressing grievances and correcting errors in AI-driven decision-making processes.

Building and Maintaining Public Trust

Public trust in AI systems is critical for their successful integration into public sector decision-making. However, the literature underscores that trust is often undermined by the lack of transparency, accountability, and fairness in AI systems. As public sector decisions directly impact citizens' lives, maintaining public trust is essential to ensuring the legitimacy of AI deployment. Trust is not merely a byproduct of

technological efficiency; it must be cultivated through transparency, inclusivity, and clear communication about how AI systems are being used and what ethical standards govern their operations (Madanchian et al., 2025). Governments must adopt a participatory approach to AI governance, involving diverse stakeholders—including the public, civil society organizations, and marginalized groups—in the decision-making process. This would help build a sense of ownership and collective responsibility over AI systems. Moreover, public sector agencies should regularly report on the performance of AI systems, including their ethical performance, to reassure citizens that these systems are being used responsibly and in alignment with societal values.

Inclusivity and Representation in AI Governance

Inclusivity remains a significant issue in the governance of AI systems, particularly in the public sector. The literature reveals that many AI systems are developed by homogenous teams, which can lead to blind spots in the design and deployment of AI technologies (Vatamanu, 2025). For example, AI systems used in healthcare, criminal justice, and social services often fail to account for the diverse needs of different populations. The result is that these systems may inadvertently disadvantage certain groups, particularly those that are already marginalized. An inclusive approach to AI governance requires a concerted effort to integrate diverse perspectives into the design, development, and deployment of AI systems. Public sector institutions must ensure that AI governance frameworks prioritize the participation of underrepresented groups in the decision-making process. This could include the establishment of advisory boards or ethics committees that consist of diverse stakeholders, including ethicists, sociologists, legal experts, and community representatives.

6. Conclusion

The integration of AI into public sector decision-making presents both significant opportunities and ethical challenges. While existing AI governance frameworks have made notable progress in addressing issues such as fairness, accountability, transparency, and inclusivity, they still fall short in providing comprehensive solutions to the ethical dilemmas posed by AI technologies. This paper has critically examined the current state of AI governance in the public sector, highlighting the

ethical implications of algorithmic bias, transparency, accountability, public trust, and inclusivity. To address these challenges, this study proposes the development of a more robust and adaptable AI governance framework that prioritizes transparency, accountability, and inclusivity while actively mitigating algorithmic bias. Public sector institutions must implement clear, practical guidelines for ensuring that AI systems are explainable, auditable, and accountable. Furthermore, AI governance frameworks must be dynamic, evolving with the rapidly changing technological landscape to address emerging ethical concerns in real-time.

Ultimately, for AI to be ethically deployed in public decision-making, governments must take a proactive approach to AI governance that places the public interest at the center. By incorporating diverse perspectives and ensuring that AI systems serve the needs of all citizens, public sector AI governance can foster greater trust, equity, and societal well-being.

7. References

- Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: A systematic literature review. *AI and Ethics*. <https://doi.org/10.1007/s43681-024-00653-w>
- Camilleri, M. A. (2024). Artificial intelligence governance: Ethical considerations and frameworks. *Expert Systems*. <https://doi.org/10.1111/exsy.13406>
- Corrêa, N. K., Galvão, C., Santos, J. W., et al. (2022). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *arXiv*. <https://arxiv.org/abs/2206.11922>
- de Almeida, P. G. R., et al. (2025). Understanding how public organizations implement ethical AI governance. *Government Information Quarterly*.
- Gunasekara, L. (2025). A systematic review of responsible artificial intelligence principles and practice. *Electronics*. <https://www.mdpi.com/2571-5577/8/4/97>
- Madanchian, M., et al. (2025). Ethical theories, governance models, and strategic AI implementation. *Frontiers in AI*.
- Ribeiro, D., Rocha, T., Pinto, G., et al. (2025). Toward effective AI governance: A review of principles. *arXiv*. <https://arxiv.org/abs/2505.23417>

Zhu, X. (2025). Public administration with, of, and through AI. *International Journal of Public Sector Management*.

Mišić, J., van Est, R., & Kool, L. (2025). Good governance of public sector AI: A combined value framework. *AI & Ethics*.
<https://link.springer.com/article/10.1007/s43681-025-00751-3>

Vatamanu, A. F. (2025). Integrating AI into public administration: Governance and ethical challenges. *Business Systems Research*.
<https://www.mdpi.com/2076-3387/15/4/149>

Kim, B.-J., Jeong, S., Cho, B.-K., & Chung, J.-B. (2025). AI governance in the context of the EU AI Act: A literature review. *arXiv*.
<https://arxiv.org/abs/2502.03468>