International Academic Institute for Science and Technology

International Academic Journal of Social Sciences
Vol. 13, No. 2, 2023, pp. 22-31.

ISSN 2454-3918

International Academic Journal of Social Sciences

www.iaiest.com

Responsible Artificial Intelligence Governance for Social Equity and Public Trust

Arjun Menon [a*], Kavya Nair [b]

[a*] *Bharat Institute for Future Technology Research, Bengaluru, Karnataka, India.*
[b] *Bharat Institute for Future Technology Research, Bengaluru, Karnataka, India.*

Abstract

The current study is concerned with responsible artificial intelligence (AI) governance and the role that responsible AI governance can play in social equity and fostering social trust in AI technologies. Specifically, the increasing need for AI governance mechanisms that would ensure fairness, transparency, accountability, privacy, explainability, and human oversight in the design and use of AI technologies was considered. A conceptual research approach was applied in the course of this study, and relevant literature was explored and analyzed on the topics of responsible AI, AI governance, social equity, algorithmic fairness, and public trust. Based on the findings obtained from the thematic analysis of the literature on the above topics, a framework for AI governance was developed, which incorporated ethical principles, stakeholder involvement, bias management, transparency measures, privacy preservation, and human oversight as determinants of socially responsible AI. Specific attention was paid to the part played by stakeholder engagement in making sure that the processes of AI governance took into account the interests of diverse social groups. Institutional responsibilities, algorithm assessment, and mechanisms for handling complaints about AI decisions were viewed as particularly important in this context. Responsible AI governance was seen as an institutional way of aligning technological development with social values. This framework presented valuable recommendations for organizations and policymakers who wish to use AI for increasing social equity and fostering public trust. In general, this research showed that proper governance can help mitigate risks related to the use of AI in decision-making.

Keywords: Responsible Artificial Intelligence, AI Governance, Social Equity, Public Trust, Algorithmic Fairness, Transparency and Accountability, Human Oversight.

DOI: 10.71086/IAJSS/V13I2/IAJSS1312
Received: 03 March 2026/Revised: 21 April 2026/Accepted: 13 May 2026/Published: 30 June 2026

Introduction

Responsible Artificial Intelligence and Emerging Governance Needs

Artificial intelligence has increasingly been used in healthcare, finance, education, employment, public administration, transport, and other fields, where decisions directly impact individuals and communities. While AI has offered the possibility for more efficient work, forecasting, automation, and decision-making, increasing application of AI technologies has raised issues connected to bias, privacy, transparency, accountability, security, and inequality. In this regard, responsible AI governance has become crucial for ensuring that AI applications are designed and implemented in such a way that would correspond to the values and interests of society (Camilleri, 2024) (Mura & Stehlíková, 2025) (Sadanand & Sadanand, 2025).

The implementation of AI governance requires the organization to set up principles, responsibilities, monitoring, and controls during all stages of using AI. It also involves regular assessment of possible social risks and methods to deal with the harm.

The increasing involvement of AI in decision-making has also transformed relations between technology developers, organizations, users, and impacted communities. AI systems can influence decisions without individuals fully understanding how those decisions were generated. As such, there is an increase in the necessity for governance mechanisms that ensure the objectives, constraints, data handling, and accountability are clearer and identifiable. Thus, responsible AI governance must go beyond accuracy in terms of technicality and look into the impacts of deploying the AI technology socially, ethically, legally, and institutionally (Winfield & Jirotka, 2018; Cojocaru, 2025).

The governance mechanisms may be integrated into different phases of the AI life cycle, which includes system design, data acquisition, modeling, testing, deployment, monitoring, and system retirement. The early identification of the risk of harm enables organizations to put necessary measures in place before the deployment of the systems. Monitoring continuously helps in the detection of new risks emanating from changes in data, system behavior, interaction, and deployment environment. Therefore, responsible AI governance should be seen as a continuous process (Gao et al., 2025).

Social Equity and Public Trust in AI Systems

Social equity in AI refers to ensuring that AI-enabled systems provide fair and non-discriminatory treatment to individuals and different social groups. If AI systems are developed on the basis of incomplete, partial, or biased information, they may reproduce or intensify existing inequalities. Such cases of AI development may be related to the issues of access to employment, education, financial services, healthcare, public services, etc. (Paglieri et al., 2025). The issue of trust is equally important in this regard, since people will be more likely to accept AI if they consider it to be fair, transparent, secure, and accountable (Robles & Mallinson, 2025). Transparency, explainability, privacy protections, human oversight, and stakeholder participation can increase such trust in AI decision-making (Tsarouhas & Grigoriadis, 2025).

Another aspect which needs to be considered in case of social equity is related to whether there are similar opportunities for different individuals to receive benefits from AI technology. The digital divide, technological literacy, and socio-economic and institutional factors can impact this matter. Thus, governance needs to pay attention not only to the consistency of results produced by algorithms but also to their implementation in particular cases (Bittencourt et al., 2024). People's trust in people can be influenced in cases where people feel that there is unpredictability, opacity, intrusion, and challenges in AI systems. Effective communication on the nature and limitations of AI systems can enable people to appreciate what these technological systems do and how they work. Effective complaints and redress mechanisms can also provide people with an avenue to air their grievances. This way, the link between technology and the community can be strengthened (Buhmann & Fieseler, 2023).

Research Problem, Objectives, and Significance

The fast-paced implementation of AI poses a problem in terms of governance because there is always a possibility that technological developments will occur at a speed faster than what organizations and institutions are able to cope with. The absence of adequate governance can lead to such problems as discriminatory decision-making, lack of accountability, violation of privacy, and no chance for the people affected by the decisions to dispute them. Thus, the study concentrated on developing a responsible AI governance model that combined ethics and social justice. The objectives of the study were to identify important aspects of governance, to analyze methods of ensuring fairness and accountability, and to integrate them into one framework along with the aspects of transparency, privacy, human involvement, and participation. In addition, the research took into account the necessity of developing governance

mechanisms that would work in the course of the AI life cycle. In contrast to the previous approaches that separated ethical considerations from technological innovations, the suggested framework related ethical and social aspects to the design of the system, its evaluation, deployment, and monitoring. Thus, it allowed recognizing the risk factors and taking measures in response to their occurrence at various stages.

Moreover, the study aimed at showing the possibility of integrating the governance dimensions into one framework instead of realizing them separately. The framework could become a basis for systematic identification of risks, establishment of the safeguards, allocation of responsibilities, stakeholder engagement, and governance assessment. Therefore, the research helped to understand the ways in which the governance of AI could ensure social justice and public trust.

Literature Review

Responsible AI and AI Governance

The concept of responsible AI is being used as one of the frameworks for ensuring that AI systems function in accordance with ethical, legal, and societal expectations. Fairness, transparency, accountability, privacy, safety, security, and human involvement are some of the common principles that are associated with responsible AI (Mahl et al., 2025; Kannike & Fahm, 2025). Governance of AI expands on such principles by outlining the organizational and institutional procedures for setting out responsibilities, decision-making procedures, risk management controls, monitoring procedures, and other aspects of accountability. In order for the governance framework to be effective, it must address all stages of the AI lifecycle from design, data preparation, development, testing, deployment, monitoring, and eventually retirement.

The governance of responsible AI should also encompass setting up the procedures for identification and mitigation of risks prior to their impact on society. The governance framework may be incorporated into the development standards, documentation procedures, risk assessments, auditing processes, and monitoring procedures. There could be several organizational roles involved in governance, so coordination becomes essential to ensure implementation of ethical considerations. It is also crucial to take into account the interplay of technical effectiveness and responsible outcomes. The technical effectiveness of the system could be achieved but, at the same time, raise issues related to fairness, privacy, accountability, or accessibility. Thus, responsible AI governance needs a multi-dimensional assessment that takes both technical and social factors into account.

AI, Social Equity, and Inclusive Decision-Making

AI-powered decision-making can affect people belonging to different demographic, economic, and social categories. If the training data used for an automated system contains inherent inequalities, then the system itself might be prone to discriminatory behavior. As such, socially responsible AI must consider issues related to the representativeness of the data, evaluation of unfairness, accessibility, and inclusive system design. Moreover, when making decisions, an inclusive approach must take into account the views of all stakeholders potentially affected by these AI systems. It is widely known that fairness is not a problem that can be solved merely from the technical side; the social context and decision-making processes can play a role (Chan & Nurrosyidah, 2025).

In order to make an AI governance system inclusive, one must take into account who takes part in the development and evaluation of AI algorithms. These stakeholders can include developers, organizational decision-makers, system users, affected individuals, domain experts, and others (Ter-Minassian, 2025). There is also an aspect of social equity, which requires an analysis of how the benefits and risks of AI technology are distributed. While on the one hand, AI can help to enhance the effectiveness of service provision and decision-making processes, on the other hand, there might be a development of new types of exclusions if the implementation is not handled properly.

Transparency, Explainability, and Accountability

The concept of transparency entails providing useful information regarding the process of developing, utilizing, monitoring, and governing AI systems. The concept of explainability, on the other hand, refers to making the outputs generated by AI systems understandable for specific users or stakeholders. Such mechanisms may assist users in recognizing the limits of such systems, raising queries where necessary, and understanding the factors that influence the automated output. Accountability supplements transparency and explainability through the identification of individuals who should be responsible for the development, implementation, monitoring, and taking remedial action of the AI system. Suitable documentation, auditing, impact assessment, and assignment of responsibility may improve accountability. The four mechanisms may assist in improving AI systems by making them more understandable and responsible.

Transparency may occur at various levels. An organization may be required to convey different information regarding the purpose of the AI system, the type of data utilized, the automation element, the limitations of the system, and the process of monitoring the system. The aspect of explainability is especially important when the output of an AI system has an impact on any significant decision made in relation to individuals. It is possible that sufficient data will need to be provided for users to comprehend why such an output has been generated.

Public Trust, Privacy, and Human Oversight

Factors that may affect trust in AI include issues of fairness, reliability, transparency, privacy protection, and accountability. When AI systems operate in critical settings, there is also the concern of the collection and handling of personal data. Privacy measures should then be able to protect not only the collection and storage of personal data but also the access, handling, transfer, and retention of the data. Human control will also serve as another layer of protection by giving the opportunity for qualified people to check on AI outputs and rectify any wrong decisions made by the system. Privacy governance is significant because AI systems can be dependent on a lot of data in order to train, test, and operate. It is necessary to set up proper governance of data procurement, processing, access, and retention. At the same time, it is necessary to ensure compliance of data practices with applicable requirements and the expectations of concerned people.

The process of human supervision can vary based on the purposes and effects of an AI system. It can involve analysis of AI suggestions, approval of decisions, monitoring of AI activity, and intervention in case of unexpected results. To be effective, human supervision should provide human decision-makers with enough rights, information, and knowledge about AI output.

Research Gap

There is extensive existing research on such aspects of responsible AI as responsible AI principles, algorithmic fairness, transparency, privacy, and accountability. Nevertheless, each of these aspects is often considered independently, which raises a need for an integrated approach towards AI governance that would encompass technical, organizational, and social aspects of the matter. There is especially a need for theoretical frameworks that take into account such dimensions as social equity, public trust, stakeholder engagement, human oversight, and ongoing governance.

This research study aims to fill this gap through the development of a governance framework that takes into consideration all these dimensions.

This gap is also connected with the relationship between governance tools and social impact. The use of each governance principle may serve as a safeguard; however, in some cases, the implementation of both of them is necessary in order to mitigate certain risks caused by the use of artificial intelligence technologies. Thus, for example, transparency does not necessarily establish responsibility, and fairness assessments without stakeholder engagement do not take into account community experience. Thus, this research study developed a framework that considers governance as an ongoing and interconnected process.

Responsible AI Governance Framework

Ethical and Social Equity Principles

The proposed framework made ethics and social equity the basis of responsible AI governance. Fairness, anti-discrimination, human dignity, inclusivity, autonomy, and respecting human rights were defined as the necessary principles of responsible governance. AI technologies should be evaluated in terms of their possible influence on diverse demographic groups, as well as accessibility and equity of AI benefits and risks distribution. The mentioned principles would serve as the base for the creation of governance mechanisms that would address not only the effectiveness of AI operation but its possible social implications as well. Ethics should be integrated into the process of AI development from its very beginning and not be addressed after some ethical problems emerge. Organizations can evaluate the social impact of systems in planning stages, data collection and selection, model development, evaluation, and deployment. Such an approach would allow identifying risks before they become a part of organizational operations. The equity principle, moreover, implies that organizations should evaluate whether any design and deployment of AI systems can create barriers for specific groups. Thus, accessibility, representation, and participation can be taken into account when assessing AI systems from the perspective of social equity.

Transparency, Explainability, and Accountability Mechanisms

The approach involved having mechanisms of transparency and explainability within the entire lifecycle of the AI system. Organizations were expected to keep proper documentation about the sources of data used, the objective of

the system, the development process, limitations, and circumstances under which the system is deployed. Explainability mechanisms could be utilized to make it easier for both end-users and impacted individuals to comprehend the outputs of the AI system, especially in situations with potentially serious consequences. Accountability mechanisms include responsibility, auditing, reporting, impact assessment, and remedies.

Documentation could act as one of the means through which organizations could achieve transparency in AI development processes. Documentation would also allow for reviews and audits when necessary. The mechanism of explainability could be modified in accordance with the needs of stakeholders. Accountability also involved the organization having proper mechanisms for dealing with identified problems with the system. This could include having escalation procedures, reviews, and remedy actions to deal with harmful or inappropriate outcomes.

Privacy, Fairness, and Bias Management

Privacy and fairness were addressed as two intertwined aspects of governance. It was necessary to use privacy protection to ensure that personal and sensitive data were safe throughout their lifecycle. Bias management implied the need to analyze training data, model performance, results, and disparities among relevant populations. Regular testing and monitoring could help to detect any discriminatory pattern and facilitate the implementation of necessary measures. Thus, this framework encouraged data assessment, testing for fairness, privacy-by-design, access control, and continuous monitoring instead of one-time testing prior to deployment.

Bias management required taking into account data-related and system-related factors. Data might include some patterns or representation problems that might influence the outcomes of a model. Therefore, the evaluation of the model could consist of analyzing whether there are differences in outputs depending on relevant populations. If any issues occurred, then the organization could analyze data quality, design of a model, deployment process, and decision-making. Privacy and fairness had to be addressed together since data management practices could hinder identification of unequal outcomes.

Human Oversight and Stakeholder Participation

Human oversight was included in the process to make sure that AI systems would not function without adequate review in cases where there were socially significant implications. It was assumed that human decision-makers should have enough powers and information, as well as sufficient knowledge about the AI system, in order to challenge or reverse any suggestions of this system. Stakeholder participation in such circumstances would also be important in terms of knowing their experience and interests. It was also important to define the scope of human responsibilities in such cases, as people who were supposed to review AI-generated suggestions had to know at least what this system is intended for, its limitations, and possible risks. Stakeholder participation could go beyond the initial design phase in case of feedback and complaint mechanisms. Thus, information that is obtained through users and other stakeholders can be used for improving AI governance frameworks.

Integrated Governance Framework for Public Trust

The proposed framework took into account the alignment of ethical considerations, social justice, transparency, privacy, fairness, accountability, human involvement, and stakeholder engagement. The governance process involved identifying the risks and social implications, assessing them, implementing the control measures, deploying, monitoring, and improving them. Public trust was seen as a result of the joint effectiveness of all the governance processes rather than only the transparency process. Continuous monitoring and feedback helped the organization to detect new risks and adjust its governance controls if needed.

Table 1 below highlights the governance dimensions that are essential in the responsible deployment of AI, which include fairness, transparency, accountability, privacy, human control, stakeholder engagement, and monitoring. These dimensions will help in achieving social justice and public trust through maintaining AI systems' fairness, transparency, accountability, inclusiveness, and monitoring.

This approach may also be viewed as a continuous cycle, not a process that follows a certain linear pattern. The need for governance may be determined at the stage of planning, assessed during development, tested prior to deployment, monitored during operation, and adjusted due to emerging threats or stakeholder needs. Such an approach will enable the governance methods to adjust to new developments in the AI systems and their environments. The presence of multiple governance dimensions will allow establishing connections between different technical and organizational processes. For instance, fairness assessments may help determine accountability methods; stakeholder engagement may serve as a foundation for monitoring; and transparency practices may assist in human oversight.

Table 1: Key Governance Dimensions and Their Contributions to Responsible AI

Governance Dimension	Main Focus	Contribution to the Framework
Fairness and Bias Management	Identifying and Reducing Discriminatory or Unequal Outcomes	Supports Social Equity
Transparency and Explainability	Making AI Processes, Outputs, Limitations, And Decisions Understandable	Supports Understanding and Trust
Accountability	Defining Responsibilities, Audits, Reporting, And Corrective Actions	Establishes Responsibility for AI Outcomes
Privacy Protection	Responsible Collection, Processing, Access, Sharing, And Retention of Data	Protects Individual Rights and Information
Human Oversight	Human Review, Intervention, and Override of AI-Supported Decisions	Maintains Meaningful Human Control
Stakeholder Participation	Consultation, feedback, complaints, and involvement of affected groups	Supports Inclusive Governance
Continuous Monitoring	Ongoing Assessment of Ai Systems, Risks, Data, And Deployment Conditions	Enables Continuous Improvement
Social Equity and Public Trust	Fair Access, Reduced Inequality, Confidence in AI Systems	Represents the Intended Social Outcomes

Methodology and Framework Evaluation

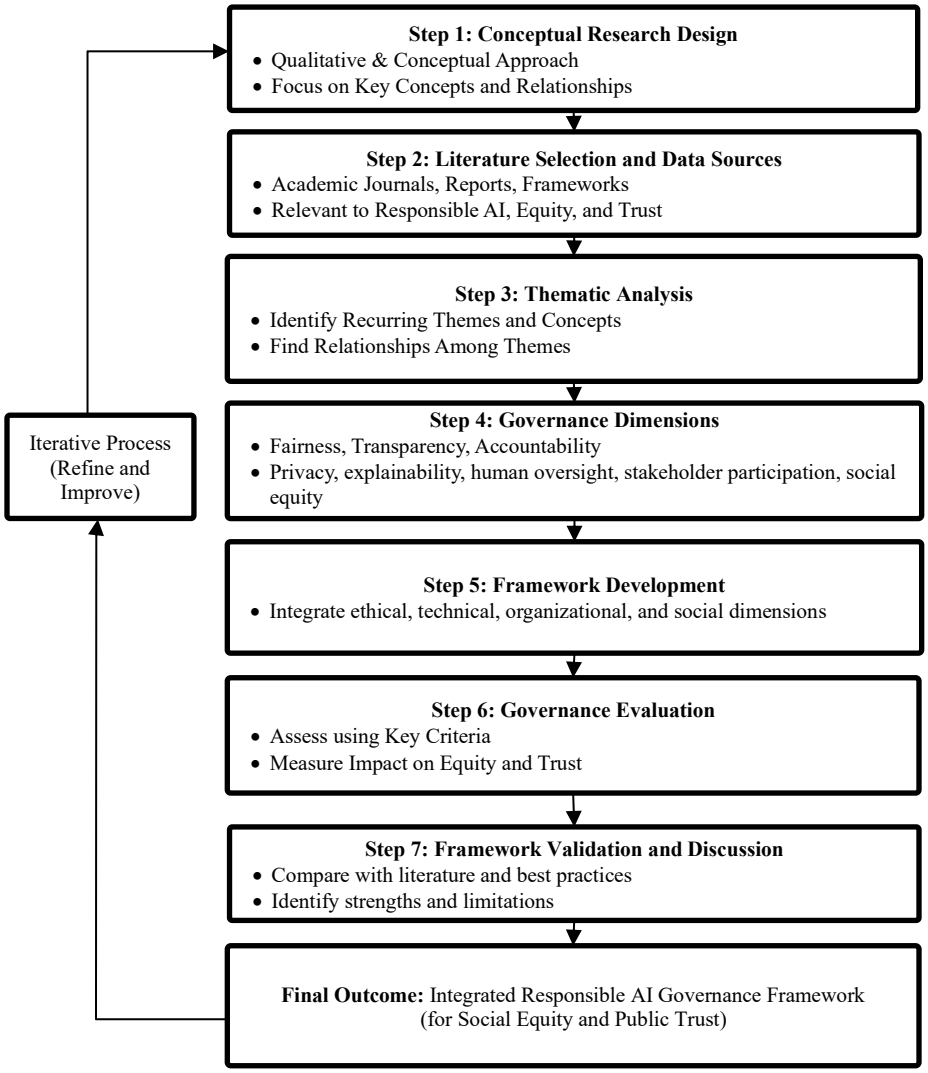


Fig. 1: Methodology Flowchart for Responsible AI Governance for Social Equity and Public Trust

Fig. 1 shows the process methodology starting from conceptual research design and literature selection to thematic analysis, identifying governance dimensions, developing the framework, evaluation, and validation. Through the iterative process, it becomes possible to continuously improve the framework, which leads to the creation of the Responsible AI Governance Framework for Social Equity and Public Trust.

Research Design and Conceptual Approach

The research was carried out through a conceptual research design, which involved a qualitative review and synthesis of the available literature on the themes. This research design was chosen due to the nature of the research topic since the research aimed at establishing relationships between responsible AI governance, social equity, and public trust and not a statistical hypothesis. Key concepts and governance dimensions were generated from the literature and were put together under interconnected categories that were used in constructing the responsible AI governance framework.

This approach enabled the exploration of various dimensions of governance simultaneously. Unlike when looking at one particular aspect of AI technology or application, the research took an overview of general governance issues that could apply to various deployment settings. The process of developing the framework involved generating recurring concepts, exploring their relationships, and grouping them into a governance framework.

This research design also provided a framework for further empirical studies since the framework is conceptual and thus needs to be empirically tested in various organizations.

Selection and Data Sources

The secondary sources relevant to the topic of study have been drawn from academic literature, scholarly databases, institutional reports, governing frameworks, and other reliable sources dealing with issues related to responsible AI and its governance. Literature that dealt with issues of AI ethics, algorithmic fairness, transparency, explainability, accountability, privacy, human oversight, social equality, and public trust was given preference. The sources selected were based on their relevance and their contribution to the understanding of the governance needs. The selection of sources for analysis was based on their contribution to understanding the issues related to governance of responsible AI and the link between the two. The sources were assessed based on their conceptual relevance and contribution towards one or more of the dimensions of governance. The selected body of literature was used to find the recurring themes and relationships rather than calculating any statistical effect.

Thematic Analysis and Framework Development

The literature selected was analyzed using thematic analysis in order to identify any recurring themes and patterns within the literature. Some of the initial themes identified include responsible AI, fairness, transparency, accountability, privacy, explainability, human oversight, stakeholder participation, social equity, and public trust. Concepts related to each theme were put into broader governance dimensions, and their interactions were analyzed in order to understand complementary and overlapping needs. This analysis involved the identification of individual concepts and how these concepts relate to each other. For instance, transparency was analyzed in relation to explainability and accountability, and fairness was analyzed in relation to data quality and bias mitigation. Human oversight and stakeholder participation were seen as tools for keeping the human factor alive. After thematic synthesis, the governance dimensions identified were incorporated into a framework. The framework development process hence started with individual concepts, progressed to governance dimensions, and ended with the interaction of governance mechanisms.

Governance Evaluation Criteria

The framework was assessed conceptually against criteria that pertained to fairness, transparency, explainability, accountability, privacy, inclusiveness, human oversight, stakeholder participation, and continuous monitoring. Fairness assessed whether AI systems would generate discriminatory or unequal results. Transparency and explainability assessed whether information regarding AI decisions could be shared effectively. Accountability assessed whether there would be defined responsibilities and means for redress. Privacy involved responsible handling of data, whereas human oversight involved the capacity of qualified personnel to oversee and intervene in AI-driven decisions. These evaluation criteria were developed to cover different aspects of technology, organization, and society. Fairness and privacy looked at the impact on people, whereas transparency and explainability looked at the understanding of AI-driven decisions. Accountability and human oversight looked at organizational responsibility and control. Inclusiveness and stakeholder participation went further than internal organizational procedures to look at the engagement of impacted groups. Continuous monitoring was included in order to take into consideration that governance effectiveness can vary with changing AI technologies, data, and environment.

Framework Validation and Discussion

Framework validation was done on the basis of conceptual comparison with the governance issues discovered in the literature review. The proposed dimensions were assessed in terms of compatibility with responsible AI standards and social equity and trust. Validation showed that a combination of technical, organizational, and social controls is more efficient than considering individual responsible AI principles separately. It was also noted that constant monitoring of risks is needed due to the dynamic nature of data, models, AI users, and environments. Thus, the framework provides a possibility for further empirical investigation and application in organizations. Moreover, validation of the framework was based on an assessment of whether the framework takes into account the governance issues identified during the literature review. The framework addresses the topics of fairness, transparency, privacy, accountability, human oversight, and stakeholder participation as a way of connecting responsible AI principles and social consequences. Nevertheless, conceptual validation does not prove the empirical efficiency of the framework in various organizations and domains. This limitation creates the basis for future empirical investigation.

Key Findings and Contributions

It has been noted that responsible governance of AI requires a comprehensive perspective that takes care of both the technical issues and the wider social implications. The key aspects identified in the literature include fairness, transparency, explainability, accountability, privacy, human oversight, and stakeholder engagement as dimensions of responsible AI governance. The framework provided the connection between these dimensions and social equity and public trust, offering a systematic approach to responsible AI deployment. Thus, the major contribution is the establishment of a connection between these dimensions within the same framework.

The framework highlighted the need to connect governance principles to the actual tools used to implement them in practice, such as impact assessments, documentation, monitoring, audits, stakeholder engagement, and corrective measures. Also, responsible AI governance should be continuous and not limited to the development process.

The second contribution is the connection of responsible AI governance and public trust. The framework took into account that trust is a multifactorial concept that includes, among other things, fairness, transparency, privacy, accountability, and human and stakeholder involvement.

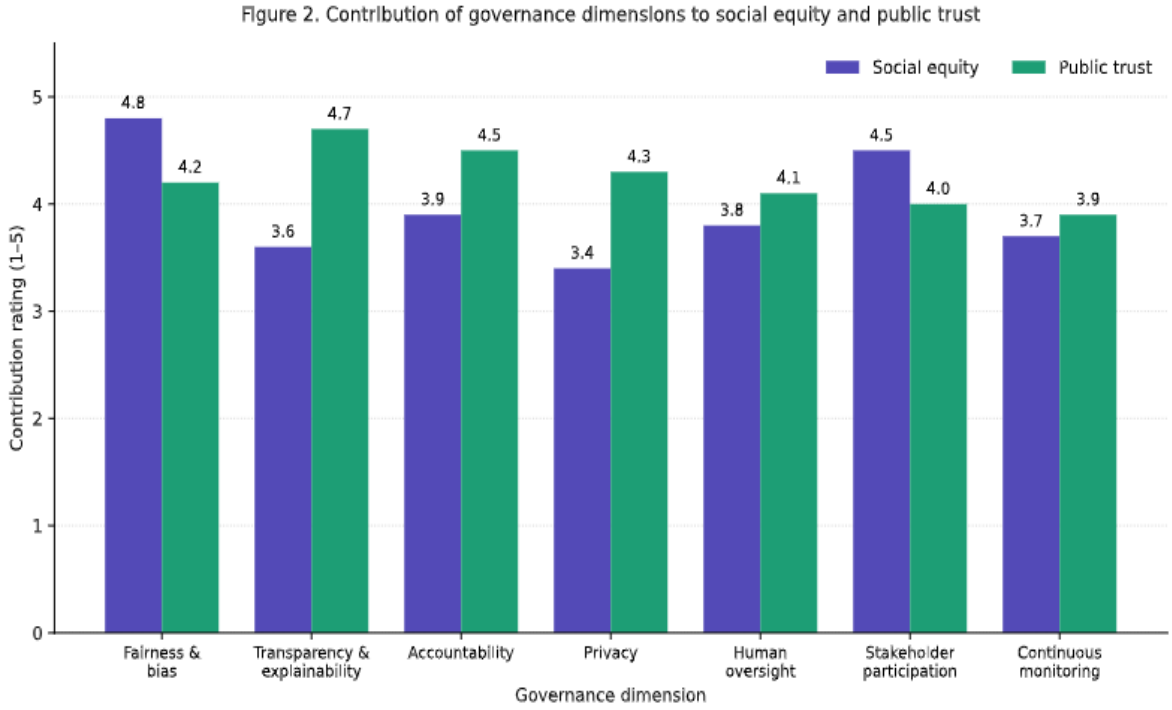


Fig. 2: Contribution of governance dimensions to social equity and public trust

Fig. 2 provides the comparative contributions of the essential governance dimensions to the realization of social equity and trust on a rating scale of 1-5. This shows that all of these contribute towards the achievement of both intended results.

Implications for Responsible AI Governance

The framework had implications for organizations, tech developers, governments, and other stakeholders working with AI technologies. Organizations would benefit from using the proposed dimensions to develop governance policies, assess the impact of AI, monitor the performance of systems, and assign accountability roles. Increased participation of stakeholders could also assist organizations in recognizing the social aspects of concern that cannot be addressed via technical assessment only. The framework suggests that responsible AI governance requires cooperation between technical staff, organizational management, compliance, the legal department, and stakeholders.

Also, organizations would have an opportunity to embed governance requirements into AI development and operation procedures. Clearly defined responsibilities could ensure that people responsible for monitoring the systems, analyzing the risks, handling complaints, and making decisions on major changes will be identified. Assessment could help in identifying new risks and adjusting governance frameworks in time. For stakeholders and communities affected by AI, mechanisms such as accessible information and feedback could serve as the means of participating in AI governance. These mechanisms would allow improving communications between AI developers, organizations, users, and stakeholders.

Challenges and Future Research Directions

Other issues in governing the development and use of AI included the difficulty of measuring fairness, explaining sophisticated AI models, balancing privacy and data usability, the assignment of responsibility and accountability, as well as the importance of human oversight. Different uses and contexts of AI could also necessitate distinct types of governance control. Therefore, future research could empirically test the framework in sectors such as health care, finance, education, work, and public administration. Future research could also explore developing governance indicators, conducting stakeholder analysis and evaluation, and assessing the effectiveness of continuous monitoring and accountability. Future research could also explore how organizations could operationalize the framework by way of governance maturity models, assessment instruments, auditing, and performance measures. Comparative research across various sectors could help discover what types of governance mechanisms would need to be modified in light of the specific sensitivity and impact of particular applications of AI. Other research could analyze how stakeholder involvement and human oversight could be measured in practice. Empirical studies could also analyze how the evolution of data, models, organizational processes, and environments could affect governance requirements over time.

Conclusion

A framework for the Responsible AI Governance Framework for Social Equity and Public Trust was constructed by identifying the interrelationship between the ethical, technical, organizational, and social aspects of AI governance. The conceptual review revealed that the key AI governance dimensions were fairness and bias management, transparency and explainability, accountability, privacy, human oversight, stakeholder participation, and monitoring. The identified AI governance dimensions formed a holistic basis for considering social and institutional aspects of AI usage. It was shown that responsible AI governance goes far beyond the technical functioning and needs to take into account the impact that AI has on people, societies, and social groups. Inclusive practices and fairness may contribute to social equity, whereas transparency, explainability, privacy, accountability, human oversight, and stakeholder participation may help to build public trust in AI technologies. The research indicated that all governance mechanisms have an interrelationship and should not be applied in isolation. These results also showed that AI governance should be seen as an ongoing process within the whole lifecycle of AI. Risk assessment, monitoring, auditing, feedback, and correction in light of changes in data, models, technologies, and deployment settings can be helpful for responding to changes in the context. Appropriate allocation of responsibilities and human involvement can be vital for making sure that decisions assisted by the use of AI will remain under appropriate control and supervision. Thus, this proposed framework is conceptual and serves as a basis for organizations, policy makers, technology developers, and other interested parties who aim at increasing the responsible AI governance and improving social equity and public trust related to AI. Nevertheless, since the framework was conceptually derived from the secondary literature, its practical efficiency needs to be empirically tested.

References

Bittencourt, J. C. N., Costa, D. G., Portugal, P., & Vasques, F. (2024). Smart cities for equitable societies: promoting digital equality, trust, and ethical governance. *2024 IEEE Smart Cities Futures Summit (SCFC)*, 37-40. https://doi.org/10.1109/SCFC62024.2024.10698207

Buhmann, A., & Fieseler, C. (2023). Deep learning meets deep democracy: Deliberative governance and responsible innovation in artificial intelligence. *Business Ethics Quarterly*, *33*(1), 146-179. https://doi.org/10.1017/beq.2021.42

Camilleri, M. A. (2024). Artificial intelligence governance: Ethical considerations and implications for social responsibility. *Expert systems*, *41*(7), e13406. https://doi.org/10.1111/exsy.13406

Chan, C., & Nurrosyidah, A. (2025). Democratizing artificial intelligence for social good: A bibliometric–systematic review through a social science lens. *Social Sciences*, *14*(1), 30. https://doi.org/10.3390/socsci14010030

Cojocaru, A. (2025). Governing AI with trust: An adaptive framework for institutional legitimacy in the UK public sector. *Transforming Government: People, Process and Policy*, 1-22. https://doi.org/10.1108/TG-05-2025-0125

Gao, Q., Chen, L., & Huang, Z. (2025). Opportunities and challenges of artificial intelligence in public health: a systematic review on technological efficacy, ethical dilemmas, and governance pathways. *Frontiers in Public Health*, *13*, 1748797. https://doi.org/10.3389/fpubh.2025.1748797

Kannike, U. M. M., & Fahm, A. O. (2025). Exploring the ethical governance of artificial intelligence from an Islamic ethical perspective. *Jurnal Fiqh*, *22*(1), 134-161. https://doi.org/10.22452/fiqh.vol22no1.5

Mahl, D., Schäfer, M. S., Voinea, S. A., Adib, K., Duncan, B., Salvi, C., & Novillo-Ortiz, D. (2025). Responsible artificial intelligence in public health: a Delphi study on risk communication, community engagement and infodemic management. *BMJ global health*, *10*(5), e018545.

Mura, L., & Stehlíková, B. (2025). The ethics of artificial intelligence: Safeguarding human dignity, social justice and environmental stability in the age of AI. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *20*(2), 479-507. https://doi.org/10.24136/eq.3743

Paglieri, L., Costumato, L., & Bonomi Savignon, A. (2025). Does artificial intelligence improve social equity in public services? A systematic literature review on opportunities and pitfalls. *International Journal of Public Sector Management*, 1-23. https://doi.org/10.1108/IJPSM-10-2024-0333

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

Sadanand, V. K., & Sadanand, T. (2025). Artificial intelligence in India: Navigating social justice, environmental ethics, and legal governance. In *Exploring AI Implications on Law, Governance, and Industry* (pp. 85-116). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-3384-7.ch004

Ter-Minassian, L. (2025). Democratizing AI governance: Balancing expertise and public participation. *arXiv preprint arXiv:2502.08651*. https://doi.org/10.48550/arXiv.2502.08651

Tsarouhas, P., & Grigoriadis, K. (2025, May). Building trust in AI for public administration: A strategic framework for transparency, XAI, participation, and digital literacy. In *2025 7th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (ICHORA)* (pp. 1-9). IEEE. https://doi.org/10.1109/ICHORA65333.2025.11017116

Winfield, A. F., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, *376*(2133), 20180085. https://doi.org/10.1098/rsta.2018.0085