Postprint: Algorithmic Discrimination Risks and Prevention Strategies in AI-Embedded Government Data Governance

Authors: Peng Lihui, Zhang Qiong, Li Tianyi

Date: 2024-10-06T00:00:00+00:00

Abstract

[Purpose/Significance] This study aims to explore the application of artificial intelligence in government data governance and the algorithmic discrimination issues it entails, and to propose corresponding solution strategies to safeguard citizens' legitimate rights and interests as well as government credibility. [Method/Process] Through the literature induction method, this paper analyzes the specific applications of artificial intelligence algorithms in government data governance, identifies the causes of algorithmic discrimination—including data one-sidedness, designers' perspectives, and social prejudices—and further examines the potential risks of algorithmic discrimination, proposing corresponding prevention and control measures. [Results/Conclusion] Research indicates that while the integration of artificial intelligence into government data governance enhances efficiency, it also introduces risks of algorithmic discrimination. Accordingly, this study proposes prevention and control measures such as clarifying algorithmic fairness, formulating industry standards, improving accountability mechanisms, and optimizing the data environment, to ensure that artificial intelligence effectively benefits the people in government data governance.

Full Text

Preamble

Journal of Agricultural Library and Information Science
DOI: 10.13998/j.cnki.issn1002-1248.24-0353

Research on Algorithmic Discrimination Risks and Prevention Strategies in AI-Embedded Government Data Governance

Peng Lihui[1], Zhang Qiong[1], Li Tianyi[2]
(1. School of Public Administration, Xiangtan University, Xiangtan 411105;

2. Network Security Detachment, Shenyang Railway Public Security Office, Shenyang 110167)

Abstract:
[**Purpose/Significance**] This study investigates the application of artificial intelligence in government data governance and the attendant problem of algorithmic discrimination, proposing corresponding solutions to safeguard citizens' legitimate rights and government credibility. [**Method/Process**] Through literature review and analysis, this paper examines specific applications of AI algorithms in government data governance, identifies causes of algorithmic discrimination—including data partiality, designer biases, and social prejudices—and explores potential risks and preventive measures. [**Results/Conclusions**] The research demonstrates that while AI integration in government data governance enhances efficiency, it also introduces algorithmic discrimination risks. Accordingly, this study proposes prevention and control strategies including clarifying algorithmic fairness, establishing industry standards, improving accountability mechanisms, and optimizing data environments to ensure AI effectively benefits the public in government data governance.

Keywords: artificial intelligence; government data governance; algorithmic discrimination; risk prevention and control

Citation: Peng Lihui, Zhang Qiong, Li Tianyi. Research on algorithmic discrimination risks and prevention strategies in AI-embedded government data governance[J]. Journal of Agricultural Library and Information Science, 2024, 36(5): 23-31.

---

0 Introduction

In an era of rapid technological advancement, artificial intelligence algorithms are playing an indispensable role in government data governance. Data and algorithms collectively constitute the underlying logic of AI, offering exceptional information processing capabilities, efficient decision-making, and precise predictive power. However, due to inherent algorithmic characteristics, social biases may trigger concerns about fairness and equal treatment. This discrimination is covert, and allowing such risks to go unchecked will exacerbate power alienation, undermine citizens' legitimate rights and interests, and ultimately precipitate a severe crisis of trust in government. AI has reshaped public decision-making models and effectiveness, yet the algorithmic discrimination it brings has aroused widespread societal concern. On July 10, 2023, seven ministries including the Ministry of Education issued the *Interim Measures for the Administration of Generative AI Services* to safeguard national security and public interests.

1.1 The Connotation of Algorithmic Discrimination

Algorithmic discrimination, a compound concept combining algorithms and discrimination, refers to the unfair phenomenon that easily arises when algorithms analyze and process data, particularly affecting specific groups. In everyday contexts, discrimination denotes unjust differential treatment based on membership characteristics within particular groups or categories. In government data governance, this is termed algorithmic bias, with roots traceable to algorithm design flaws, data biases, and external factor interference. These biases may manifest in various forms, posing challenges to policy formulation and social justice, including unfairness based on gender or ethnicity. To mitigate this issue, governments should strengthen data management and optimize algorithmic regulatory frameworks to enhance decision-making fairness and transparency.

1.2 Characteristics of Algorithmic Discrimination

1.2.1 Covertness of Algorithmic Discrimination

Due to the difficulty for ordinary people to comprehend algorithmic decision-making processes, scholars have aptly termed this the "black box." The sources of algorithmic opacity may stem from three aspects: opacity arising from commercial secrets or state confidentiality, opacity caused by data literacy disparities, and opacity resulting from extreme complexity in algorithmic processes. Algorithms, the core of AI, involve a series of computational steps that take certain values as input and generate certain values as output. This black box enables highly covert application of unjust treatment to individuals. Since algorithmic operations are considered objectively neutral, and since discriminatory models can easily be created by deliberately adding certain features as input options or deliberately omitting/reducing certain variables during training, most groups cannot promptly detect and respond to discrimination. The complexity of algorithmic models and ambiguity of decision-making processes provide feasible pathways for developers to conceal discriminatory behavior, leading to algorithmic discrimination. A well-known example is the COMPAS recidivism prediction program implemented in the early 2000s, which represents a real case of algorithmic discrimination. Individuals cannot understand the algorithmic decision-making process and can only make intuitive judgments based on external forms, which undoubtedly increases the covertness of algorithmic discrimination.

1.2.2 Precision of Algorithmic Discrimination

Algorithmic discrimination exhibits precision in two aspects. First, based on big data analysis, algorithmic decision-making possesses natural advantages in analytical capability, overturning traditional discrimination that could only target small groups based on 显性特征 such as race. Algorithms can uncover deeplevel hidden variables—such as webpage browsing time and mobile terminal information—to conduct more precise and diversified classifications for discriminatory treatment. For instance, commercial banks' credit algorithms in the United States, driven by racial discrimination and group default rates, often grant Black individuals and other people of color fewer loan opportunities under equivalent circumstances. Second, algorithmic discrimination can render discriminatory judgments on virtually all internet users' real-time online behaviors. Big data algorithms can constantly monitor individual digital footprints, breaking down barriers between individuals and making the "granular society" phenomenon increasingly prominent. When enterprises design recruitment algorithm systems, they often use built-in algorithms to filter out employees with longer commute times, depriving them of equal competition opportunities.

1.2.3 Complexity of Accountability in Algorithmic Discrimination

AI is not an isolated programmatic island. Behind algorithmic decision-making, designers, data providers, and users all play essential roles, making accountability extremely complex. For example, during the 2016 U.S. election, a data analysis company mined Facebook historical data to manipulate voters' political choices through algorithms. After deployment, decision-making functions rely on real-time updated data, meaning erroneous input information or misleading information obtained during self-improvement phases can cause algorithmic bias. Data providers are also frequently listed as suspects in algorithmic discrimination. Whether data providers act intentionally (as when Microsoft's chatbot was taught racist concepts by malicious users) or unintentionally (as when surveyors cannot obtain valid data from Black communities due to historical underrepresentation, leading to sampling frame errors), poorly handled non-sampling errors can cause severe algorithmic bias.

1.3 Causes of Algorithmic Discrimination

1.3.1 Subject Dimension

Algorithm designers exhibit inertial thinking. As the crystallization of mathematics and computer science, algorithms have deeply integrated into human social production and life. The design process is inevitably influenced by designers' knowledge, experience, and viewpoints, sometimes reflecting certain preferences or tendencies. Even without intentional prejudice, designers' implicit values are often consciously or unconsciously embedded during data collection, algorithmic filtering, and purposive analysis, subsequently superimposing this unfairness in model construction and producing biased outputs that externalize as unjust decisions or behaviors targeting individuals or social groups.

1.3.2 Object Dimension

Big data, with its massive volume advantage, often leads people to arbitrarily assume its samples are highly representative. However, as Gallup's classic case from the 1930s demonstrated, simply expanding sample size without addressing sampling frame errors and non-response errors can lead to severe algorithmic bias. The premise that sample statistics adequately represent population parameters is crucial for objective and effective algorithmic decision-making. Once data sources lack comprehensiveness, errors accumulate in subsequent modeling. U.S. health informaticist Deas Kimberly highlighted a skin cancer screening algorithm that effectively detected white positive cases but struggled to accurately identify positive cancer in Black individuals, with particularly pronounced errors for darker-skinned patients. This occurred because Black people, comprising 13% of the U.S. population, have minimal industry representation in healthcare, causing extreme distrust of medical professionals in Black communities and making it difficult for investigators to obtain valid data, forcing focus on white communities instead.

1.3.3 Model Dimension

Statistical inference plays a crucial role in algorithmic models. If tool rationality is pursued exclusively while neglecting value rationality, the internal logic of making predictive assessments about population parameters based on sample statistics will inevitably create probabilistic associations between individuals and groups, causing individuals to be labeled with group characteristics and receive differential treatment. This can expand the Matthew effect and sow seeds of social disorder. Even with authentic and comprehensive data samples, algorithmic discrimination may still occur.

---

2. Risks of Algorithmic Discrimination in AI-Embedded Government Data Governance

2.1 Exacerbating Group Polarization and Expanding the Matthew Effect

In AI-embedded government data governance, algorithms may neglect or marginalize minority group needs due to training data biases or designers' subjective tendencies. This neglect can cause minority voices and interests to be overlooked, exacerbating social inequality. With powerful computing capabilities and massive data support, algorithms can quickly obtain "optimal solutions" in complex social governance scenarios, yet optimal solutions do not always equate to most appropriate solutions. AI may tend to protect majority interests at the expense of minority rights when facing moral dilemmas like the trolley problem. Similar issues exist in resource allocation—if algorithms make decisions based on students' historical performance and school resource allocation, they may prioritize already advantaged schools and students while neglecting those needing more support, thereby intensifying educational resource imbalances and expanding social inequality.

2.2 Strengthening Information Cocoons and Damaging Citizens' Rights

In AI-embedded government data governance, targeted data screening and intentional classification rules often grant citizens priority ranking based on group characteristics. This represents direct discrimination. For example, Carnegie Mellon University researchers concluded through experiments that Google's advertising algorithm exhibits significant gender discrimination when pushing recruitment information, mirroring real-world employment environments. Under deep learning, algorithmic self-improvement creates more opportunities for embedding social biases. As data-driven information systems flood government decision-making processes, opacity issues become more severe. When governments rely on biased data for decision-making, it may lead to misallocation of public resources and, in some cases, trigger social injustice. If algorithms incorrectly assess a community's needs, that community's public services may suffer, affecting residents' quality of life. This information barrier not only damages citizens' right to know and choose but may also cause misunderstanding or prejudice toward government decisions and actions.

2.3 Difficulties in Regulating Loopholes and Inducing Power Alienation

Due to its covertness and complex accountability, algorithmic discrimination is difficult to effectively regulate legally. While AI algorithms significantly improve work efficiency in government data governance, they also pose security concerns that cannot be ignored. Algorithmic vulnerabilities and defects, combined with high technical complexity and continuous updating characteristics, create extreme challenges for regulatory departments, making effective regulation difficult. This regulatory lag and incompleteness may provide opportunities for criminals and breed power alienation. If individuals manipulate algorithms to quietly influence public resource allocation, it will cause social injustice. As the most important protector of vulnerable groups, government must uphold fairness and justice values. The opacity of algorithms provides hidden soil for power abuse, making it difficult for the public and supervisory institutions to understand internal algorithmic mechanisms, thus hindering effective oversight.

2.4 Weakening Public Trust and Increasing Governance Difficulty

If the public questions algorithmic fairness and transparency, it may trigger social dissatisfaction and increase governance difficulty. Generative AI, led by new technologies, has pushed government data governance to a new stage but also brought new risks. Its powerful generative capabilities have broadened algorithmic applications in government decision-making, but its high autonomy makes behavior prediction and control more difficult, potentially exacerbating the covertness of algorithmic discrimination and further complicating social governance. According to the 48th China Statistical Report on Internet Development, the number of tampering incidents on Chinese government websites increased by nearly 30.9% compared with the same period in 2019, already causing actual impacts on public security. When adopting AI technology, government must carefully consider its potential impact on public trust and actively take measures to enhance decision-making transparency and fairness.

---

3. Prevention and Control Strategies for Algorithmic Discrimination Risks

3.1 Clarifying Algorithmic Fairness and Adhering to Results-Oriented Approach

When embedding AI in government data governance, clarifying algorithmic fairness principles is crucial. Due to algorithmic discrimination's covertness and difficulty in definition, government should take the leading role. Research shows that fairness in real life pursues structural fairness rather than static equality. When formulating fairness standards, government must consider data diversity, algorithmic decision logic, and potential impacts on different groups. A results-oriented thinking mode should run through the entire algorithm design and implementation process to ensure outputs do not favor any specific group, thereby avoiding discriminatory decisions. To achieve this, government should establish independent algorithmic fairness assessment agencies to conduct regular reviews and monitoring of algorithms used in data governance. During development and application, government should actively invite diversified stakeholders—including enterprises, social organizations, and the public—to participate through hearings and other forms, collecting opinions to ensure algorithm design fully reflects different groups' interests and needs, with enhanced attention to minority groups.

3.2 Establishing Industry Standards and Upholding Algorithmic Transparency

In the journey to enhance algorithmic transparency in government data governance, algorithm developers play a central role. They should provide detailed documentation explaining algorithm principles, data sources, expected outputs, and establish public platforms to display algorithm details used in government data governance for public supervision and review. Media plays a key role in promoting algorithmic transparency by reporting cases to reduce public alienation from technological unfamiliarity and help more people understand algorithms' complexity and limitations. However, media should avoid over-reliance on automatically generated content or oversimplifying complex issues, instead maintaining independent thinking and in-depth analysis to guide public rational attitudes toward algorithmic technology. If no unified industry standards exist, even perfect laws will fall into the dilemma of "not punishing the multitude." Therefore, the industry should spontaneously organize to jointly develop a detailed set of algorithm development and application standards, implementation norms, and regulatory measures, clarifying algorithm design principles, moral and legal responsibilities, and forming unified industry norms.

3.3 Advocating Social Fairness and Optimizing Data Environment

In any era, the meaning of law is to ensure the realization of norms rather than mere punishment or relief. Social biases themselves are often the fundamental source of algorithmic discrimination. As previously mentioned, cases of algorithmic gender discrimination and racial discrimination embed social prejudices into 偏激 algorithm design, making the supposedly objective and neutral algorithmic world present situations similar to human society. To fundamentally solve such problems, we must restrain inherent human biases. Data, as the carrier of human behavior, is relatively easier to change. In data collection, we should ensure authentic data is fed into algorithms; in data preprocessing, private information should be consciously desensitized; in data transmission, maximum filtering of society's short-sightedness should occur to create a more just, transparent, and efficient data environment for AI. Only by filtering social prejudices can AI be cleansed of discrimination and serve society with a new look.

3.4 Improving Accountability Mechanisms and Promoting Multi-Participant Governance

Given algorithmic discrimination's covertness and complex accountability, responsibility should be clearly and fairly distributed among different subjects. Algorithm design companies should actively improve algorithmic transparency, focusing on authenticity and understandability, and report to government departments in the form of manuals. They should also explain data sources and collection mechanisms to the public through short videos on platforms like Douyin, enabling necessary perceptual understanding of algorithms. Government departments, as initiators and important regulators of government algorithms, should supervise AI embedding and usage. Although different management models can integrate institutional resources, they often lead to fragmentation and lack of coordination in practice. Only by establishing specialized responsibility departments can reasonable demands from enterprises not be rejected and citizens' rights protection actions be facilitated. Citizens, as the largest beneficiaries of government algorithms and direct victims of algorithmic discrimination, should actively participate in algorithmic governance by learning algorithm knowledge and improving data literacy. Through multi-party collaborative governance, we can prevent the emergence of "algorithmic leviathan."

4 Conclusion

The embedding of AI in government data governance, while achieving intelligent integration and accelerating governance, also intensifies information cocoon phenomena and Matthew effects, damaging citizens' vital interests. Without regulation, algorithmic discrimination risks may trigger severe social trust crises. This study first elaborated on algorithmic discrimination's connotation, characteristics, and causes, then conducted a comprehensive, multi-level analysis of algorithmic discrimination risks in AI-embedded government data governance from subject, object, and model dimensions, and finally proposed prevention measures including clarifying responsibilities of AI stakeholders, unifying industry standards, promoting data fairness, strengthening algorithmic transparency and explainability, and accelerating the improvement of multi-participant governance. These measures aim to safeguard effective AI benefits for the people while avoiding high-risk algorithmic discrimination patterns. The proposed regulatory path represents an exploration of AI 业态, aiming to provide theoretical support and practical guidance for related research. Although algorithmic discrimination risks cannot be completely eliminated in the short term under current technical conditions, we should neither exaggerate these risks nor negatively deny technology's value.

References

[1] XIE Y J, YANG Y X. Algorithmic discrimination and its governance in era of artificial intelligence[J]. Journal of Beijing University of Posts and Telecommunications (Social Sciences Edition), 2022, 24(5): 18-25.

[2] Ministry of Justice of the People's Republic of China. Interim Measures for the Administration of Generative AI Services[EB/OL]. [2024-03-18]. https://www.moj.gov.cn/pub/sfbgw/flfggz/flfggzbmgz/202401/t20240109_{493171}.html.

[3] BURRELL J. How the machine "thinks": Understanding opacity in machine learning algorithms[J]. Big Data & Society, 2016, 3(1): 2053-9517.

[4] ROMEI A, RUGGIERI S. A multidisciplinary survey on discrimination analysis[J]. The Knowledge Engineering Review, 2014, 29(5): 582-638.

[5] YAN K R. Algorithm deviation of artificial intelligence and its avoidance[J]. Jianghai Academic Journal, 2020(5): 141-146.

[6] ANGWIN J. Machine bias: There's software used across the country to predict future criminals and it's biased against blacks[EB/OL]. [2024-03-19]. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[7] PASQUALE F. The Black Box Society: The Secret Algorithms That Control Money and Information[M]. Beijing: China Citic Press, 2015.

[8] BURRELL J. How the machine "thinks": Understanding opacity in machine learning algorithms[J]. Big Data & Society, 2016, 3(1): 2053-9517.

[9] AZIZ Z H. Racial equity in algorithmic criminal justice[J]. Duke Law Journal, 2019, 68(6): 1043-1134.

[10] DIAKOPOULOS N. Accountability in algorithmic decision making[J]. Communications of the ACM, 2016, 59(2): 56-62.

[11] CAIRE C M. When algorithms discriminate[N]. The New York Times, 2015-07-10(12).

[12] CITRON D K, PASQUALE F. The scored society: Due process for automated predictions[J]. Washington Law Review, 2014, 89(1): 1-33.

[13] STEININGER M, JUR S. German Supreme Court stands firm on liability for unjustified warning letters[J]. Journal of Intellectual Property Law & Practice, 2006, 1(4): 247-248.

[14] Nytimes. How Trump consultants exploited the Facebook data of millions[EB/OL]. [2024-03-26]. https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html.

[15] LIU P, CHI Z J. Ethical reflections on algorithmic discrimination[J]. Journal of Dialectics of Nature, 2019, 41(10): 16-23.

[16] JIA J P. Statistics[M]. 7th ed. Beijing: China Renmin University Press, 2018.

[17] VEALE M, KLEEK M V, BINNS R. Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making[C]//Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, 2018: 1-14.

[18] ZHAO T Y. The forking paths for the trolley problem[J]. Philosophical Research, 2015(5): 96-102.

[19] JIA K. Artificial intelligence and algorithm governance research[J]. Chinese Public Administration, 2019(1): 17-22.

[20] LIU Y B. The isolated island metaphor: The invasion of algorithmic ethics in the era of AI[J]. Administrative Tribune, 2019, 26(6): 121-128.

[21] FU Z Q, SUN W P. Science and technology towards "goodness": The value follow of artificial intelligence development[J]. Gansu Social Sciences, 2021(2): 97-103.

[22] KAPLAN J. Humans Need Not Apply: A Guide to Wealth and Work in the Age of Artificial Intelligence[M]. New Haven: Yale University Press, 2015.

[23] BU S. On the algorithm discrimination in artificial intelligence and the relevant review criteria[J]. Journal of Shanxi University (Philosophy and Social Science Edition), 2019, 42(4): 124-129.

[24] Cyberspace Administration of China, China Internet Network Information Center. The 48th China Statistical Report on Internet Development[EB/OL]. [2024-03-27]. https://www.cac.gov.cn/2021-02/03/c_{1613923423079314}.htm.

[25] SKIRPAN M, GORELICK M. The authority of "fair" in machine learning[J/OL]. arXiv preprint arXiv:1706.09976, 2017.

[26] NICHOLAS L. Sociology of Law[M]. Shanghai: Shanghai People's Publishing House, 2013: 188.

[27] LIU P, CHI Z J. Ethical issues of algorithms and their solutions[J]. Journal of Northeastern University (Social Science), 2019, 21(2): 118-125.

[28] ZHANG J. Reflection and comment on "technology neutrality" in copyright law[J]. Intellectual Property, 2008, 18(1): 72-76.

Note: Figure translations are in progress. See original paper for figures.