

Collaborative Governance and Algorithmic Accountability: Regulating Risks and Restructuring Organizations for AI Integration

Xiaoran Ma*

School of Liberal Arts, Nanjing University, Nanjing, 210023, China

**Corresponding author: Xiaoran Ma.*

Abstract

The widespread application of artificial intelligence (AI) in corporate resource planning and public decision-making has provided impetus for improving management efficiency and creating social value. However, the complex structure and opacity of algorithms have led to a crisis of trust, posing challenges to traditional public management accountability mechanisms. Drawing on socio-technical systems theory, this paper provides a normative analysis of thirteen recent studies on the challenges of technology implementation, ethical trust, and legal regulation. The findings suggest that the current governance dilemma stems not only from technological limitations but also from institutional neglect, which enables accountability avoidance. Although the EU AI Act proposes a preliminary form of collaborative governance, it still has shortcomings in terms of procedural justice and the feasibility of human oversight. The governance logic should shift from individual oversight to an organization-in-the-loop approach, to achieve sustainable and responsible AI governance through the construction of a procedural justice framework.

Keywords

artificial intelligence, collaborative governance, algorithmic accountability, EU AI act, socio-technical systems, procedural justice

1. Introduction

Artificial intelligence is being widely applied to various organizational activities, becoming a core element driving digital transformation. Through automated processes and stronger data insight capabilities, AI helps enterprises and entrepreneurs create significant business and social value [1, 2]. This technology-driven management change not only improves the efficiency of administrative output but also reshapes the underlying logic of organizational operation. However, the leap in efficiency does not mean that risk governance capabilities grow in tandem.

In fact, the root cause of governance lag lies in the complexity of algorithmic systems and the lack of transparency in their construction, thus constituting a typical black box dilemma. Traditional public management theory is based on the traceability of administrative actions and requires clear attribution of

responsibility. However, when algorithms are deeply involved in the decision-making process, the output of the system is seriously disconnected from the existing administrative accountability mechanism. Since the derivation process of machines is difficult for humans to perceive directly, organizations often adopt a “systemic neglect” strategy, attributing errors to algorithms and codes to cover up management loopholes and evade the consequences they should bear [3]. Such behavior greatly weakens public trust and shakes the legal basis of administrative legitimacy, causing algorithms to even evolve into a means of evading social supervision in some contexts [4].

In response to the aforementioned governance oversights, the EU introduced policies such as the AI Act to attempt to classify and manage algorithms based on their risk levels, clarify how high-risk algorithms are trained and supervised, and make them as visible as possible to provide the possibility for external accountability in the future [5]. However, existing tools still seem inadequate when dealing with the multidimensional value conflicts of fairness, transparency, and efficiency, especially for resource-constrained SMEs, which still face numerous obstacles in fulfilling compliance obligations in complex application environments [6]. At the same time, administrative agencies transfer some powers to algorithms, but due to the limitations of legal regulation and technical supervision, when algorithms systematically malfunction, the intervention and checks and balances in human-computer interaction are still insufficient [7, 8]. If the governance system cannot delve into the underlying operating logic of algorithms, the fairness of public decision-making will be gradually eroded by technological forces.

This study, based on the perspective of public management, examines the organizational risks of artificial intelligence applications from a global perspective and introduces the socio-technical systems (STS) theory to explore how to embed accountability mechanisms into the entire lifecycle of algorithm applications through collaborative governance and organizational restructuring [9, 10]. Meanwhile, this study aims to demonstrate that the governance logic should shift from relying solely on individual supervision to incorporating supervision into organizational system design (i.e., an organization-in-the-loop approach), and through the framework of procedural justice, ensure the openness and transparency of the decision-making process as much as possible [11, 12]. This is related to the maintenance of administrative efficiency and is a necessary way to rebuild the credibility of administrative accountability in the era of automation. This paper hopes to provide theoretical support and policy recommendations for building a responsible and sustainable governance system, and to ensure that technological progress continuously enhances the public interest [13].

2. Theoretical Foundation

The socio-technical systems (STS) theory argues that technological application is not an isolated tool change, but rather a mutual shaping of technological structure and social organizational dimensions. The effectiveness of the system depends on the synergy of four dimensions: motivation, technology, human interaction, and organizational drive [10]. In the field of AI governance, this theory extends the traditional individual oversight perspective to an organization-in-the-loop architecture, emphasizing the need to elevate the responsibility system from the individual level to the organizational institutional level by designing scientific workflows and auditing procedures, thereby ensuring that algorithmic agency is effectively checked and balanced by the collective management mechanism [12].

In response to the accountability needs arising from system interactions, the theory of algorithmic accountability and organizational responsibility further clarifies the path of migrating management responsibility to the organizational institutional level. This theory emphasizes elevating management responsibility to the institutional level, thereby touching upon the core issue of power regulation in the digital age. This reconstruction shifts responsibility from vulnerable individual gatekeepers to robust organizational institutional levels, ensuring that algorithmic agency is effectively checked and balanced by the collective management mechanism through the design of scientific workflows and auditing procedures. When management responsibility is established at the institutional level, organizations can overcome the black box dilemma of decision-making caused by technological complexity, thereby providing legitimate institutional support for the operation of digital technology in the public domain.

To resolve the decision-making game faced by organizations when assuming responsibility, the procedural justice framework provides normative guidelines for establishing a legitimate foundation for governance. This

framework does not pursue a single optimal outcome, but rather ensures a normative balance between values such as efficiency and fairness in the application of technology by establishing open, inclusive, and relevant decision-making processes with appeal channels [11]. Through open procedures, administrative agencies can rebuild public trust in algorithmic administration, ultimately achieving a closed loop from theoretical principles to management practice, provided that the technological logic cannot be completely dismantled.

Collaborative governance theory constitutes the mechanism for realizing procedural justice from principle to action, emphasizing cross-sectoral collaboration between the government, the private sector, and social organizations. Faced with governance loopholes caused by algorithmic systems, administrative agencies establish specialized agencies and introduce multiple stakeholders to build a defense system based on risk assessment [5]. This mechanism aims to transform the legal principles of digital constitutionalism into practical operation, enhancing the adaptability of rule-making through the participation of multiple stakeholders, and providing dynamic support for administrative decision-making in complex environments.

The structure of digital constitutionalism examines the regulation of digital power from a more macro-level social and legal perspective, providing a fundamental legal foundation for collaborative governance. When algorithmic power intervenes in public decision-making, traditional state-centered constitutional theories show limitations in dealing with decentralized digital power. The social constitutional perspective advocates addressing the potential infringement of fundamental rights by technology through deconstructing and reconstructing traditional legal principles [9]. This requires extending the constraint mechanism of administrative power into the internal logic of algorithms to ensure that the operation of digital technology in the public domain complies with the rule of law, thereby maintaining fairness at the social and legal level.

3. Literature Review

3.1 Research on AI Governance Efficiency and Value Creation

The core of academic discussions on AI-enabled management lies in revealing the evolutionary logic of technology-driven organizational value generation. Research has found that AI has triggered multi-dimensional transformative effects within organizations. The first-order effect achieves improved administrative efficiency and reduced operating costs through automated processes, while the second-order effect induces a redistribution of power within organizations through a profound reshaping of knowledge management and decision-making logic [1]. This transformation has fundamentally shifted the management paradigm, and organizations are gradually evolving from a simple stage of automated task execution to a higher-level governance stage dominated by algorithm-driven decision support and demand forecasting.

In the field of entrepreneurial management research, artificial intelligence is regarded as a key enabling tool that significantly reduces market entry barriers through technological spillover effects. AI has the ability to transform massive amounts of unstructured data into precise actionable insights, which effectively compensates for the inherent lack of industry experience and material resources in startups, thereby helping them optimize their decision-making paths [2]. This technological dividend has completely changed the way entrepreneurs identify market opportunities, making rational analysis based on objective evidence gradually replace the traditional intuition-driven model. However, as algorithms become deeply involved in the management center, how to ensure that the decision-making process conforms to the norms of fairness and justice while obtaining technological efficiency has become a key issue in establishing a legitimate foundation for governance [11].

3.2 Research on the Algorithm Black Box and Accountability Dilemma

The diffusion and application of artificial intelligence show significant imbalance at the public policy level, which directly restricts the governance collaboration among organizations of different sizes. For small and medium-sized enterprises with limited resources, the application of algorithms faces 27 specific challenges, such as capital shortage, lack of professional talent, and high compliance costs [6]. This state of resource scarcity directly restricts the endogenous growth of organizational governance capabilities, making them often passive in the face of complex regulatory requirements and technological iterations. From the perspective of public regulation, the uneven state of technology diffusion means that a single policy template is difficult to

cover the diversified organizational needs, which leads to a further widening of the governance gap in the micro and small enterprise group.

The actual effectiveness of algorithms is highly dependent on how well the internal ecosystem of the organization adapts to the interaction between technological structures and social organizational dimensions. Hardware upgrades and algorithm introductions only constitute the basic elements of technological evolution. Without a matching knowledge management system and institutional support, organizations will find it difficult to effectively digest the data insights brought by technology [10]. Studies have shown that the interaction patterns among people deeply affect the transmission path of decision-making information, while the organizational driving mechanism provides the necessary legal guarantee for the operation of technology. If the organic combination of these dimensions is ignored, even if a high-precision algorithm system is introduced, the organization will often fall into efficiency stagnation due to the rejection reaction of the internal mechanism, and even induce chaos in the management process. This interactive attribute determines that the governance system must go beyond a purely instrumental perspective and turn to the overall coordination and reconstruction of the organization's internal environment [12].

3.3 Review of Research on Collaborative Governance and Regulatory Tools

Ethical biases in algorithmic systems are a direct cause of organizational trust collapse, and their empirical impact often transcends the scope of traditional crisis management. When algorithms exhibit bias or errors in decision-making, the damage to organizational trust is clearly irreversible. Unlike psychological repair in interpersonal communication, traditional crisis public relations methods such as public apologies are often ineffective in dealing with algorithmic errors [4]. This repair dilemma stems from the public's deep understanding of the impersonal nature of algorithmic decision-making. Simple statements of attitude are insufficient to effectively offset the systemic damage caused by technological logic, leading to the shaking of the foundation of social trust.

In fact, the underlying logic of this trust crisis lies in the construction and maintenance of an “institutional black box.” When faced with technical failures or decision-making biases, public institutions often tend to adopt a collective neglect strategy, emphasizing the complexity and non-intervention of algorithms to evade administrative accountability [3]. This behavior of using technological barriers to cover up management loopholes leaves damaged public interests in a vacuum where no one takes responsibility for a long time, thereby shaking the foundation of administrative procedural fairness. When the accountability mechanism is dissolved by the black box of technology, the application of algorithms may evolve into an invisible means of circumventing social supervision, thereby triggering broader governance risks and social injustices [8]. This phenomenon requires that the research on governance tools must shift its focus from technical parameters to the institutional arrangements behind algorithmic decision-making in order to address the crisis of administrative legitimacy caused by the breakdown of accountability.

4. Analysis & Discussion

4.1 Regulatory Tools: Collaborative Governance Framework and the EU AI Act

The EU's AI Act of 2025 established a collaborative governance organizational structure, building a cross-departmental, multi-level regulatory network by setting up the EU AI Office, Advisory Forum, and Scientific Group. These institutional setups aim to integrate the decision-making power of technical experts, industry representatives, and regulatory agencies, and alleviate the lack of algorithm expertise in administrative agencies [5]. By establishing a normalized communication mechanism, the collaborative governance model enhances the dynamic adaptability of rule-making and provides institutional support for addressing governance loopholes brought about by technological evolution.

However, risk-based static classification is limited to focusing more on large-scale, high-risk incidents, while neglecting the protection of basic individual rights. This framework mainly pre-determines risk levels based on application scenarios, and this static classification is difficult to cover the emergent risks of algorithms in complex social environments [8]. When a technology application is classified as low-risk but actually causes cumulative damage, the existing classification regulation often shows its ineffectiveness. In addition, the Act

is still insufficient in identifying and dealing with systemic risks, and the regulatory focus is mostly on the compliance performance of individual systems, ignoring the global impact on market fairness and social ethics when multiple systems operate collaboratively [9]. This logical paradox means that the goal of protecting fundamental rights may be diluted by technical parameters in practice.

4.2 Accountability Mechanism: From Procedural Justice to Organizational Loop

The governance of artificial intelligence often faces a value game between accuracy and transparency. When the algorithmic structure that pursues extremely high prediction accuracy becomes increasingly complex, making it difficult to fully disclose its operational logic to the public, the procedural justice framework provides a legitimate basis for resolving such normative conflicts. This framework argues that the core of governance is not to pursue absolute technological disclosure, but to establish a decision-making process that includes democratic consultation and open information [11]. By allowing stakeholders to participate in setting algorithmic guidelines and providing clear appeal channels after the decision is made, administrative agencies can ensure that management behavior has a legitimate basis, thereby maintaining public trust, provided that the technical black box cannot be completely dismantled.

In highly complex algorithmic application scenarios, the traditional human-in-the-loop model has reached the limits of human cognition. Individual supervisors often become passive recipients of algorithmic decisions in practice because they cannot understand the deep code logic, causing the administrative accountability mechanism to become a mere formality [8].

Therefore, the governance logic must be reconstructed to an organization-in-the-loop approach. This model requires deeply embedding accountability requirements into the overall decision-making and workflow of the organization, and achieving checks and balances at the collective level through institutionalized auditing, cross-departmental verification, and dynamic risk monitoring [2]. This restructuring shifts responsibility from vulnerable individual gatekeepers to robust organizational systems, which not only alleviates regulatory fatigue in technology applications, but also ensures that the application of artificial intelligence complies with ethical standards for sustainable governance through full-cycle supervision [13].

4.3 Sustainable Governance Perspective

From a macro-social and legal perspective, AI governance has transcended the jurisdiction of a single sovereign state. Golia points out in her article that digital constitutionalism advocates establishing global collaborative norms to address the potential erosion of individual rights by transnational algorithmic power [9]. This framework establishes digital justice as the fundamental logic of global governance, preventing regulatory arbitrage through the construction of unified technological ethical norms. It is certain that establishing a global algorithmic trust system requires transcending geographical boundaries, reaching a legal consensus in multilateral agreements, and ensuring that digital power is rigorously constrained by the core values of human civilization.

The concept of sustainable AI requires expanding the evaluation criteria of public management from purely administrative efficiency to the dimensions of environmental integrity and social justice [13]. Therefore, the governance system needs to incorporate the energy consumption costs of algorithm operation, the carbon footprint of data centers, and the long-term impact of technological applications on social structures into the assessment indicators [14], aiming to ensure that digital transformation does not come at the expense of natural resources or exacerbate social inequality. By incorporating sustainable principles throughout the entire algorithm lifecycle, public institutions can simultaneously safeguard technological advancement and public interests, thereby establishing a resilient and enduring governance ecosystem.

5. Conclusion

This study, from the perspective of public administration, demonstrates that the core of algorithm governance lies in building a transparent organizational structure and reshaping the accountability system that has deviated due to technological intervention. The main value of this study lies in proposing the organization-in-the-loop governance paradigm, advocating that administrative agencies should regard algorithms as an

organic component of the organizational system, rather than an independent tool detached from supervision. By establishing a robust responsibility allocation mechanism, organizations can effectively overcome the black box dilemma of decision-making caused by technological complexity, ensuring that the efficiency dividends brought by technological empowerment do not come at the expense of administrative legitimacy [3]. This paradigm shift fills the institutional gap in the application of algorithmic administration in socio-technical systems theory, providing new theoretical support and practical guidelines for power regulation in the digital age.

True algorithmic justice should be based on procedural fairness. The government should use the rule of law to deprive organizations of their right to use technological complexity as a cover for black box defense, and enforce the obligation to explain algorithmic logic. By establishing standardized open processes and democratic consultation mechanisms, public regulation can ensure that every decision in the process of technology application has a traceable chain of evidence, thereby protecting the basic rights of citizens in an uncertain environment [11]. The focus of policy design should shift from simple result review to full-cycle process monitoring, and provide a legal basis for the operation of algorithmic administration and a foundation of public trust through institutionalized constraints.

Although this paper constructs a normative framework for algorithmic governance, the research is limited by the distribution of existing empirical cases, mainly focusing on theoretical deduction and lacking in-depth research on specific industries, and it is still insufficient in the analysis of legal conflicts in cross-border regulations. In the context of the popularization of generative artificial intelligence and general artificial intelligence, the public management system must have agile response capabilities and promote the governance paradigm from static compliance inspection to dynamic collaborative governance [5]. Future exploration should focus on establishing real-time monitoring and multi-party collaboration mechanisms to ensure that the governance logic is upgraded synchronously with the evolution of algorithmic models. By establishing sustainable and inclusive regulatory standards, human society can rebuild the authority of administrative accountability in the era of automation, and enable technological progress to continuously improve public welfare [13].

References

- [1] Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial intelligence and business value: A literature review. *Information systems frontiers*, 24(5), 1709-1734.
- [2] Giuggioli, G., & Pellegrini, M. M. (2023). Artificial intelligence as an enabler for entrepreneurs: a systematic literature review and an agenda for future research. *International Journal of Entrepreneurial Behavior & Research*, 29(4), 816-837.
- [3] Kronblad, C., Essén, A., & Mähring, M. (2024). When justice is blind to algorithms: Multilayered blackboxing of algorithmic decision-making in the public sector. *MIS Quarterly*, 48(4), 1637-1662.
- [4] Schelble, B. G., Lopez, J., Textor, C., Zhang, R., McNeese, N. J., Pak, R., & Freeman, G. (2024). Towards ethical AI: Empirically investigating dimensions of AI ethics, trust repair, and performance in human-AI teaming. *Human Factors*, 66(4), 1037-1055.
- [5] Cancela-Outeda, C. (2024). The EU's AI act: A framework for collaborative governance. *Internet of Things*, 27, 101291.
- [6] Oldemeyer, L., Jede, A., & Teuteberg, F. (2025). Investigation of artificial intelligence in SMEs: a systematic review of the state of the art and the main implementation challenges. *Management Review Quarterly*, 75(2), 1185-1227.
- [7] Holzinger, A., Zatloukal, K., & Müller, H. (2025). Is human oversight to AI systems still possible?. *New Biotechnology*, 85, 59-62.
- [8] Kusche, I. (2024). Possible harms of artificial intelligence and the EU AI act: fundamental rights and risk. *Journal of Risk Research*, 1-14.

- [9] Golia Jr, A. (2024). Critique of digital constitutionalism: Deconstruction and reconstruction from a societal perspective. *GlobCon*, 13, 488.
- [10] Thomas, A. (2024). Digitally transforming the organization through knowledge management: a socio-technical system (STS) perspective. *European Journal of Innovation Management*, 27(9), 437-460.
- [11] de Fine Licht, K. (2025). Resolving value conflicts in public AI governance: A procedural justice framework. *Government Information Quarterly*, 42(2), 102033.
- [12] Herrmann, T., & Pfeiffer, S. (2023). Keeping the organization in the loop: a socio-technical extension of human-centered artificial intelligence. *Ai & Society*, 38(4), 1523-1542.
- [13] Van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics*, 1(3), 213-218.
- [14] Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial intelligence and business value: A literature review. *Information systems frontiers*, 24(5), 1709-1734.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

