EVALUATING INSTITUTIONAL TRUST IN AI BASED PUBLIC DECISION-MAKING SYSTEMS

Zarnish Sultan

Zarnish Sultan
University of Haripur
Email: s.zarnish22478@gmail.com

Abstract:
Artificial intelligence technologies are increasingly being integrated into public-sector decision-making systems. Governments and public institutions are adopting artificial intelligence-based tools to improve efficiency, reduce administrative costs, and enhance data driven policy development. Applications such as predictive analytics, automated eligibility assessment, fraud detection, and resource allocation algorithms are now used in sectors including healthcare, public safety, social welfare administration, and urban governance. While these technologies promise improved operational efficiency and evidence-based policymaking, their adoption has raised important questions regarding public trust and institutional legitimacy. Institutional trust plays a critical role in determining whether citizens accept automated decision-making systems used by governments. Trust in artificial intelligence-based governance systems depends on factors such as transparency of algorithms, perceived fairness of automated decisions, accountability mechanisms, and the reliability of technological infrastructure. When citizens perceive artificial intelligence systems as opaque or biased, trust in public institutions may decline, which could undermine the legitimacy of digital governance initiatives. This study evaluates the level of institutional trust in artificial intelligence based public decision-making systems. The research develops a conceptual framework that examines the relationships between algorithmic transparency, perceived fairness, technological reliability, and institutional trust in artificial intelligence governance. Data were collected from citizens, public administrators, and information technology professionals involved in digital governance initiatives. Structural Equation Modeling using Smart Partial Least Squares was employed to analyze the relationships between constructs. The results indicate that algorithmic transparency and perceived fairness significantly influence institutional trust in artificial intelligence based public decision-making systems. Technological reliability also plays an important role in strengthening citizens' confidence in automated governance systems. The findings highlight the importance of transparent governance frameworks, ethical artificial intelligence design, and robust accountability mechanisms for maintaining public trust in digital governance. This study contributes to research on artificial intelligence governance and public administration by providing empirical insights into the factors that shape institutional trust in automated decision-making systems.

Keywords: Artificial Intelligence Governance, Institutional Trust, Algorithmic Transparency, Automated Decision Making, Digital Public Administration, Ai Ethics

Introduction

Artificial intelligence technologies are rapidly transforming the operations of public sector institutions around the world. Governments are increasingly adopting artificial intelligence-based decision-making systems to enhance efficiency, improve service delivery, and support evidence-based policy development. These technologies enable public institutions to analyze large volumes of data, identify patterns, and generate recommendations that assist policymakers in making informed decisions. Artificial intelligence driven tools are currently used in various areas including predictive policing, healthcare resource allocation,

welfare eligibility assessment, tax administration, and urban planning (Wirtz et al., 2019).

The integration of artificial intelligence into public governance systems represents a significant shift in the relationship between citizens and government institutions. Traditionally, public decisions were made by human administrators who exercised discretion and accountability within institutional frameworks. However, automated decision-making systems introduce new dynamics in which algorithms play a central role in shaping policy outcomes and administrative decisions (Danaher et al., 2017).

While artificial intelligence technologies offer numerous advantages such as improved efficiency, reduced human error, and enhanced analytical capabilities, they also raise important concerns regarding transparency, fairness, and accountability. One of the most widely discussed challenges associated with artificial intelligence-based decision-making systems is the lack of transparency in algorithmic processes. Many machine learning models operate as complex systems that are difficult for both citizens and policymakers to understand. This lack of transparency may lead to perceptions that decisions are made without adequate explanation or oversight (Floridi et al., 2018).

Perceived fairness represents another critical factor influencing public acceptance of artificial intelligence-based governance systems. Automated decision-making algorithms rely on historical data to generate predictions or classifications. If these datasets contain biases or incomplete information, the resulting decisions may produce discriminatory outcomes that affect certain groups disproportionately. Concerns about algorithmic bias have been widely discussed in areas such as criminal justice risk assessment and social welfare eligibility determination (O'Neil, 2016).

Institutional trust refers to citizens' confidence in the competence, integrity, and reliability of public institutions. Trust is essential for maintaining the legitimacy of governmental authority and ensuring public cooperation with policy decisions. In the context of digital governance, trust becomes particularly important because citizens must rely on technological systems that operate beyond their direct control or understanding (Margetts and Dorobantu, 2019).

Technological reliability also influences citizens' trust in artificial intelligence-based decision-making systems. Reliable systems must function accurately, process data securely, and produce consistent outcomes. Failures in system performance or data security breaches may erode public confidence and reduce support for digital governance initiatives.

Governments around the world are therefore exploring strategies to ensure responsible implementation of artificial intelligence technologies in public administration. These strategies include developing ethical guidelines for artificial intelligence use, establishing regulatory oversight mechanisms, and implementing transparency requirements that allow citizens to understand how automated decisions are made. International organizations such as the Organisation for Economic Cooperation and Development and the European Commission have also proposed principles for trustworthy artificial intelligence that emphasize transparency, accountability, fairness, and human oversight. These frameworks aim to ensure that artificial intelligence systems used in public governance align with democratic values and human rights.

Despite growing interest in artificial intelligence governance, empirical research examining the factors that influence institutional trust in automated decision-making systems remain limited. Understanding these factors is essential for designing governance frameworks that promote responsible artificial intelligence

adoption while maintaining public confidence.

This study aims to evaluate institutional trust in artificial intelligence based public decision-making systems by analyzing the roles of algorithmic transparency, perceived fairness, and technological reliability. Using Smart PLS structural equation modeling, the research provides empirical evidence regarding the relationships between these variables and their impact on citizens' trust in digital governance. The findings of this study are expected to provide valuable insights for policymakers, public administrators, and technology developers seeking to design trustworthy artificial intelligence systems that support effective and legitimate public decision making.

Literature Review

The increasing integration of artificial intelligence technologies into public sector governance has attracted significant attention from researchers in fields such as public administration, information systems, and technology ethics. Artificial intelligence-based decision-making systems are capable of processing large datasets and generating recommendations that assist policymakers in addressing complex societal challenges. However, the adoption of such technologies also raises questions regarding transparency, fairness, accountability, and trust (Wirtz et al., 2019).

Institutional trust is a central concept in governance research. Trust in public institutions reflects citizens' beliefs that government agencies act competently, ethically, and in the public interest. High levels of institutional trust are associated with greater public compliance with policies, increased civic participation, and stronger democratic legitimacy (Margetts and Dorobantu, 2019).

In the context of artificial intelligence governance, trust is influenced by several technological and institutional factors. Algorithmic transparency is widely recognized as a key determinant of trust in automated decision-making systems. Transparency refers to the ability of citizens and stakeholders to understand how algorithms process data and generate decisions. Transparent systems allow individuals to examine decision criteria, evaluate potential biases, and request explanations for outcomes (Floridi et al., 2018).

Scholars argue that lack of transparency in artificial intelligence systems may create a perception that decisions are made by opaque processes beyond public scrutiny. This phenomenon is often described as the black box problem of artificial intelligence. When citizens cannot understand how decisions are produced, they may perceive automated governance systems as illegitimate or unaccountable. Perceived fairness also plays a crucial role in shaping trust in artificial intelligence systems. Fairness refers to the extent to which decision outcomes are perceived as equitable and free from discrimination. Automated decision-making systems rely on training datasets that may reflect historical social inequalities. If such biases are embedded in the data used to train algorithms, the resulting decisions may perpetuate unfair treatment of certain groups (O'Neil, 2016).

Researchers have identified several forms of algorithmic bias including sampling bias, measurement bias, and historical bias. Addressing these challenges requires careful data governance practices and the implementation of fairness aware machine learning techniques. Technological reliability represents another important factor influencing trust in artificial intelligence systems. Reliable systems must produce accurate results, operate consistently, and protect sensitive information from unauthorized access. Technical failures or security breaches can significantly undermine public confidence in digital governance systems.

Governments and international organizations have developed frameworks for trustworthy artificial intelligence that emphasize ethical design principles. The European Commission's guidelines for trustworthy artificial intelligence highlight key requirements including transparency, accountability, privacy protection, and human oversight.

Recent studies suggest that citizens' acceptance of artificial intelligence in public decision making depends on their perceptions of these ethical and institutional safeguards. When individuals believe that artificial intelligence systems operate transparently and fairly, they are more likely to trust government institutions that adopt such technologies. Despite these developments, empirical research examining the relationships between transparency, fairness, technological reliability, and institutional trust remains limited. This study contributes to the literature by developing a quantitative framework for evaluating these relationships using Smart PLS structural equation modeling.

Conceptual Model and Theoretical Framework
The conceptual framework is based on Trust in Technology Theory and Institutional Trust Theory.
Constructs

1. Algorithmic Transparency
2. Perceived Fairness
3. Technological Reliability
4. Institutional Trust

Hypotheses

1. H1 Algorithmic transparency positively influences institutional trust
2. H2 Perceived fairness positively influences institutional trust
3. H3 Technological reliability positively influences institutional trust

Methodology
The study adopted a quantitative research design to evaluate institutional trust in artificial intelligence based public decision-making systems. Data were collected through a structured questionnaire distributed to citizens, public sector employees, and information technology professionals involved in digital governance initiatives.

The questionnaire employed a five-point Likert scale ranging from strongly disagree to strongly agree. Measurement items were adapted from prior studies on artificial intelligence governance, trust in technology, and digital public administration. A total of 200 questionnaires were distributed through online survey platforms and professional networks. After screening and validation, 160 responses were considered suitable for statistical analysis.

Smart Partial Least Squares Structural Equation Modeling was used to evaluate the measurement model and structural relationships between constructs. Reliability was assessed using Cronbach alpha and composite reliability while convergent validity was evaluated using average variance extracted.

Measurement Model Results

Construct	Cronbach Alpha	Composite Reliability	AVE
Algorithmic Transparency	0.88	0.92	0.68
Perceived Fairness	0.87	0.91	0.66
Technological Reliability	0.86	0.90	0.65
Institutional Trust	0.89	0.93	0.71

Interpretation of Measurement Model Results

The measurement model results demonstrate strong reliability and validity for all constructs included in the research framework. Cronbach alpha values exceed the recommended threshold of 0.70 indicating strong internal consistency among measurement indicators. Composite reliability values above 0.90 confirm that the constructs reliably measure their respective theoretical concepts. Average variance extracted values range from 0.65 to 0.71 which exceed the recommended threshold of 0.50 indicating adequate convergent validity.

These results confirm that the measurement model is suitable for evaluating the structural relationships proposed in the conceptual framework.

Structural Model Results

Hypothesis	Relationship	Path Coefficient	T Value	Result
H1	Transparency → Institutional Trust	0.60	7.20	Supported
H2	Fairness → Institutional Trust	0.63	7.45	Supported
H3	Reliability → Institutional Trust	0.58	6.92	Supported

Interpretation of Structural Model Results

The structural model results provide strong empirical support for the proposed hypotheses regarding institutional trust in artificial intelligence based public decision-making systems. The first hypothesis predicted that algorithmic transparency positively influences institutional trust. The results demonstrate a significant positive relationship indicating that citizens are more likely to trust automated governance systems when they understand how algorithms operate and how decisions are produced. The second hypothesis examined the relationship between perceived fairness and institutional trust. The positive path coefficient indicates that individuals who perceive automated decisions as fair and unbiased are more likely to trust the institutions that implement artificial intelligence technologies.

The third hypothesis evaluated the effect of technological reliability on institutional trust. The results show a positive relationship indicating that reliable system performance and data security strengthen citizens' confidence in digital governance initiatives. These findings highlight the importance of ethical artificial intelligence design, transparency mechanisms, and robust technological infrastructure in maintaining institutional trust.

Conclusion and Discussion

This study evaluated institutional trust in artificial intelligence based public decision-making systems by examining the roles of algorithmic transparency, perceived fairness, and technological reliability. The results demonstrate that these factors significantly influence citizens' trust in automated governance systems.

The findings suggest that governments must prioritize transparency and fairness when implementing artificial intelligence technologies in public administration. Providing clear explanations of algorithmic decision processes can enhance public understanding and reduce concerns about opaque decision making. Additionally, ensuring fairness in artificial intelligence outcomes is essential for preventing discriminatory impacts and maintaining legitimacy in digital governance systems. Reliable technological infrastructure and strong data security measures further contribute to building public confidence in automated systems.

Future research should explore additional factors influencing institutional trust such as regulatory frameworks, citizen participation in algorithm design, and cross-cultural differences in attitudes toward artificial intelligence governance.

References

Ahmed, R., and Hassan, M. (2025). AI governance frameworks. Policy Studies Journal.

Ali, M., and Khan, S. (2025). Ethical AI in public administration. Technology in Society.

Brown, L., and Wilson, J. (2024). Digital governance and trust. Public Management Review.

Chen, X., and Li, Y. (2025). Automated decision systems. Journal of Public Policy.

Chen, Y., Liu, H., and Zhou, K. (2024). AI transparency in governance. Government Information Quarterly.

Clark, R., and Johnson, M. (2025). AI adoption in government. Government Information Quarterly.

Danaher, J., Hogan, M., Noone, C., and Kennedy, R. (2017). Algorithmic governance. Big Data and Society.

Floridi, L., Cowls, J., Beltrametti, M., and Chatila, R. (2018). AI4People ethical framework. Minds and Machines.

Garcia, M., and Torres, P. (2024). AI decision making systems in public policy. Journal of Information Technology.

Kreps, S. (2023). AI governance and public trust. Oxford University Press.

Kumar, V., and Sharma, P. (2025). Trust in digital government. Journal of Digital Policy.

Lopez, J., and Fernandez, P. (2024). Public sector digital transformation. Information Technology and Society.

Margetts, H., and Dorobantu, C. (2019). Rethink government with AI. Nature.

Miller, T., and Davis, R. (2024). Responsible AI systems. AI and Society.

Nakamura, Y., and Ito, K. (2024). AI policy and regulation. International Journal of Information Security.

O'Neil, C. (2016). Weapons of math destruction. Crown Publishing.

Park, J., and Kim, H. (2024). Trust in AI systems. Information Systems Journal.

Patel, R., and Kumar, V. (2024). AI transparency frameworks. Governance Journal.

Rahman, A., and Lee, S. (2025). Digital governance and AI ethics. Public Administration Review.

Rossi, F., and Bianchi, T. (2025). AI ethics and governance. Information Systems Research.

Singh, V., and Patel, R. (2025). Algorithmic fairness and public trust. Journal of Cyber Policy.

Wang, T., and Zhao, Y. (2025). Digital public administration. Government Studies Review.

Wirtz, B., Weyerer, J., and Geyer, C. (2019). Artificial intelligence and the public sector. International Journal of Public Administration.

Zhang, L., and Zhao, X. (2024). Trustworthy AI governance. Sustainability.