ISSN 2616-7433 Vol. 5, Issue 1: 68-72, DOI: 10.25236/FSST.2023.050113

Legal Governance of Discrimination Risks in Algorithmic Automated Decision Making

Wu Xue

School of Law, Shandong University of Technology, Zibo 255049, Shandong, China

Abstract: *With the booming development of digital technology, algorithmic automated decision making is applied in both public and private fields because of its efficient data collection and processing capabilities as well as certainty and uniformity. The wide application of algorithmic automated decision-making brings great convenience to personal life, business operation and public governance, but also brings the risk of algorithmic alienation and discrimination. In order to regulate the discrimination risk of automatic algorithm decision making, the first thing is to adhere to the consciousness of human's subject status. Automatic algorithm decision making is still a kind of intelligent advanced tool serving human beings. On this premise, the transparency of decision making is improved and the relief method of responsibility distribution after automatic decision discrimination infringement is clarified.*

Keywords: *algorithmic automated decision making; discrimination; coping strategies*

Nowadays, the profound impact of the information revolution on human beings has far exceeded that of the agricultural and industrial revolutions, and China is accelerating the construction of a "digital China". The widespread use of digital technologies such as the Internet, big data and cloud computing is affecting the way of life and survival of human beings, as well as reshaping the operation of state organs in an increasingly in-depth digital way. Algorithms are profoundly affecting the production, social and legal relations of intelligent society, and human beings are quietly entering the "algorithmic society". Compared with traditional manual decision-making, algorithmic automated decision-making can improve service quality, provide more personalized services, shorten processing time, and allocate resources more effectively, these characteristics make it widely used in the development of various sectors of society.

1. Algorithmic automated decision making and discrimination

With the increasing development of big data and artificial intelligence, algorithms are given a new connotation. Algorithms go beyond the form of mainly computer programs, and have the ability to process information on the basis of big data collection and autonomous machine learning, and have powerful prediction ability and decision-making ability. The closely related algorithmic automated decision-making refers to the process of algorithms analyzing and processing the input data and making decisions by automated means, and these processes do not require human participation, and algorithms with decision-making capabilities are no longer pure tools.

Discrimination is usually considered to be discrimination, which is opposed to equality and justice, but it does not mean that the behavior that satisfies discrimination is discrimination. The law does not oppose reasonable, relevant and necessary differences, but only prohibits unjustified differences. If the difference is justified and serves a legitimate purpose, it does not constitute discrimination, but a reasonable difference permitted by law[1]. The distinctive feature of discrimination is unreasonable and unjustified differentiation, which is a lack of respect for individual rights. Thus, legal discrimination is any unjustified differentiation in treatment of a specific group or individual that is prohibited by law and is aimed at derogating from, limiting or depriving them of their legal rights.

The powerful algorithm cannot conceal the essence of discrimination, and the discrimination of algorithmic automated decision making is not a product of technological innovation, but a new expression of traditional social conflicts shaped by digitalization. When the "decision maker" automatically analyzes and processes the input data types through algorithms, the output results with different standards of subjective discrimination are produced. The risk of discrimination by algorithmic automated decision making is still different from the risk of traditional social discrimination, which relies on artificial intelligence technology and presents characteristics of the times such as mechanization, tolerance, and difficulty of regulation in the specific expression of discrimination.

Algorithm automatic decision discrimination has mechanism. Although the human brain decision-making in the past has been unable to exclude discrimination, control and unfair value preference, it is basically individual, random and constantly changing, while the algorithm decision-making is universal and continuous. If the system itself contains discriminatory factors or the input data contains discriminatory factors, it will inevitably output discriminatory decisions, which may be upgraded and extended through independent learning and automatically form discrimination against non-specific people. In other words, once discrimination occurs, social inequity will be systematically and routinely caused. Lead to a wider range of institutionalized infringement results.

Algorithmic automatic decision discrimination has tolerance. The wide application of automatic algorithm decision-making makes the public increasingly rely on algorithms for analysis, prediction and decision-making. For example, when shopping in Douyin、Toutiao and Taobao, personalized recommendation enables users to obtain the information and services they need in the first time, which not only saves time and improves efficiency, but also gives users a full sense of user experience. People will continue to use these algorithms to enjoy convenience even when they know that there may be differential pricing and discrimination risk. When people give up some existing rights and interests in order to enjoy convenience, passively immersed in the convenience brought by technology, this makes users have more tolerance to the potential discrimination risk of algorithm automatic decision-making system than before.

Algorithm automatic decision discrimination regulation is difficult. The difficulty of monitoring whether there is discrimination risk in algorithm automatic decision-making lies in that the data and algorithm rules are mastered by designers and users, so it is difficult to obtain external supervision. Even if they can be obtained, non-professionals cannot judge whether there is discrimination because automatic decision-making is written by a professional algorithm language. Internally, users will not release data or run programs in order to protect "trade secrets" for maximum profit. In addition, when discrimination risks occur, the actual operators of the algorithm automated decision-making system may also use the black box of the decision-making process to claim that the decision-making behavior is automatically made by the algorithm and advocate technology neutrality to evade responsibility, which further increases the difficulty of supervision.

2. The causes of algorithmic automated decision discrimination are explored

2.1. The original "black box" nature of algorithmic automated decision making

In cybernetics, the unknown region or system is usually referred to as a "black box", and the algorithmic automated decision itself is like an unknown "black box" - the user has no way to know the inner mechanism of the decision output, and it is not clear how the decision After the decision is generated, it is not clear how the algorithm designer or the actual controller is responsible for it. In other words, we can directly observe the input data and the output of the algorithm after analysis and processing, but we do not know how the decision is analyzed inside the model and even the output provided is uninterpretable. The "black box" nature of the algorithmic decision-making system is externalized by the fact that the automated decisions are made quickly and intrinsically by the algorithm, while the operational process is ambiguous and not intuitively perceivable, which is likely to bring about invisible discrimination. In fact, "black box" is not necessarily negative, it is primarily a neutral descriptive expression, and human decision makers are as black boxes as the algorithms that are meant to replace them[0].

2.2. Lopsided data collection

According to the meaning of algorithm mentioned before, we can see that if there is no input data, it will not automatically generate decisions, only input data, the algorithm can calculate and analyze to establish the instruction model to solve a specific problem. In the automated decision making based on big data and machine learning systems, the model needs to be supported by a large amount of data, i.e., a large amount of input data for "learning", and this learning process is also the process of algorithm training. Each group of training data will have a clear set of corresponding result data or judgment data. When building the prediction model, the actual result obtained from each input training data will be compared with the expected prediction result, and the algorithm will reinforce the learning to adjust the prediction model automatically, after repeated training, until the training result of the model reaches an expected accuracy rate, the algorithm has built the preliminary model. The input data can be used to obtain matching decision results. Therefore, one of the major causes of discrimination in automated decision making is the discrimination in the training data or input data provided, i.e., the so-called "garbage in, garbage out", and it is impossible for an algorithm that contains discrimination to output fair and just decisions.

2.3. Pre-existing prejudices

Algorithm is a mathematical program or computer code to express opinions, any algorithm is designed by the designer to achieve a certain purpose, it reflects the designer's will , its design philosophy, data use norms, decision output standards, etc. are the subjective choice of the developer , algorithm automated decision-making system also reflects the subjective thinking of the designer, developer, if the designer at the beginning of the design will be their own values Even after continuous learning and reinforcement, this inherent discrimination cannot be eliminated, that is to say, even if the data input is clean, after the analysis and processing of the automated decision model with the discrimination gene, the output data will be contaminated with the discrimination gene decision. Behind the technology must stand technical experts or specific interest groups, it can be seen that the quality and value level of the designer has an important impact on the equality and fairness of the output of the algorithm automated decision making. For example, COMPAS, a crime-risk assessment algorithm used in America's federal criminal-justice system, has an element of discrimination against blacks. In its analysis of non-reoffending reports, blacks were more than twice as likely as whites to be labeled by COMPAS as having a moderate or high risk of reoffending (42% versus 22%)[0].

2.4. Discrimination arising from acquired learning

Discrimination may also be the result of acquired learning, and many algorithm-related ethical issues are built on the characteristics of machine learning algorithms and deep learning algorithms , with the increasing maturity of big data algorithm technology, deep learning algorithm technology has been widely used, compared to machine learning without prior construction of models, its random generation of neurons and weights, according to the input data for self-training, without monitoring. In traditional algorithms, the one-to-one correspondence between data and rules needs to be done manually by the designer, and the algorithm will only process data according to the pre-settings without learning ability. In the deep learning stage, humans only need to input practice data with features and inform the output results corresponding to them, and the algorithm can construct the model by itself, and this learning process is not limited to the practice stage, but also applicable in the application stage. With the increase of input data in practical applications, the model is automatically optimized and improved by continuously drawing information from relevant data, but the algorithm itself does not understand the relevance and risk of alienation because the algorithm itself mines new data with discrimination factors, and the model automatically learns into a discrimination decision system.

In summary, the current algorithmic automated decision-making system is not absolutely neutral and can help human beings get rid of the risk of discrimination completely, and even itself may be influenced by various discrimination phenomena and information to make discriminatory behaviors. We should acknowledge the limitations of automated decision making, but the risk of discrimination is not inherent and unavoidable, and we should adopt targeted strategies to deal with the causes of discrimination risk from automated decision making.

3. Strategies for coping with the risk of discrimination in algorithmic automated decision making

3.1. Adhere to the main position of human

We generally believe that human beings should be the final decision makers in all important matters, and that human choices are an integral and fundamental part of private and public life[0]. Artificial intelligence is still a human tool, at most an upgraded version of the intelligent tool that can "dance with humans and machines", so the current artificial intelligence is still characterized as a highly automated, intelligent tool created by humans. Although the development of algorithmic automated decision-making to a certain extent to replace the human automatically make decisions, but it is ultimately not human, is still a kind of intelligent advanced tools for human services, its development should be human-centered, people-oriented, respect for human dignity and value of the principle of priority. Due to the current level of technology, artificial intelligence is still unable to encode the unique human emotional values such as empathy and moral sense into computer language for use in algorithmic automated decision making, therefore, the subject of law is of course human. It is important to clarify the subject status of the users of algorithmic automated decision-making and the attributes of the tools of automated decision-making for the reasonable and scientific formulation of relevant legal norms and supporting facilities. Therefore, the application of algorithmic automated decision-making should set strict restrictions and legal reservation of the restricted area, especially the important matters about life, human dignity and value, race, religion, etc. should not be easily left to the algorithmic automated system to make decisions, and even if it is left to the decision-making system to decide, strict procedures and strict legal framework should be established, so that timely remedies can be made when the decision is discriminatory infringement.

3.2. Improving the transparency of automated decision-making

"The complexity of algorithms and the opaqueness of technology lead to algorithmic discrimination in automated decision making. The improvement of the transparency of algorithms in data processing can effectively curb the occurrence of algorithmic discrimination. However, only after understanding the design principle, operation rules and decision formation logic of the automated decision-making system, users can know what can be done and what cannot be done, the promotion and potential risks to users' interests can regulate the automated decision-making process, and the external behavior and results of the decision can be attributed. There are two main reasons for non-transparency: the owners and users of automated decision making often equate algorithm transparency with exposing trade secrets and believe that it will destroy their competitive advantage, so users and owners tend to limit the transparency of algorithms; secondly, users and the public cannot understand the technical aspects of the code and the resulting non-transparency.

In response to the first kind of opacity arising from deliberate secrecy, some people worry that if the law compulsorily discloses the whole process of algorithmic automated decision making will lead to the infringement of trade secrets and patent protection, which is equivalent to enterprises making the results of their own hard research public. In order to realize this principle, mandatory information disclosure and information sharing mechanisms need to be established under the framework of legal regulation, and the operation of algorithmic systems should be publicized in a timely, open and continuous manner. For the second type of algorithms, which are opaque due to their complexity and incomprehensibility, mandatory disclosure by law or not is of limited use. Due to the level of technology and social structure, reading and understanding the code of algorithmic automated decision making is still a specialized skill, and even if the user obtains sufficient information related to the algorithm automatic decision-making, the data subject is unable to overcome the technical obstacles to effectively exercise the rights due to the limitation of time, resources and the lack of necessary professional knowledge[0]. Therefore, it is possible to set up a special regulatory body for automated algorithmic decision making, which will be staffed by professionals who will determine whether there is discrimination in automated algorithmic decision making.

3.3. Confirmation of the mechanism for allocating responsibility for

In terms of the principle of attribution of discrimination in algorithmic automated decision-making, the information of the rights of the designers, users and users of algorithmic automated decision-making is in a serious asymmetry, and there exists a natural even barrier, and the no-fault principle as the principle of attribution of discrimination in algorithmic automated decision-making can make the risk burden more reasonable and maximize the protection of users' rights. It is impractical to judge the liability based on the subjective fault of the designer and user of the automated algorithm decision, and we should focus on the dangerousness of the actor. Therefore, as long as the output decision contains discrimination, it can be considered that the algorithm has produced infringement damage, and the subjective existence of fault of the designer and user of the automated decision of the algorithm is not considered in the scope of liability for damages, that is, the designer and user of the automated decision of the algorithm bear no-fault responsibility for the consequences of discrimination caused by the automated decision.

Algorithm automated decision-making is developed and designed by the designer of the algorithm, is put into operation by the user of the algorithm, so the designer and user of the algorithm is the main responsibility to bear the main body, the algorithm of automated decision-making discrimination infringement bear the corresponding infringement responsibility. The object of analysis of algorithms is data, and data is ultimately a mirror image of society. If society itself is harmonious and free of discrimination, naturally the data algorithm will not be manifested. If the designer of the algorithm automatic decision has embedded the discrimination risk into the program at the beginning of the design, the designer should bear the no-fault liability for the infringement result. If the designer cannot detect and exclude the discriminatory gene in the algorithm according to the existing technical level when designing the algorithm automatic decision, but the discriminatory gene appears in the operation and use, in this case, the user shall bear the main infringement liability, and the designer shall bear the supplementary liability. Finally, if the infringement damage is intentionally caused by the victim, the designer and the user of the algorithm should not be responsible. Finally, if the infringement damage is intentionally caused by the victim, the designer and the user of the automated algorithmic decision making are not responsible, except for the intentional infringement by the infringer, which, according to the general view, can only reduce but not exclude the liability of the aggressor.

4. Conclusion

Discrimination in human society will not disappear, and algorithmic automated decision making is ultimately an advanced tool for humans, from humans, and the risk of discrimination in automated decision making will not be completely excluded. In order to regulate the risk of discrimination in automated decision-making, we should insist on the status of human subjects, and that discrimination is the product of interaction between humans and humans, not between humans and technology. Regulation of algorithmic automated discrimination in the application of specific areas should be distinguished and moderate, taking into account both the maintenance of individual rights and the operability of measures, future industry development, constantly reviewing algorithmic automated decision-making technology, enhancing their own risk prevention capabilities, and maximizing the protection of the legitimate rights and interests of users.

References

[1] Niyi, Basukoski Artie, Chaussalet Thierry," (2021). An Exploration of Ethical Decision Making with Intelligence Augmentation", Social Sciences.

[2] Bonezzi, A., Ostinelli, M., & Melzner, J. (2022). The human black-box: The illusion of understanding human better than algorithmic decision-making. Journal of Experimental Psychology: General.

[3] Tolan, S. (2019). Fair and Unbiased Algorithmic Decision Making: Current State and Future Challenges. 10.48550/arXiv.1901.04730.

[4] Gal, & Michal, S. . (2018). Algorithmic challenges to autonomous choice. Michigan Technology Law Review.

[5] Lilian, Edwards, Michael, & Veale. (2017). Slave to the algorithm? why a 'right to an explanation' is probably not the remedy you are looking for. Duke Law & Technology Review.