Proceedings of ICGPSH 2026 Symposium: Global Governance, Digital Politics, and the Transformation of Public Policy Systems
DOI: 10.54254/2753-7048/2026.ZJU34383

A Study on the Risks and Ethical Regulation of Artificial Intelligence in Public Decision-Making

Yixuan Chen

Faculty of International Media, Communication University of China, Beijing, China
chenyixuanecho@163.com

Abstract. As generative artificial intelligence, machine learning, and predictive analytics tools continue to enter government governance scenarios, public decision-making is shifting from "digital assistance" to "algorithmic embedding." While artificial intelligence can enhance the efficiency of policy identification, resource allocation, and risk early warning, its use in public decision-making also brings about compound risks, such as data bias, model opacity, outsourcing of responsibility, algorithmic bias, and compromised procedural justice. Based on interdisciplinary research at the intersection of public administration and technology law, this paper employs a methodology combining normative analysis, literature review, and comparative institutional analysis to address three central questions: "How does AI reshape the risk structure of public decision-making?" "Why are existing ethical and legal regulations insufficient?" and "How can a regulatory framework be established that balances efficiency and rights?" The study argues that the core gap in current regulation lies not merely in the absence of "human oversight," but in the lack of procedural intensity designs commensurate with decision-making risks. Based on this, the paper proposes a "tiered risk–procedural intensity matching" framework: according to the functional role and degree of rights impact of AI in public decision-making, differentiated measures such as registration and disclosure, algorithmic impact assessments, external audits, objection remedies, procurement transparency, and prohibition lists should be applied. This framework helps translate ethical principles into enforceable institutional arrangements.

Keywords: public decision-making, AI governance, algorithmic risk, ethical regulation, procedural justice

1. Introduction

In recent years, artificial intelligence has evolved from a tool for process automation in government services to high-impact scenarios such as administrative law enforcement, social assistance reviews, credit assessments, public safety early warning systems, resource allocation, and policy implementation monitoring [1, 2]. This signifies that AI is no longer merely a technological adjunct for the government to "improve efficiency," but is increasingly embedded within the decision-making chain of state power. As algorithmic models begin to participate in risk identification,

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2026.ZJU34383

individual ranking, recommendation generation, and even influence final decisions, the structure, basis, and boundaries of accountability in public decision-making are undergoing transformation [3].

Existing research from the perspectives of public administration, law, and technology ethics has already pointed out that algorithmic systems may replicate or even amplify existing inequalities through training data, proxy variables, objective functions, and human-machine interaction mechanisms, thereby undermining administrative discretion, procedural transparency, and democratic legitimacy [4, 5]. However, current discussions still suffer from two shortcomings. First, many studies treat "transparency," "explainability," and "human oversight" as universal solutions, without sufficiently distinguishing the varying levels of risk across different public decision-making contexts. Second, existing norms often emphasize the articulation of principles but rarely address how these principles can be translated into enforceable, accountable, and remedial institutional procedures.

In light of these issues, this paper adopts a methodology combining normative research, literature review, and comparative institutional analysis. It first delineates the functional boundaries of AI in public decision-making; second, analyzes risk-generation mechanisms and their ethical consequences; third, examines the progress and limitations of current regulations in the EU and China; and finally, proposes an innovative regulatory framework based on "tiered risk–procedural intensity matching." The theoretical significance of this paper lies in integrating legitimacy analysis in public administration with procedural safeguards in technology law; its practical significance lies in providing China's public sector with a more operational path for compliance and ethical governance when introducing artificial intelligence.

2. Modes and boundaries of AI embedding in public decision-making

2.1. From tool-assisted to embedded governance

The public sector's initial motivation for adopting AI is typically to improve administrative efficiency, reduce costs, and enhance predictive capabilities. In practice, however, algorithms quickly shift from "assisting in data processing" to "influencing judgment criteria" [1, 2]. Young et al. summarize this phenomenon as "artificial discretion," meaning that algorithmic systems do not eliminate discretion but rather reshape the conditions for its exercise, its visibility, and its distribution in new forms [1]. When algorithms participate in public decision-making, what truly changes is not merely the technical means, but how state power perceives individuals, classifies them, and treats them differentially.

2.2. Three functional roles in the public decision-making chain

Functionally, artificial intelligence generally plays three roles in public decision-making: First, information screening, which involves aggregating, cleaning, and identifying anomalous patterns in large-scale data; second, advisory support, which provides public officials with risk scores, rankings, or policy recommendations; and third, the outcome-determining role, which essentially decides whether an individual receives eligibility, resources, or penalties [2, 3]. The risk intensity of these three roles varies: the closer the role is to the final decision and the more it involves fundamental rights and significant interests, the less efficiency alone can serve as the basis for its implementation.

DOI: 10.54254/2753-7048/2026.ZJU34383

2.3. Boundary requirements from a public value perspective

Public decision-making differs from general commercial recommendations; its goal is not solely profit maximization, but the simultaneous pursuit of public values, such as efficiency, fairness, legitimacy, responsiveness, and accountability [4]. Therefore, determining whether AI is suitable for a particular public context cannot rely solely on accuracy rates or cost savings; it must also assess whether it undermines procedural justice, restricts the opportunities for vulnerable groups to express themselves, or leads to "black-box administration" where accountability is difficult to establish. This constitutes the fundamental boundaries for the application of AI in the public sphere.

3. Risk generation mechanisms of AI in public decision-making

3.1. Data bias and agency discrimination

The risks associated with artificial intelligence systems stem primarily from the data layer. Historical data used by the public sector often contains systemic biases; for example, records of law enforcement frequency, welfare claims, or credit defaults do not necessarily reflect "true risk," but may simply be the result of past governance practices. If models continue to rely on this data, they will transform historical inequalities into the basis for future decisions, thereby creating a situation where "bias is reproduced in technological form" [5]. A more subtle issue lies in proxy variables: even if a system does not directly use sensitive attributes, it may indirectly result in differential treatment through variables such as ZIP codes, occupation, or educational background.

3.2. Opacity and the shift of discretion

The second risk of artificial intelligence lies in the opacity of decision-making. The more complex the algorithmic model, the harder it is for administrative parties to understand why they are flagged, ranked, or rejected; even frontline civil servants may struggle to fully grasp the model's limitations [2]. On the surface, algorithms appear to reduce human discretion; in reality, however, discretion has merely shifted from front-end case-by-case judgments to back-end data selection, model configuration, threshold design, and vendor interfaces [1, 2]. This creates a new governance paradox: administrative actions still exercise power, yet the judgmental processes that truly determine outcomes have become even less visible.

3.3. Algorithmic bias and selective reliance

Even if institutional requirements stipulate that "the final decision rests with humans," the problem does not automatically disappear. Empirical studies show that when faced with algorithmic recommendations, public sector personnel may exhibit "automation bias"—over-reliance on algorithms even when warning signals are present—or "selective reliance"—a greater willingness to accept recommendations when algorithmic conclusions align with existing stereotypes [6]. This indicates that human-machine collaboration does not inherently guarantee greater fairness; without evidence-based support, training mechanisms, and organizational constraints, human oversight is likely to amount to little more than a token endorsement.

DOI: 10.54254/2753-7048/2026.ZJU34383

3.4. Outsourcing of responsibility and the vendor black box

In real-world governance, a large number of public algorithms are developed, maintained, or updated by third-party technology companies. This results not merely in technical dependence, but in the externalization of the accountability structure: governments may evade their duty to explain by claiming that "the system was developed by the supplier," while suppliers refuse to disclose model details on the grounds of trade secrets, ultimately creating a vacuum of accountability [7]. When algorithms enter the public decision-making chain through government procurement, private entities effectively participate in the exercise of state power. Consequently, their accountability cannot be confined to the level of traditional contractual liability but must be incorporated into the framework of public law review and public responsibility.

4. Ethical conflicts in AI-driven public decision-making

4.1. The tension between efficiency and procedural justice

Proponents of AI use in the public sector typically emphasize its advantages in identifying high-risk individuals, accelerating administrative processes, and conserving fiscal resources [3]. However, the fundamental difference between public decision-making and private business operations lies in the fact that the former must not only pursue effective outcomes but also uphold due process. Citron has long pointed out that automated decision-making may formally enhance consistency but substantively undermine individuals' opportunities for defense, informed consent, and correction [8]. Therefore, what the public sector truly needs to guard against is not "technology being too slow," but rather "procedures being circumvented by technology."

4.2. The legitimacy crisis: the three dimensions of input, process, and output

From the perspective of legitimacy, algorithmic public decision-making faces at least three risks: at the input level, the public often does not know which decision-making stages have incorporated algorithms and lacks opportunities to participate in rule-setting; at the process level, models, data, and evaluation metrics lack sufficient transparency and auditability; at the output level, even if efficiency improves, public trust in the government may be eroded due to biases and unintended harm [5]. This implies that the governance of artificial intelligence is not merely a matter of accuracy but a question of institutional legitimacy.

4.3. The limitations of the "human oversight" myth

Many current political documents regard "retaining human final decision-making authority" as a key safeguard. However, after a systematic review, Green points out that policies relying solely on human oversight are often based on an unproven assumption: namely, that humans can effectively identify and correct algorithmic errors [9]. Under the combined pressures of time constraints, performance evaluations, and reliance on expertise, supervisors may lack both the capacity and the incentive to override system recommendations. Therefore, what truly needs to be designed is not abstract human oversight, but institutional oversight that is auditable, reviewable and accountable.

DOI: 10.54254/2753-7048/2026.ZJU34383

5. Progress and shortcomings of existing ethical and legal regulations

5.1. The European Union: from data protection to risk-based regulation

The EU's existing framework provides two key pillars for public algorithm governance. First, Article 22 of the General Data Protection Regulation (GDPR) imposes restrictions on decisions "based solely on automated processing which produce legal effects or similarly significant effects," and requires the provision of opportunities for human intervention, expression of views, and challenging such decisions [10]. Second, the EU AI Act classifies various systems in scenarios such as biometrics, education, employment, essential public services, law enforcement, immigration, and the judiciary as high-risk AI, requiring compliance arrangements including risk management, data governance, technical documentation, logging, transparency obligations, and human oversight [11]. A key contribution of this framework is that it begins to link technical risks with institutional accountability.

5.2. China: incremental governance through principles, regulations, and filing

China has established a preliminary framework for AI governance, ranging from overarching principles to specific regulations. The 2019 "Principles for the Governance of Next-Generation Artificial Intelligence" set forth requirements such as harmony and friendliness, fairness and justice, respect for privacy, security and controllability, shared responsibility, and agile governance [12]; the 2021 *Ethical Guidelines for Next-Generation Artificial Intelligence* further embedded fairness, privacy, controllability and trustworthiness, and accountability into the entire lifecycle of AI [13]; the 2022 *Regulations on the Management of Algorithm-Based Recommendations in Internet Information Services* promotes algorithm governance through measures such as filing, user rights protection, and security assessments [14]. However, these regulations focus more on platform governance and internet information services, and specific regulations for public decision-making scenarios—such as administrative approval, social assistance, public safety, and administrative law enforcement—remain insufficient.

5.3. Three shortcomings of current regulations

First, the regulatory targets do not fully align with the actual risks in public decision-making. Many regulations center on "algorithm recommendations" and "generated content," while procedural requirements for administrative agencies' use of decision-support systems remain insufficiently detailed. Second, the existing system emphasizes ex post accountability over ex ante justification and does not yet universally require public departments to demonstrate the necessity, appropriateness, and availability of alternatives before procuring or deploying algorithms. Third, external oversight remains weak; the public finds it difficult to promptly identify which administrative scenarios utilize AI, and lacks access to substantive channels for objection and redress [15].

DOI: 10.54254/2753-7048/2026.ZJU34383

6. Pathways to innovation: establishing an ethical regulatory framework based on "risk tiering and procedural intensity matching"

6.1. Proposal of the framework: shifting from "whether to use AI" to "how to govern AI based on risks"

This paper argues that the regulation of artificial intelligence in public decision-making should not remain stuck in a binary debate of "for or against," but should shift to the question of "what procedural safeguards should be allocated according to the intensity of risk?" Unlike universal ethical principles, the "risk-tiered—procedural intensity matching" framework emphasizes that the closer an algorithm is to the final decision, the greater its impact on citizens' fundamental rights, and the more it involves vulnerable groups, the higher the procedural intensity should be. In other words, the focus of regulation should shift from abstract technical attributes to concrete power implications.

6.2. Risk classification: a dual assessment based on functional role and impact on rights

Public sector AI systems can be categorized into four types. The first is the low-risk information-assistance type, used solely for internal data organization and trend identification, with no direct impact on individual rights; the second is the medium-risk advisory-support type, which influences resource allocation or law enforcement priorities but does not directly make final decisions; the third is the high-risk decision-impact type, where although human approval is retained, algorithmic scoring substantially shapes the final outcome; the fourth category is the unacceptable-risk type, which involves implementing unchallengeable, fully automated decisions in scenarios with significant implications for rights, such as social assistance, law enforcement penalties, and disqualification. This classification better reflects actual risks than the simple distinction of "human involvement or not."

6.3. Configuration of procedural stringency: translating ethical principles into operational mechanisms

Corresponding to the above classifications, low-risk systems should at a minimum establish user registration, basic disclosure, and data security requirements; medium-risk systems should add algorithmic impact assessments, data bias testing, records of human review, and periodic performance evaluations; high-risk systems should be required to conduct independent external audits, provide explanations for decision-making, establish appeal review procedures, maintain logs, and ensure procurement transparency; for systems classified as "unacceptable risk," explicit prohibition lists or substantive restrictions should be established, prohibiting their use as the sole or decisive basis for decisions. This design incorporates the EU's risk-tiered approach while also addressing Green's critique of the "myth of human oversight," as it calls for institutionalized, evidence-based oversight rather than mere sloganeering [9, 11].

6.4. Supporting Mechanisms: procurement transparency, public participation, and continuous accountability

To ensure the effective implementation of the aforementioned framework, three supporting mechanisms are required. First, establish a transparent public algorithm procurement system that discloses, at a minimum, the system's purpose, supplier, core functions, scope of application, evaluation conclusions, and update records to prevent accountability from ending with contract

DOI: 10.54254/2753-7048/2026.ZJU34383

outsourcing [7]. Second, introduce public participation and expert review in high-impact scenarios, ensuring that models undergo democratic scrutiny prior to deployment rather than relying on ex post facto corrections. Third, establish a mechanism for continuous accountability by incorporating algorithm logs, performance deviations, appeal outcomes, and deactivation criteria into routine oversight, thereby transitioning from "static compliance" to "dynamic governance."

7. Conclusion

This paper analyzes the risks and ethical regulatory issues surrounding the application of artificial intelligence in public decision-making, arguing that the integration of AI into the public sector is not merely a technological upgrade but a restructuring of how state power operates. The primary risks do not lie solely in the algorithms themselves but within the governance chain formed by data, models, human-machine interaction, organizational incentives, and vendor relationships. The resulting data bias, lack of transparency, algorithmic bias, and outsourcing of responsibility will further undermine procedural justice, public trust, and administrative legitimacy [5-7].

A comparison of existing systems reveals that the European Union has established a regulatory framework centered on data protection and high-risk regulation [10, 11], while China has gradually built a governance system grounded in governance principles, ethical norms, and algorithm registration [12-14]. However, both China and the EU currently face a common challenge: while ethical principles are abundant, the institutional density required to translate them into specific procedural obligations, remedial mechanisms, and procurement constraints remains insufficient.

Based on this, the innovation proposed in this paper lies in replacing the vague calls for "enhancing transparency" or "maintaining human oversight" with a "tiered risk–procedural intensity matching" framework. This framework emphasizes that, depending on the functional role of AI within the public decision-making chain and the extent of its impact on rights, differentiated mechanisms for prior justification, impact assessment, external auditing, explanation, remedies for objections, and deactivation should be implemented. The value of this framework lies in its ability to bridge the analysis of legitimacy in public administration with procedural safeguards in technology law, while better aligning with the reality of public decision-making scenarios characterized by unequal power dynamics and highly sensitive consequences. This paper remains primarily a normative study and lacks empirical comparisons of algorithmic practices across different departments and decision-making types; future research could further examine the applicability and cost-effectiveness of this framework by integrating local government cases, algorithmic procurement documents, and samples of administrative reconsideration/litigation, while exploring specific pathways for constructing China's public algorithm inventory system, impact assessment system, and specialized redress mechanisms.

References

[1] Young, M. M., Bullock, J. B., & Lecy, J. D. (2019) Artificial discretion as a tool of governance: a framework for understanding the impact of artificial intelligence on public administration [J]. Perspectives on Public Management and Governance, 2(4): 301-313. https: //doi.org/10.1093/ppmgov/gvz014

[2] Bullock, J., Young, M. M., & Wang, Y.F. (2020) Artificial intelligence, bureaucratic form, and discretion in public service [J]. Information Polity, 25(4): 491-506. https: //doi.org/10.3233/IP-200223

[3] Madan, R., & Ashok, M. (2023) AI adoption and diffusion in public administration: a systematic literature review and future research agenda [J]. Government Information Quarterly, 40(1): 101774. https: //doi.org/10.1016/j.giq.2022.101774

[4] Selten, F., & Meijer, A. (2021) Managing algorithms for public value [J]. International Journal of Public Administration in the Digital Age, 8(1): 1-16. https: //doi.org/10.4018/IJPADA.20210101.oa9

DOI: 10.54254/2753-7048/2026.ZJU34383

[5] Grimmelikhuijsen, S., Meijer, A. (2022) Legitimacy of algorithmic decision-making: six threats and the need for a calibrated institutional response [J]. Perspectives on Public Management and Governance, 5(3): 232-242. https: //doi.org/10.1093/ppmgov/gvac008
[6] Alon-Barkat, S., Busuioc, M. (2023) Human-AI interactions in public sector decision making: “automation bias” and “selective adherence” to algorithmic advice [J]. Journal of Public Administration Research and Theory, 33(1): 153-169. https: //doi.org/10.1093/jopart/muac007
[7] Crawford, K., Schultz, J. (2019) AI systems as state actors [J]. Columbia Law Review, 119(7): 1941-1972. https: //columbialawreview.org/content/ai-systems-as-state-actors/
[8] Citron, D. K. (2008) Technological due process [J]. Washington University Law Review, 85(6): 1249-1314. https: //openscholarship.wustl.edu/law_lawreview/vol85/iss6/2/
[9] Green, B. (2022) The flaws of policies requiring human oversight of government algorithms [J]. Computer Law & Security Review, 45: 105681. https: //doi.org/10.1016/j.clsr.2022.105681
[10] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation), Article 22 [EB/OL]. https: //www.legislation.gov.uk/eur/2016/679/article/22
[11] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (EU AI Act) [EB/OL]. https: //eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
[12] Ministry of Science and Technology of the People's Republic of China. (2019) New Generation Artificial Intelligence Governance Principles - Developing Responsible Artificial Intelligence [EB/OL]. June 17. https: //www.most.gov.cn/kjbgz/201906/t20190617_147107.html
[13] Ministry of Science and Technology of the People's Republic of China. (2021) Ethical Norms for New Generation Artificial Intelligence [EB/OL]. September 26. https: //www.most.gov.cn/kjbgz/202109/t20210926_177063.html
[14] Cyberspace Administration of China. (2022) Provisions on the Management of Algorithmic Recommendations in Internet Information Services [EB/OL]. January 4. https: //www.cac.gov.cn/2022-01/04/c_1642894606364259.htm
[15] Bignami, F. (2022) Artificial intelligence accountability of public administration [J]. The American Journal of Comparative Law, 70(Supplement_1): i312-i346. https: //doi.org/10.1093/ajcl/avac012