

Research on Algorithmic Risks and Resilient Blocking Mechanisms of AIGC Embedded in Political and Legal Public Opinion Governance

Peipei Liang

School of Media and Communication, Shandong University of Political Science and Law, Jinan, Shandong, China

Abstract: **Objective:** The integration of AIGC into the governance of political and legal public opinion has given rise to new risks, such as algorithmic bias and deepfake technologies. This study aims to clarify the underlying technical logic and risk mechanisms associated with these developments and to explore resilient governance pathways. **Methods:** Using resilience theory as a framework and combining literature analysis with theoretical deduction, this study constructs a risk transmission model characterized by technological trigger-scenario amplification-social diffusion, and derives a resilient mitigation mechanism across four dimensions-technology, institutions, actors, and ecology. **Results:** The integration of AIGC follows a logic of technological empowerment-scenario adaptation-power restructuring, giving rise to four types of risks-cognitive manipulation, algorithmic bias, information cocooning, and security erosion-which exhibit characteristics of chain-like transmission and phased evolution. Based on these findings, this study establishes a four-dimensional mitigation mechanism encompassing technological resilience, institutional resilience, actor resilience, and ecological resilience. **Conclusion:** The prevention and control of algorithmic risks in political and legal public opinion governance should move beyond rigid, reactive responses toward the proactive construction of resilience, achieving a paradigm shift from rigid stability maintenance to resilient governance, thereby providing an analytical framework and institutional solutions for addressing complex governance crises.

Keywords: AIGC; Public Opinion Governance; Algorithmic Risks; Resilient Mitigation Mechanism; Algorithmic

Governance

1. Introduction

With the remarkable performance in generative artificial intelligence technology, the governance of political and legal public opinion is changing its way from manual judgment and analysis toward governance driven by intelligent means. By utilizing technological frameworks like natural language understanding, profound synthesis, and algorithmic recommender systems, AIGC has embedded in the complete lifecycle of political and legal data, from gathering and analysis to assessment and distribution. This integration incontrovertibly advances governance effectiveness, yet it concurrently presents new risk categories, including algorithmic bias, cognitive manipulation, and deepfake technologies. Since political and legal public opinion holds distinctive value, it sets especially strict requirements on the reliability, controllability, and transparency of algorithmic technologies. Previous studies predominantly centers on the adoption of these technologies or the governance of algorithms, yet it gives insufficient attention to the algorithmic risks systematic identification within the specific settings of the political and legal domain; in addition, it is rare to utilize a resilience viewpoint to examine risk mitigation approaches. Resilience theory highlights a system's competence to adapt, recover, and evolve when faced with outward disturbances, providing a new conceptual framework that transcends traditional, inflexible regulatory approaches [1]. On this basis, this study examines the AIGC integration into the of political and legal public opinion governance as its core research focus, systematically studies the formation mechanisms and propagation logic of algorithmic risks, and builds a multidimensional risk reduction mechanism integrating technological, institutional, stakeholder, and ecological

resilience, with the goal of offering theoretical foundations and practical guidance for the modernizing political and legal public opinion governance in the digital and intelligent technologies age.

2. The Technical Logic and Scenario Reconstruction of Integrating AIGC into Political and Legal Public Opinion Governance

2.1 Technical Embedded Triple Logic

The field of government governance possesses unique characteristics: while it should not reject the adoption of new technologies, it must also approach the risks associated with their deployment with caution [2]. AIGC does not enter the governance of political and legal public opinion all at once; rather, it does so along a progressive path—first empowering the technology, then adapting it to specific scenarios, and finally reshaping the distribution of power. At the level of technology empowerment, AIGC draws on large language models to automate the production of public opinion texts and to carry out semantic analysis. It further employs computer vision technologies for the intelligent recognition of multimodal information, which in practice broadens the reach of public opinion monitoring and accelerating response times. At the level of scenario adaptation, AIGC is adjusted to fit the particular needs of the political and legal sectors. This involves, for instance, building knowledge graphs and analytical models that vary according to the scenario at hand—whether it concerns public opinion around a specific case, around law enforcement, or around policy. At the power restructuring layer of, the AIGC involvement leads a governance authority redistribution; algorithms are no longer simply instrumental tools but have become quasi-actors possessing autonomous decision-making abilities, moving the conventional human-machine binary framework toward a tripartite interactive model including humans, algorithms, and information.

2.2 Four Dimensions of Scenario Refactoring

AIGC is now being used to work context in ways that reach deep into the key scenarios of public opinion governance across the political and legal departments. In the dimension of information collection, AI-driven web crawling and semantic analysis technologies make it

easier to collect public opinion data in real time across multiple platforms. Yet the bias in selection built into these algorithms can smoothly cause critical data to be systematically sidelined. In the layer of content production, AIGC can turn out guiding communications, official response statements, and public opinion briefings on its own. The issue is that content created inside the algorithmic black box goes without effective value assessment, which renders it highly susceptible to ideological drift or misdirected public opinion guidance. In the layer of distribution, personalized recommendation algorithms depend on user profiling to precisely target public opinion. However, this approach further reinforces the information cocoon effect, deepening the fragmentation of online public discourse and the tendency toward group polarization. In the decision-support dimension, AIGC leverages affective computing and trend forecasting to provide intelligent analytical support for political and legal authorities; yet the lack of interpretability in these algorithmic models may, to some extent, undermine the legitimacy and justification underlying public opinion governance decisions.

2.3 The Dual Nature of the Embedding Effect

The deep integration of AIGC into the governance system for public opinion in the political and legal spheres exhibits a distinct “double-edged sword” effect. The positive aspect manifests as a paradigm shift in governance efficacy: the public opinion monitoring model has evolved from post-event retrospective analysis to pre-event early warning, the logic underlying public opinion analysis has shifted from being experience-driven to data-driven, and public opinion response strategies have moved from adopting a uniform messaging approach to implementing precise, tailored responses [3]. The negative aspect, however, lies in how the excessive expansion of technological rationality exerts a structural pressure on the formation of political and legal value rationality: the purported technical neutrality of algorithms carries a clear degree of obfuscation, as these algorithms inherently carry embedded value biases; simultaneously, the unchecked proliferation of instrumental rationality poses a potential erosion of core political and legal values—such as procedural justice and substantive justice—and this risk carries

significant, real-world consequences that cannot be ignored.

3. The Generation Mechanism and Typed Identification of Algorithmic Risks

3.1 Three-dimensional Mechanism of Risk Generation

The emergence of AIGC algorithmic risks within the field of political and legal public opinion governance is the result of the synergistic interaction among three factors: endogenous technological limitations, lagging institutional provision, and an imbalance in stakeholder capabilities.

3.1.1 Technical dimension: The endogeneity and contagion of algorithmic flaws

First, the cumulative amplification of data bias: the training data used in AIGC systems carry historical value-based biases and structural inequalities, which can easily give rise to systematic misjudgments regarding specific groups or types of cases in the context of political and legal public opinion analysis. Second, the inherent limitations of model interpretability: the black-box nature of deep learning models makes it difficult to trace the logical chain underlying public opinion analysis, thereby undermining the intrinsic requirements for transparency and accountability in political and legal decision-making. Third, the asymmetric tension arising from technological iteration: the rapid evolution of AIGC technologies and the relatively slow development of governance frameworks create a “scissors gap” effect, whereby new types of risks often emerge first in regulatory vacuum zones [4].

3.1.2 Institutional dimension: The lagging nature and fragmentation of rule provision

Current algorithmic governance frameworks are predominantly characterized by a post-event response approach—namely, providing passive regulation in response to already emerging risks—while there is a notable lack of forward-looking institutional design for the application of AIGC in the political and legal spheres. Specialized adaptation mechanisms for systems such as algorithmic filing and security assessments within the political and legal domains have not yet been established; meanwhile, foundational institutional arrangements—such as data rights confirmation and liability attribution—remain unaddressed.

3.1.3 Subjective dimension: differentiation of competency structures and insufficient synergy

In the Global Artificial Intelligence Governance Initiative, China clearly articulates the principle of consultation, joint contribution and shared benefits as well as the mission of Intelligence for Good, emphasizing that artificial intelligence should serve the well-being of humanity [5]. A cognitive gap separates the algorithmic literacy of political and legal authorities from the business logic that drives technology enterprises. The public’s ability to grasp of algorithmic risks is weak and exhibits significant intergroup differences; and an action collaborative governance network involving multiple stakeholders has not yet taken shape.

3.2 Fourfold Risk Type Identification

Judged by how they manifest and what drives them, the algorithmic risks that AIGC brings to the governance of political and legal public opinion fall into four categories.

First, the cognitive manipulation risk: AIGC uses deepfake technology to produce fake audio and video that look and sound entirely convincing, and it applies computational propaganda techniques to churn out mass-produce pseudo-grassroots public opinion, thereby enabling precise manipulation of public perception [6]. In the political and legal domain, this risk shows up in several ways: case details get fabricated or distorted, court rulings are given deliberate readings, and enforcement actions are invoked selectively. All these factors undermines public trust in judicial and law enforcement authorities.

Second, the algorithmic bias risk: value-driven biases lie within algorithmic models—in the training data, feature selection, how weights are allocated—and they can lead to systematic misjudgments or selectively suppress public opinion concerning particular domains or groups within political and legal public opinion governance. What defines these risks is how intensely they hide and how hard they are to correct. Quietly, they can erode the social order on which opinions of judicial fairness rest.

Third, the information cocoon risk. Recommendation algorithms and AIGC-driven content generation in tandem to drive users into a highly uniform data environment, one in which different social groups perceive the same legal and political event through a parallel reality framework [7]. This risk deepens divisions in

social consensus and weakens the public communication and value integration functions that public opinion management is intended to perform.

Fourth, the security erosion risk. AIGC technology can be used to malicious purposes—fabricating manufactured social trends or false disaster scenarios, thereby undermining the normal function of legal and political authorities. Technical weaknesses within algorithmic systems may likewise be capitalized, causing the leaking or tampering of sensitive legal and political information. Risks of this nature bear directly on social stability and national security. They constitute a critical threshold that the legal and political public opinion governance must guard with particular.

4. The Algorithmic Risks Evolution and Transmission in the Political and Legal Public Opinion Governance

4.1 Chain-like Mechanism of Risk Dissemination

AIGC algorithmic risks move through the public opinion domain in a sequence: a technological catalyst sets things off, the scenario magnifies them, and social diffusion propagates them further. At the layer of technological trigger, what makes propagation possible is AIGC's capacity for rapid dissemination and high-fidelity generation. The visual deception inherent in deepfake materials, meanwhile, it leaves the public unable to distinguish authentic from fake by intuition alone. At the level of contextual magnification, the public, sensitivity, and authority character of public opinion give algorithmic risks an outsized amplifying impact: for example, a single forged law enforcement video, can trigger a nationwide trust crisis in the space of several hours. At the level of social communication, the networked model of social media and the magnifying role played by opinion leaders at critical nodes, allows risk-related content to spread across domains and cross boundary barriers. As events unfolds, emotional and group polarization are steadily reinforced. Through these circulation mechanisms, deepfake risks can produce a cross-domain, multi-level ripple effect, whose deleterious impact is persistently amplified during dissemination and extends beyond the confines of any separate field [8].

4.2 Phased Features of Risk Evolution

Algorithmic risks in the political and legal public opinion governance evolve in distinct steps. In the latent step, they stay dormant—technical vulnerabilities or information bias, for instance—without having set off any tangible public opinion event, yet they already form the foundational source of structural vulnerability within governance fragility. In the outbreak step, a particular triggering event activates these latent risks; disinformation generated by AIGC rapidly gains dominate public discourse, while algorithmic recommendation systems hasten its viral propagation, driving public sentiment to escalate exponentially. In the deadlock step, rumor and truth become intertwined in a complicated interaction; different accounts of the same context consolidate their own audiences through algorithmic dissemination, leaving public opinion governance caught in a cyclical quandary of factual rectification, renewed doubt, and further clarification. In the decline step, the authoritative information eventual release brings public opinion intensity down; the deterioration of trust wrought by algorithmic risks, however, endures, and may be reactivated in later events.

4.3 The Inherent Logic of Risk Evolution

The algorithmic risks evolution is not merely a technical fact; it derives from the interweaving of technological, capital, and power logic. Technological logic manifests in AIGC's inherent pursuit of maximum productivity and minimum cost. Capital logic shows up in the platform economy's single-minded grab for traffic and attention, with recommendation systems placing commercial logic above public value. Power logic concerns how algorithmic power becomes and reshapes administrative authority—where the principal-agent problem inherent in automated judgment can generate both a blurring of responsibility and power estrangement. The interaction of these three logics poses algorithmic risks beyond the simple scope of technical malfunctions, transforming them into a complicated governance crisis that encompasses civil rights, social equity, and political security.

5. The Construction Pathway for the Resilience Disruption Mechanism

5.1 Theoretical Foundation of Resilience Disruption

The resilience-based prevention mechanism differs from traditional rigid control paradigms; its core lies in leveraging the system's adaptive adjustment and evolutionary learning capabilities to achieve the dynamic absorption and transformation of risks. Drawing upon complex systems theory and risk governance theory, resilience-based risk prevention in the governance of political and legal public opinion should encompass four dimensions: technical resilience emphasizes the robustness and error-correction capability of algorithmic systems; institutional resilience refers to the flexible adaptation and dynamic updating of governance rules; actor resilience focuses on the cognitive enhancement and capability evolution of governance participants; and ecological resilience seeks the synergistic coexistence of technology, institutions, actors, and the environment [9].

5.2 Four-Dimensional Resilience Prevention Mechanism

5.2.1 Technical Resilience

The key of technological resilience relies on the deployment of technology as a countermeasure against itself, thereby constructing a technical safeguard framework for algorithmic governance within the dynamic interaction between defense and offense. Regarding traceability and detection, it is important to develop recognition algorithms capable of identifying deepfake information and AI-generated content, and to construct a comprehensive technical framework for tracing AIGC-generated content within the political and legal domains, thereby building a closed-loop logic of regulating algorithms through algorithms. When it comes to algorithmic explainability, XAI technologies should be embedded in critical scenarios like public opinion analysis, thereby the decision-making logic of algorithms remains transparent and traceable for human overseers—a strategy that directly confronting the legitimacy challenges arising from black-box governance. For risk prevention and control, a pre-deployment evaluation mechanism for technical iteration should be established: any new AIGC technology designed for political or legal settings should be subject to risk assessment and security verification, so as to avert the hazards of deploying systems that carry underlying vulnerabilities.

5.2.2 Systemic resilience

In terms of systemic resilience, governance regimes ought to be refined in step with dynamic rule-formulation practices, so that governance standards do not lag behind technological progress. First and foremost, dedicated regulatory rules regulating algorithmic applications in the political and legal parts require stricter enforcement, specifying the boundaries of application and negative lists for AIGC in domains like public opinion monitoring, content generation, and decision-aiding. Next, a full-lifecycle regulatory regime—covering assessment, monitoring, filing, and accountability—ought to be put in place, with algorithmic risk assessment integrate into the prior approval procedures for political and legal informatization projects. Last but not the least, an flexible governance mechanism should be introduced; given how quickly AIGC technologies update, flexible means like regulatory sandboxes could facilitate achieve an adaptive balance between containing risk and encouraging innovation.

5.2.3 Subjective Resilience

A core means of enhancing stakeholder resilience consists in bridging the cognitive gaps among heterogeneous actors and fostering collaborative governance abilities in the age of algorithms [10]. For judicial and law enforcement institutions, a routine training mechanism on AI and algorithmic ethics ought to be adopted, one that enhances the algorithmic literacy of law enforcement officers and insists developing versatile legal and political practitioners. For platform operators, institutional regimes including algorithmic auditing and disclosure reporting need to be introduced to reinforce their sense of social responsibility, encouraging them to embed public values into the optimization objectives of their algorithmic design instead of relying merely on traffic indicators. For the general public, algorithmic literacy development should be incorporated into the national education framework, strengthening citizens' capacity to recognize and resist risks including information cocoons and deepfakes, and fortifying a societal line of defense against these risks.

5.2.4 Ecological resilience

The pursuit of ecological resilience lies in the synergistic coexistence of technology, institutions, actors, and the environment—thereby fostering a healthy ecosystem for algorithmic governance. First, establish a four-

pillar collaborative governance network encompassing the government–platform–public–experts stakeholders, clearly defining each party's role and the boundaries of their responsibilities in algorithmic risk governance. Second, create a cross-sectoral and cross-regional mechanism for sharing algorithmic risk-related information and coordinating responses on public opinion matters, thereby addressing the challenges of data silos and fragmented governance. Third, cultivate a social ecosystem for algorithmic governance by supporting independent third-party institutions in conducting algorithmic audits and assessments, and by shaping a diversified co-governance framework that integrates government regulation, industry self-discipline, and public oversight.

5.3 Prerequisites for the Effective Operation of the Mechanism

The effective operation of the resilience-breaking mechanism depends on the synergistic support of multiple conditions, including data, value, and regulatory discourse. At the data governance level, the rights confirmation, circulation, and security governance of political and legal data provide high-quality data support for identifying algorithmic risks. At the value normative level, core political and legal values—such as procedural justice, substantive justice, and human rights protection—should be translated into ethical constraints for algorithm design. At the global governance front, China should taken an active part in shaping global rules for algorithmic governance, including those governing cross-border data flows and the mutual recognition of algorithmic standards, and bring its own positions and experience into the international algorithmic governance agenda. This approach will widen the external institutional space for resilient governance and enhance the international compatibility of institutional arrangements.

6. Conclusion

Bringing AIGC into the governance of political and legal public opinion is a complex undertaking, one in which technological empowerment and the risks that come with it exist side by side. By systematically unpacking the technical logic through which AIGC restructures the landscape of political and legal public opinion regulation, this research identifies four key kinds of risks—cognitive manipulation,

algorithmic bias, information cocooning, and security erosion—investigates the transmission pathway and stage-specific developmental attributes of these risks, and builds a multi-layered risk mitigation regime spanning technological, institutional, stakeholder, and ecological resilience. The study reveals that governing algorithmic risks in public opinion management cannot remain at passive reaction; it should move to the proactive construction of resilience. Through the concurrent technological progression, institutions, stakeholders, and the ecosystem, a paradigm shift becomes attainable. Going forward, as AIGC technology continues to evolve and the digital transformation of the political and legal domains move further along, the essence and dynamic evolution of algorithmic hazards will grow in complexity. How to preserve security thresholds while cultivating technological innovation, and how to promote the key values of the political and legal regime while improving governance effectiveness, these questions still call for further in-depth inquiry. The formulation of a resilience mitigation framework is not a permanent solution, but rather an permeable system characterized by dynamic adjustment and sustained evolution; it must be continually enhanced within regulatory practices to truly accomplish the organic embedding of robust algorithmic governance and the stability of the political and legal architecture.

Acknowledgments

This paper is supported by Jinan Municipal Philosophy and Social Sciences Planning Research Project entitled “Research on the formulation of a Resilient Governance regime for Intelligent Guidance of Political and Legal Public Opinion in Jinan amid the Coexistence of AIGC Empowerment and Risks” (No. JNSK2026C143).

References

- [1] Wu, C., Lee, M., Lin, C., et al., A Longitudinal Study on Psychological Reactions and Resilience Among Young Survivors of a Burn Disaster in Taiwan 2015-2018, *Journal of Advanced Nursing*, 2020, (4), p.113.
- [2] Zeng, Y., Shi, J. Generative Artificial Intelligence in Government Governance: Logical Frameworks and Risk Regulation. *China's Administrative Management*, 2023,

- 39 (09):90.
- [3] Shen, Y. (2025). Research on a new method of generating artificial intelligence to empower ideological and political education. *Journal of Sociology and Education*. vol.2
- [4] Tsarouhas, P., & Grigoriadis, K. (2025). A socio-technical framework for AI trust in public administration. Transforming Government: People, Process and Policy. *Advance online publication*.
- [5] Cai, C. Promote the establishment of a new global governance paradigm for artificial intelligence. *National Governance*, 2025, (14):57.
- [6] Mandal, U., Setua, S. K., & Sarma, S. S. (2026). Deepfakes and beyond in the era of AI-generated disinformation: A systematic literature review of emerging technologies, challenges, and policy recommendations. *Journal of Visual Communication and Image Representation*, (98), Article 104748.
- [7] Zhou, J. (2025). Intelligent Recommendation Algorithm User Information Cocoon Room Effect Breakthrough Strategy. *Advances in Computer and Communication*, vol.6.
- [8] Gonzalez-Argote, J., Maldonado, E., & Maldonado, K. (2025). Algorithmic Bias and Data Justice: Ethical Challenges in Artificial Intelligence Systems. *EthAlca*, 159.
- [9] Zhan, G., Chen, Y. A conceptual model and generative mechanism for deepfake risks from a public safety perspective. *Academic Monthly*, 2025, 57(08):84.
- [10] Liu, X., Liu J. The Implications of Ecological Resilience Governance and an Analytical Approach to Its Implementation Pathways. *Journal of Southwest University (Social Sciences Edition)*, 2024, 50(05):103-115.