Liability Rules for Algorithmic Discrimination

Hu Zhiju

School of Law and Political Science, Urban and Rural Cultural Development Research Center, Guangzhou College of Applied Science and Technology, Guangzhou, Guangdong, 511370, China

Abstract: Algorithmic discrimination represents the translation of human societal discrimination into the realm of algorithms. It manifests as algorithmic thinking, black-box algorithms, and non-neutrality, all stemming from the design of algorithmic technologies. Furthermore, it becomes apparent in the form of biased decision-making during algorithmic operations. This phenomenon not only infringes upon citizens' equal rights but also disrupts market economic order and jeopardizes social fairness and justice. To effectively address the real-world issues arising from algorithmic discrimination, it is imperative to approach this problem from a legal perspective. Algorithmic discrimination differs from traditional forms of infringement, necessitating a thorough analysis of its constituent elements, responsible entities, modes of liability, and the establishment of mechanisms for liability allocation. Through these legal regulations, a legal framework for the algorithmic era can be constructed, safeguarding the safety and development of humanity in the digital society.

1. Introduction

In the era of the information age, data stands as the foundational resource, while algorithmic decision-making serves as the core engine. Virtually every sector now relies on algorithms, and individuals, without exception, benefit from the conveniences enabled by algorithms. Algorithms epitomize the pinnacle of technological and mathematical rationality, enabling people to reshape the material world more objectively and efficiently using data and algorithms. If we regard digital technology as the input of big data and the output of data analysis, then algorithms become the critical step that transforms input values into output values. Data possesses inherent objectivity and neutrality, yet algorithms designed by humans unavoidably carry biases.

Discrimination has persistently existed in human thought and consciousness and continually evolves with the changing times. Algorithmic discrimination, emerging alongside the development and application of new technologies such as artificial intelligence, big data, and algorithms, represents a novel form of societal discrimination. The digital realm mirrors the human society and the inescapable issue of discrimination in human society is equally present in the digital sphere, often manifesting in more complex ways.[1] Algorithmic discrimination's impact on societal institutions, economic, political, and cultural development is increasingly profound, extending beyond traditional factors such as race and gender to affect daily human life and societal affairs.

While algorithms have brought countless opportunities and great conveniences to individuals'

lives, they have also raised a series of potential infringements. This highlights the fact that discrimination has evolved into new forms with the changing times. The application of algorithms and big data has not necessarily resolved issues of justice and equality. In some instances, it has exacerbated injustices and inequalities, giving rise to new ethical dilemmas. The challenges and issues posed by digital technology necessitate legal responses.

This paper systematically discusses the reasons behind the formation of algorithmic discrimination, summarizes the regulatory challenges associated with it and proposes corresponding legal regulatory measures. The aim is to reshape justice and equality in the digital society and promote the responsible development of algorithmic technology.

2. The Analysis of the Factors Leading to Algorithmic Discrimination

2.1 Subjective Bias in Algorithm Design

Algorithms, fundamentally created by humans to enhance the convenience of production and daily life, inevitably bear the imprint of human thought. Algorithmic discrimination, as a derivative of digital technology, is evidently rooted in subjective biases shaped by human thinking, rather than being an inevitable consequence of algorithmic technological advancement. [2] Within the contemporary human society, biased ideologies persist. If algorithm designers harbor subjective biases and embed their personal inclinations and implicit biases into algorithms, it will invariably result in algorithmic discrimination issues. Algorithmic discrimination can originate from the subjective biases of algorithm designers, whether intentional or unintentional. Regardless of the designer's efforts to exclude personal values, their own design philosophies, criteria for algorithmic output, cultural backgrounds, religious beliefs, and the societal systems they inhabit will inevitably influence the development or configuration of the algorithm program. This is an unavoidable reality.

In addition to inherent ideological biases, profit-driven motives are also a significant cause of algorithmic discrimination. To maximize profits, algorithm designers may intentionally create discriminatory algorithms by offering more favorable conditions to attract users with higher perceived value. This undoubtedly leads to unfair treatment of other users. In comparison to inherent ideological biases, groups subjected to algorithmic discrimination due to profit-driven motives are more likely to break through discrimination barriers through consumer behavior. This is a primary objective behind the intentional design of discriminatory algorithms: attracting user attention, stimulating user consumption, and ultimately generating greater economic benefits for the designers.

2.2 Data and Technical Flaws in Algorithmic Programs

In addition to issues arising during the design process, data sets are one of the primary reasons behind algorithmic discrimination problems. Like the controversies surrounding discriminatory image-editing algorithms, if data sampling disproportionately favors certain groups during the data collection process while neglecting others, algorithms will only capture the characteristics of the sampled group.[3] Consequently, the information generated by the algorithm will only be applicable to the sampled group, leaving unsampled groups susceptible to discrimination. Therefore, even if algorithms are designed with fairness in mind, an uneven distribution of data samples can lead to severe algorithmic discrimination problems. Algorithms with learning capabilities not only reflect the ideologies of their designers but also incorporate user perspectives. Machine learning algorithms learn from user interactions and may store and apply learned user behavior information to similar interaction scenarios.

Although machine learning algorithms are initially designed to improve the precision of information and data output by learning from information and data, they often lack rigorous information discernment and filtering mechanisms. Discriminatory information cannot be identified and filtered out in advance during the learning process. Once an algorithm learns and stores discriminatory information during this process, it is highly likely to produce discriminatory outputs in other interaction scenarios, exacerbating algorithmic discrimination issues. As long as inherent biases in the data used by the algorithm remain unaddressed, these biases will persist throughout the algorithmic decision-making process.

2.3 Algorithmic Black Box Leading to Information Asymmetry

Deep learning algorithms, as a more advanced form of cognition, analyze the intrinsic properties of entities through induction and summarization, constituting a more sophisticated form of cognition. This cognitive mechanism results in algorithms possessing an understanding of data that developers cannot grasp or comprehend. Developers lack control over the algorithm's execution process and are unable to provide external explanations. The lack of transparency in algorithms makes algorithmic discrimination more covert. Currently, this covert cognitive mechanism cannot be effectively resolved through technological means.

Algorithm designers, equipped with extensive data, possess an information advantage, enabling them to easily differentiate between users. However, users are often unaware of each other's information. In most cases, users can only view the outcomes of the algorithmic operation they personally undergo, without insights into the algorithm's data utilization, analytical logic, and crucial processes.[4] This creates a situation where users may not even realize they have become victims of algorithmic discrimination. For instance, in employment contexts, certain demographic groups frequently experience differential treatment in job recommendations and screening through search engines. Similarly, in marketing, internet platforms use extensive data analysis to profile consumers based on their preferences and income levels. Subsequently, they implement differentiated pricing for the same quality products or services, depending on these factors. Users can only discover that they have fallen victim to algorithmic discrimination when they compare their experiences with those of other users. Information asymmetry renders algorithmic discrimination covert and challenging to detect, leading to instances where the results of algorithmic decisions conceal the outward expression of preexisting algorithmic biases.

3. Real-world Harms Resulting from Algorithmic Discrimination.

3.1 Algorithmic Discrimination Disturbing Market Economic Order

Unreasonable differential pricing in transactions often constitutes unfair pricing. China has established relevant legislation to address unfair pricing practices in transactions. Laws such as the "Anti-Monopoly Law of the People's Republic of China" and the "Price Law of the People's Republic of China" both stipulate that trading counterparts shall not engage in differential treatment in transaction prices and other trading conditions and shall not discriminate against other operators. With the widespread application of algorithms in the business sector, algorithmic platforms have gradually gained technological and market advantages.[5] To maximize their commercial interests, these platforms have begun to use algorithms to discriminate against consumers by implementing price discrimination. This brings offline price discrimination into the online realm. Algorithmic price discrimination involves using algorithmic technology to assess the payment ability and willingness of different consumers and then formulating sales strategies with differential pricing. Algorithmic price discrimination further amplifies the influence of offline price discrimination. It

severely violates the traditional business ethics standard of "clear and fair pricing," alters the basic rules of competition, market dynamics, and market economic order, and fails to fully respect and equally protect the legitimate rights and interests of consumers.

In response to this, the "E-commerce Law of the People's Republic of China" stipulates that e-commerce operators, when providing search results for goods or services to consumers based on their interests and preferences, should simultaneously provide options that are not based on the consumer's individual characteristics. This aims to maintain market economic order and protect consumers' legitimate rights and interests.

3.2 Algorithmic Discrimination Undermining Social Equity and Justice

The emergence of algorithmic discrimination often results from algorithms prioritizing efficiency over fairness when faced with the choice between the two. Today, automated algorithmic decision-making spans a wide range of areas. Algorithms are not only applied in commercial sectors such as housing rentals, financing, credit assessment, insurance, and recruitment but also play a role in formulating and implementing public policies and in the judicial domain. In these areas, once discriminatory data or algorithms are introduced, the resulting discriminatory consequences can be even more severe than in the business sector.[6] For example, in the judicial sector, the use of algorithmic programs to predict recidivism rates can easily lead to risk assessment results tainted by discriminatory ideologies held by algorithm designers or due to inaccurate algorithms. This not only jeopardizes fairness and justice but also undermines the credibility of the judicial system.

Typically, groups subjected to discrimination often experience a vicious cycle due to their outsider status, resulting in a situation where the strong become stronger, and the weak become weaker." From the perspective of resource and benefit allocation, algorithmic discrimination seriously disrupts distributive justice. This applies to both the allocation of material resources and the distribution of labour and social systems. Society bears the heavy burden of severe discrimination, affecting social equity and justice negatively. Therefore, addressing these issues both technically and legally is imperative.

3.3 Algorithmic Discrimination Violating Citizens' Equal Rights

Equality is a cherished value and a cornerstone of law. Equality, as a fundamental right, has been enshrined in the constitutions or laws of most countries. Constitutional equality prohibits any unreasonable differential treatment based on certain inherent natural or social characteristics of individuals or groups, which aims to impair, diminish, restrict, or deprive others of their legal rights under identical circumstances. It also prohibits any measures that have the legal effect of harming specific groups or individuals based on unreasonable distinctions. The prohibition of discrimination is a central component of the right to equality. Discrimination victims often experience unreasonable differential treatment under similar circumstances, which seriously infringes on citizens' right to equality. Algorithmic discrimination, to some extent, institutionalizes discrimination and inequality within society.[7] The root cause of gender or racial discrimination, for example, often lies in the existence of algorithmic biases within artificial intelligence systems that result in differential treatment of certain individuals or groups.

In the rapidly evolving era of artificial intelligence, the utilitarian nature of algorithmic biases can accelerate and even magnify infringements on the right to equality among members of society. For instance, in the 2015 case where Google's image recognition algorithm mistakenly classified Black people as gorillas, the algorithmic system's biased orientation led it to misidentify Black individuals based on pre-existing algorithmic biases. This incident vividly illustrates how

algorithmic biases can infringe on the right to equality of Black communities. Despite Google's subsequent measures to filter related terms, the bias within the system was not eliminated and continues to deepen over time.

4. Constructing the Accountability Path for Algorithmic Discrimination.

4.1 Determining the Constituent Elements of Algorithmic Discrimination

Algorithmic discrimination, as an emerging form of infringement, exhibits diversity and complexity. It is not entirely feasible to apply the traditional elements of liability found in conventional tort law to determine its existence. However, it is also essential to stay rooted in the fundamental principles of tort law rather than creating an entirely new set of criteria for recognition. To address the issue of establishing the elements of liability in algorithmic infringement, it is necessary to expand or restrict the scope and definition of these elements appropriately.

First, we can expand the definition of behavioral elements. It should not consider behavior only as conscious actions performed by living things but should cover non-living and unconscious behaviors. Even though algorithms lack autonomous consciousness, they still reflect the intentions of their designers. Algorithms are fundamentally tools with the underlying intentionality of their creators.

Second, we need to limit the scope of harm. Algorithmic discrimination often results in damages that are difficult to ascertain and may not necessarily be remediable. When a victim alleges that algorithmic actions have violated their rights and caused harm, the burden of proof should not excessively demand conclusive and precise evidence of harm.[8] Initial proof of the existence of harm should suffice.

Third, we should apply the principle of reversal of the burden of proof. The complexity, specialization, and opacity of algorithms make it challenging for victims to prove a causal relationship between their harm and the operation of algorithms. As direct users of algorithmic outcomes, it is unreasonable to expect users to possess significant knowledge of computer algorithms. In cases of algorithmic infringement, the burden of proving causation, which may be difficult for victims to establish, should be shifted to the developers and controllers of the algorithmic program.

Fourth, developers and owners should prove their subjective status. The manifestation of will in artificial intelligence algorithms is essentially a reflection of human intentionality wearing the cloak of artificial intelligence. In the case of algorithmic infringement, proving subjective negligence, whether intentional or negligent, should not require proof of the subjective state of consciousness of the autonomous algorithm. Instead, it should involve establishing the subjective intent of the developers and owners of the algorithmic program.

4.2 Clarifying the Responsible Parties for Algorithmic Discrimination

From the perspective of traditional tort law, it is essential to identify the parties responsible for bearing liability in any infringement. Accurate identification of responsible parties is crucial for seeking effective remedies. Given the attribute of technological neutrality associated with algorithms, algorithms themselves cannot be the primary subjects of liability under tort law. Therefore, it is imperative to determine the responsible parties for algorithmic discrimination as network service providers, owners of algorithmic programs, and developers of algorithmic programs.

Firstly, network service providers should be held accountable for algorithmic infringement. It is argued that algorithmic infringement should be subject to the safe harbour principle. Following the

occurrence of algorithmic discrimination, network service providers become liable when they fail to exercise reasonable care and can be deemed to have played a certain facilitating role. Secondly, according to relevant tort law theories, when harm results from an infringement, the victim has the right to seek compensation from the person responsible for committing the infringement or the owner or manager of the infringing item. Computer software algorithms exist tangibly in the physical realm. Therefore, the harm caused by algorithmic programs should logically be borne by their owners in terms of liability under tort law.

Lastly, developers of algorithmic programs, possessing the most profound understanding of the operation, data acquisition, and attributes of the algorithmic program, should assume additional responsibility under certain conditions. All algorithmic programs inevitably reflect the subjective value judgments of their developers.[9] When it comes to developers, they should not be equated with manufacturers, and the principle of strict liability should not be automatically applied. Instead, the doctrine of negligence should be invoked. To establish that developers of algorithmic programs are at fault, it is necessary to demonstrate that they could have foreseen the circumstances leading to harm caused by the algorithmic program and that the developers did not consider these circumstances during the design and programming phases. Thus, developers of algorithmic programs should bear corresponding liability if they cannot prove that they acted without negligence regarding the harm caused by their developed algorithms.

4.3 Analysing the Modes of Liability for Algorithmic Discrimination

The operational characteristics of algorithms, characterized by their opacity and complexity, present a real challenge when attempting to determine both the existence of algorithmic infringement and the modes of liability. One fundamental reason for the difficulty in attributing responsibility lies in the lack of clarity regarding the responsible parties. If algorithmic infringement is to establish its own criteria for determining liability and regulatory frameworks, its modes of liability should naturally align with traditional ones found in tort law.

Therefore, the modes of liability for algorithmic infringement should be based on the categories established in traditional tort law, with some degree of discretionary flexibility corresponding to the severity of the harm resulting from the infringement. The primary modes of liability for prosecuting infringement should include cessation of infringement, removal of obstacles, elimination of dangers, and compensation for losses. In cases where algorithmic discrimination jeopardizes the personal safety and property of individuals, those who have suffered harm have the right to demand that the responsible parties undertake these modes of liability. In the context of algorithmic infringement, it is essential to apply the concept of compensation for harm. Given the multifaceted and covert nature of algorithmic infringement caused by the technical characteristics of algorithms, establishing the scope of actual damage can be challenging, and its duration may be prolonged.[10] This can complicate the assessment of tangible losses. Therefore, discussions of compensation for harm in the context of algorithms should encompass not only material damages but also psychological harm and potential latent damage. For instance, when algorithmic data collection infringes upon personal information rights, it is often difficult to quantify the tangible losses. In such cases, it may be necessary to introduce compensation for psychological harm.

4.4 Establishing a Mechanism for Allocating Responsibility in Algorithmic Discrimination

In the aftermath of any infringement, it is indispensable to safeguard the legitimate rights and interests of the victims. To achieve this, a comprehensive mechanism for allocating liability is essential, as it enables victims to pursue effective remedies to protect their rights and helps mitigate their losses. To address the issue of determining the responsible parties and the appropriate modes

of liability for algorithmic infringement, it is essential to enhance the mechanism for allocating liability in algorithmic discrimination. This should involve the establishment of a dual-track system of liability that encompasses platform liability and technical liability, effectively addressing the roles of both algorithm developers and platform operators.

In the current legal framework in our country, platform liability is primarily emphasized, wherein platforms that employ algorithms are held responsible for harmful outcomes resulting from the use of algorithms.[11] However, this approach overlooks the fact that algorithms have transcended their conventional role as mere tools in the era of big data. It also neglects the designers of algorithms as a pivotal component of the equation. Therefore, there is a pressing need to transition from a single-track system of accountability, which focuses solely on platform liability, to a dual-track system that emphasizes both platform liability and technical liability.

In this dual-track system, platform responsibility and technical responsibility respectively correspond to the protection of users' legitimate rights and interests and the need to let stakeholders understand the algorithm decision-making process. On one hand, this system imposes certain restrictions on algorithm developers. On the other hand, it ensures that stakeholders affected by algorithms are not left in a state of prolonged algorithmic infringement. By clearly defining the responsibilities of both algorithm developers and the platforms utilizing algorithms within this dual-track system, it establishes a reasonable model for the allocation of liability in algorithmic infringement cases. This model is conducive to addressing the complex challenges posed by the algorithmic era and fosters greater public trust in algorithms.

5. Conclusions

With the advent of the algorithmic era, various aspects of people's lives are increasingly subject to algorithmic decision-making. The existence of algorithmic discrimination poses a threat to the legitimate rights and interests of civil subjects. The proactive explorations and beneficial practices that various countries have undertaken to regulate algorithmic discrimination are worthy of study and reference. China also attaches great importance to the harm of algorithmic discrimination and its legal governance.

From the perspective of proactive prevention, legislation in China has clarified relevant principles and rules, providing a legitimate basis and a protective environment for the legitimate rights and interests of civil subjects.[12] Other scattered laws and regulations have also indicated directions and paths for remedies against algorithmic discrimination infringement. To mitigate the harm caused by algorithmic discrimination, it is necessary to establish a comprehensive framework for liability within the context of algorithmic discrimination. A complete legal process to address algorithmic infringement risks should be established promptly. This can be achieved through legislative guidance on algorithm applications, the establishment of a system for sharing algorithmic infringement liability, and other measures to address the current difficulties in regulating algorithmic infringement. These efforts will help to effectively remedy the harm of algorithmic discrimination, bring the development and application of algorithmic technology under the rule of law, protect the equal rights of individuals in society, safeguard the values of equality and justice, and at the same time reduce or overcome the risks caused.

Acknowledgement

References

[1] Shi Ying. The Origin, Challenge and Legal Response of Algorithm Discrimination [J]. Journal of Gansu University of Political Science and Law, 2022(3):58-68.

[2] Liu Yuzhuo. The Constitutional Function of Equal Rights in Governance of Algorithmic Discrimination [J]. Science Economy and Society, 2022(1):66-79.

[3] Yang Yahan. Research on Typological Infringement of Improper Application of Algorithms in the Age of Artificial Intelligence and Exploration of Differentiated Regulation Paths [J]. Industrial Technology Innovation, 2022(4):79-82.

[4] Wang Ying. On the Framework of Algorithm Infringement Liability [J]. Chinese Law, 2022(3):165-184.

[5] Li Wenchao, Wu Yifan. Research on Prior Regulation and Infringement Remedies for Algorithm Infringement Behavior [J]. Application of Law, 2023(3):119-128.

[6] Zhang Hui. Specification and Technical Analysis of Autonomous Algorithm Privacy Protection [J]. Lanzhou Academic Journal, 2021(3):120-135.

[7] Song Baozhen. The rights of "digitally disadvantaged groups" and their legal protection [J]. Legal Science (Journal of Northwest University of Political Science and Law), 2020(6):53-64.

[8] Zheng Yushuang. Solving the problem of technological neutrality—Rethinking jurisprudence on the relationship between law and technology [J]. Journal of East China University of Political Science and Law, 2018(1): 85-97.

[9] He Guoqiang. Risk society, risk allocation and tort liability law reform [J]. Guangdong Social Sciences, 2018(3): 228-237.

[10] Shen Weiwei. The Myth of the Algorithmic Transparency Principle—Criticism of Algorithmic Regulation Theory [J]. Global Law Review, 2019(6):20-39.

[11] Wang Liming. New challenges to civil law in the era of artificial intelligence [J]. Oriental Jurisprudence, 2018 (3):4-9.

[12] Zhang Linghan. Platform algorithm accountability and improvement in the "Personal Information Protection Law (Draft)" [J]. Economic and Trade Law Review, 2021(01):36-46.