ISSN: 1813-4890

Anti Algorithmic Discrimination from a Legal Perspective: Enlightenment from American Experience to China

Chuyun Wang[1] , Lei Li [2,*], Ziyun Zhao[3]

[1]School of Law, Guangzhou College of Commerce, Guangzhou, Guangdong, 511363, China

[2]Department of International Law, China Foreign Affairs University, Beijing, 100037, China

[3]College of Business and Economics, Australian National University, ACT Australia

Abstract

Algorithms are based on big data and machine learning, and have strong self-learning functions. They can continuously learn and extract rules from massive data, and then form automated decision-making. They have more and more extensive application prospects in human society. However, the complexity and opacity of the algorithm also bring many problems. More and more practical experience shows that the seemingly neutral algorithm can also lead to discriminatory results, and the problem of algorithm discrimination needs to be solved urgently. Whether it is to make technical improvements to the algorithm model or propose legal regulations to solve the problem of algorithmic discrimination, the first basic problem to be faced is the identification of algorithmic discrimination. However, the current legal identification of algorithmic discrimination is facing difficulties. On the one hand, algorithmic discrimination is concealed, and on the other hand, China's current anti discrimination legal norms are relatively general and not very maneuverable. In this regard, this paper draws on the beneficial experience of the anti discrimination law of the United States, and puts forward suggestions on the review standards of algorithmic discrimination and the construction of the legal system of anti algorithmic discrimination in China and other countries.

Keywords

Big data; automated decision making; algorithmic discrimination.

1. Introduction

The "algorithmic discrimination" discussed in this paper can be defined as the unreasonable differential treatment implemented by algorithmic automated decision-making systems based on the division of target groups in the context of big data and artificial intelligence. Algorithms are not completely objective and neutral as people usually think, and more and more practical experience shows that seemingly neutral algorithms can also lead to discriminatory results. If the algorithm discrimination is not regulated, the systematic linkage of algorithmic decision-making will make the bias and discrimination in the algorithm model continue, so as to further consolidate the discrimination and aggravate the social inequality. In order to defend equality, human dignity, and break the stereotype of class, it is urgent to build and improve China's anti-algorithmic discrimination legal system based on national conditions.

The United States paid attention to algorithmic discrimination earlier. The White House published an article "Big Data: Seizing Opportunities, Preserving Values" in 2014, which pointed out that, on the one hand, under the guise of neutral algorithms, deliberate discrimination against the most vulnerable social strata may occur, on the other hand the use of big data can lead to inadvertent discrimination due to its inherent characteristics.[1] Subsequently, in 2016, the White House released “Big Data: A Report on Algorithmic Systems,

ISSN: 1813-4890

Opportunity, and Civil Right"[2], which analyzed the opportunities and challenges brought by big data and pointed out that big data and algorithms may make discrimination more serious and difficult to detect. Scholars in the United States have also started early and are leading the way in research on algorithmic discrimination. Prof. Danielle Keats Citron and Prof. Frank Pasquale also pointed out in the article "The Scored Society: Due Process for Automated Predictions" that the use of algorithms makes all kinds of discrimination more subtle, which in turn will arbitrarily deprive some people of important opportunities. Therefore, the algorithm should be supervised by due process when predicting individual behavior and making decisions. The authors advocate giving individuals the right to inspect, revise, and dissent and other rights, and advocate the disclosure of algorithms, which should be subject to audit supervision. They have repeatedly emphasized that improving transparency is a crucial first step in the process compliance supervision of algorithms, and opening the black box to individuals or neutral experts on their behalf is the key to allowing them to challenge unfair decisions of algorithms.[3]

In recent years, the problem of algorithmic discrimination has also attracted the attention of Chinese scholars. They have analyzed the adverse effects of this problem and proposed regulatory schemes for algorithmic discrimination from different dimensions.Combined with the concealment of algorithms, scholars have mainly proposed measures such as algorithm interpretation, algorithm review, and algorithm accountability, providing macro and framework solutions.[4-6] Some scholars have also conducted research from the perspective of anti-discrimination.[7-9] On the basis of these research results, the author will focus on further research and write this article under the framework of anti-discrimination law combined with China's national conditions.

2. Causes of Algorithm Discrimination

The risk of discrimination may exist in all aspects of machine learning. The value judgment of algorithm developers always runs through the whole process of machine learning, and data, as the basis of machine learning, comes from human society, and the implicit social discrimination will also be brought into the algorithm model. Exploring the causes of algorithmic discrimination requires detailed analysis in conjunction with the working principles of machine learning.

2.1. Define Problem

The first step in machine learning is to determine exactly what you want to predict, that is, algorithm developers first need to define the problem. According to whether the problem needs to be defined clearly or not, it can be divided into structured problems and unstructured problems. Structured problems usually have a definite answer, such as calculating the area of a circle based on its radius, which are deterministic. Therefore, the formulas used to solve structured problems usually remain the same and do not change with different input variables. While for an unstructured problems, which may have multiple correct answers, the formulas used to solve the problems and the resulting answers are often not unique and vary depending on the input data.[10]

In real life, a large number of problems that algorithms need to solve are highly complex unstructured problems. For example, scoring personal credit is a non structural problem. There is no single rule for predicting a borrower's ability and willingness to repay, and it is crucial to have an appropriate and quantifiable definition of this type of problem.

In order to convert an unstructured problem into a computer-analyzable problem, the algorithm developer must convert the problem to be defined into a problem about some target variable. This process is open, and the target variables can be predicted based on the

ISSN: 1813-4890

understanding of the project objectives and requirements, referring to the past results or different data characteristics. Taking personal credit score as an example, it is not clear how to define personal credit, because there is no unified and fixed standard for the definition of credit worthiness, but it is built by the credit industry to predict the borrower's repayment ability before the loan is issued. Therefore, for algorithm developers, there are many possibilities for how to define credibility and establish target variables for prediction.

These different data features are often referred to as class attributes[10], and the process of labeling class attributes is either objective or subjective. For example, customers can be divided into two categories according to whether they respond to targeted advertisements. This criterion is objective. However, when assessing whether a job seeker is suitable for a particular job, it is subjective and diverse to label the job seeker. The judgment of the developer will affect the determination of such labels, and there may be discrimination against certain groups in the process.For example, making hiring decisions based on years of service may reduce employment opportunities for certain groups, such as married and pregnant women, even if they are as capable as other job seekers.

Identifying target variables and their associated class labels is an extremely important step in machine learning, which will determine what data mining will find. Transforming a complex problem into quantifiable target variables and class labels is challenging and subjective, and data miners may intentionally or unintentionally disadvantage protected classes. In addition, the target variable is based on a certain sample. If the initial data set cannot reflect the same feature distribution in a wider population, the sample cannot represent all potential predicted objects, which will lead to the problem of insufficient inclusiveness.[11]

2.2. Collect and Preprocess Data

After the target variables and class labels are defined, all kinds of relevant information will be collected, which will become training data. The quantity and quality of training data will determine the effectiveness of machine learning. If the collected data is biased, the machine may learn the bias and reflect it in the final algorithm system, that is, the so-called "bias in, bias out".[12] In addition to inheriting the bias in the previous data, the bias existing in the user's ongoing behavior will also become the input data. This is discovered by Professor latanya Sweeney in a study on online advertising in the Google search engine, that is, Google will predict the probability of users clicking on advertisements according to the historical Click through rate of different advertisements, so as to determine the ranking of different advertisements. In order to cater to the user's preferences, this kind of advertising may reproduce the user's existing prejudice[13], that is, prejudice and discrimination in the learning process.

In addition, not all data are collected equally. Professor Crawford pointed out that big data sets have "signal problems"[14]. In the era of big data, the collection of data is very dependent on the intelligent devices in the hands of individuals. Social network data, web browsing data, online shopping data, etc. are all important components of the data set. However, in some groups, especially vulnerable groups, such as those who are older or live in remote areas and are not well-off, the penetration rate of intelligent devices is low, and they rarely generate data, whom cannot be fully represented by the data set and will be systematically missed.

There are also risks in the process of data preprocessing. One of the characteristics of big data is that the information density is very low, and problems such as incorrect data and data loss may occur in the process of collection. In order to standardize the data and put it into the database in an appropriate form, it is necessary to preprocess the collected data, including making the format correct, making up the missing value, processing the deviation value, and converting the data[15]. This process needs to be carried out by algorithm developers according to their experience. There are subjective factors. If there are deviations or errors, bias or discrimination will occur.

ISSN: 1813-4890

2.3. Select Feature

Feature selection is an important part of machine learning, and not every input in the training data is relevant to what is being predicted. As the machine learning process progresses, the features most relevant to the defined problem need to be selected, given greater weight, and incorporated into the final model.

Algorithm developers may introduce personal bias or discrimination into the feature selection process, which may lead to the emergence of discriminatory risks. And because of the protection of trade secrets, we have no way of knowing which features are most important to the forecast and how the forecast is calculated.

3. Review Experience of Anti Algorithmic Discrimination Cases in the United States

With the further development of artificial intelligence technology and algorithms, discrimination in algorithmic decision-making is becoming more and more hidden, and it is more difficult to identify discrimination. At the same time, China's laws have not yet given a clear definition of algorithmic discrimination, and there is a lack of clear identification requirements. This chapter mainly summarizes the useful experience in the legal recognition of algorithmic discrimination in the United States, hoping to provide reference for the legal recognition of algorithmic discrimination in China. In judicial practice, disparate treatment and disparate impact are the two basic standards for U.S. courts to review algorithmic discrimination cases.

3.1. Disparate Treatment

Formal equality emphasizes that people cannot be divided into different categories and treated differently based on gender, race, color, religious belief and other factors.[16] Based on the value goal of formal equality, the review of discrimination follows the judgment standard of "disparate treatment", that is, different treatment is given to people in similar situations, and there is subjective intention of discrimination, which constitutes discrimination. In this review mode, the subjective intention of discrimination is the premise for the developer or user of the algorithm to bear the responsibility.

When the plaintiff claims that the defendant's behavior constitutes discrimination, it needs to prove that the defendant has the intention or motive of discrimination. The plaintiff can provide direct evidence, that is, evidence sufficient to prove the defendant's intention of discrimination without inference or hypothesis, such as the restriction on gender or race on the defendant's job advertisement. However, in reality, such blatant discrimination is relatively rare and it is not easy to obtain direct evidence. If the plaintiff is unable to provide direct evidence, indirect evidence can also be used to prove the defendant's intention of discrimination, which requires indirect evidence that can be inferred from various facts, data or details to prove the existence of discrimination motivation, such as:a) the plaintiff is a minority group, b) the plaintiff applied for and met the recruitment requirements for a position offered by an employer, c) the plaintiff was rejected despite meeting the eligibility requirements, d) the position remained after the plaintiff was rejected Open, employers continue to seek qualified candidates.

In addition, if race, sex, religion and other prohibited discrimination reasons are used as input data, even if they are not considered as the most important features in the final algorithm model, and thus do not have a discriminatory impact on the legally protected groups, considering that this classification method itself is in violation of the law and there is a potential risk of discrimination against the plaintiff, this behavior should still be regarded as discrimination.[17]

In conclusion, to claim algorithmic discrimination, the plaintiff would need to prove two things: first, the defendant used the race, sex, place of birth, and religion of the specifically protected

ISSN: 1813-4890

classes as input data and conducted an treated them disparately; second, the defendant had an intention to discriminate. There is a subjective intent to discriminate when a predictive model is known to discriminate against legally protected classes, but the decision-making model is still designed or used. Or if the defendant continues to use the algorithm to make predictions after seeing that the algorithm has a disproportionately bad impact on legally protected classes, so that disparate treatment occurs, it shows that the defendant also has an obvious discriminatory intent.[11]

3.2. Disparate Impact

The concept of substantive equality recognizes the differences between individuals. As the differences between people exist objectively, if we ignore these differences and only guarantee procedural equality without considering whether each individual has similar strength, it will aggravate the existing inequality in the real society. Therefore, on the basis of formal equality, the concept of substantive equality has been developed to ensure the realization of the ultimate goal of equality. The prohibition of indirect discrimination is one of the important manifestations of substantive equality in the anti discrimination law.[18]

Indirect discrimination refers to acts that, although ostensibly neutral, cause adverse consequences for individuals belonging to groups of specific gender, race or religious belief. Of course, if there is a legitimate purpose and objective reasons, and the means to achieve that purpose are necessary and appropriate, then those acts are not indirect discrimination. In the United States, indirect discrimination is called adverse impact[18], and adverse impact can also be called disparate impact.

The standard of disparate impact does not emphasize the subjective intention of discrimination, but whether there are discrimination consequences objectively. This standard originated from the case of Griggs v. Duke Power Co. in 1971. In this case, the defendant Duke Power required the applicant to have a high school education and pass the standardized intelligence test when hiring and assigning jobs. Although there is no specific race related factor in the job requirements of Duke Power, due to the long history of racial segregation, African Americans can not get a good education, and the proportion of their high school education is far lower than that of whites, and the scores of standardized intelligence tests are generally low. However, since there is no evidence to show that Duke Power has the intention of racial discrimination when putting forward the requirements for high school education and intelligence test, and this standard is applicable to blacks and whites alike, the district court held that Duke Power's behavior is permitted by law.

However, most of the judges of the Supreme Court of the United States believe that the subjective intention of non discrimination does not exempt the Employer from liability, and the employer should bear the responsibility of proving that the employment conditions are clearly related to the job. In this case, even though the above two employment conditions appear to be neutral, Duke energy cannot prove that there is a reasonable relationship between these conditions and the ability to work, and these conditions have caused most African American job seekers to suffer differential effects.

In short, according to the differential impact standard, the focus of the review is whether the defendant's behavior has adverse effects on the classes protected by the law. The plaintiff does not need to prove that he has been treated differently, nor does it need to prove that the defendant has subjective intention of discrimination. According to the provisions of article 703 (k) of the Civil Rights Act revised in 1991, the court's review steps are: the plaintiff proves that there are facts and evidence of disparate impact based on race, color, religion, gender or nationality; The defendant failed to prove that there was a reasonable connection and commercial necessity, or the plaintiff could propose the existence of an alternative approach to reduce the disparate impact and the defendant refused to adopt such alternative approach.

ISSN: 1813-4890

3.3. Comparison of Two Review Standards

Both the disparate treatment standard and the disparate impact standard play a huge role in the process of anti-discrimination and promotion of equal protection in the United States, especially in the field of employment, but the focus of the two standards is different. As shown in Table 1, their main differences are as follows:

(1)in terms of the subjective intention of the defendant, according to the standard of disparate treatment, the plaintiff needs to provide direct or indirect evidence to prove that the defendant has discriminatory intention, and the court will focus on examining its subjective intention; However, according to the disparate impact standard, the plaintiff does not need to prove that the defendant has a clear discriminatory intention.

(2)As for the standard of proof, according to the standard of disparate treatment, the plaintiff only needs to claim that he has been treated differently because he belongs to a specific class; According to the disparate impact standard, the plaintiff has to prove that it has caused disproportionate impact on a specific legally protected group.

(3)In terms of liability exemption, according to the disparate treatment standard, as long as the defendant's behavior is identified as discrimination, there are no other reasons for exemption; However, according to the disparate impact standard, the defendant can propose that there is a reasonable connection between the setting of conditions and the purpose to be realized and it is in line with commercial necessity, and thus the defendant is exempted from legal liability for the defense reason.

Table 1. Difference between disparate treatment standard and disparate impact standard

Review standard	subjective intention	proof standard	exemption from liability
Disparate treatment	Intention	Individuals are treated disparately	None
Disparate impact	None	Classes are adversely affected	Reasonable connection&Commercial necessity

The process of algorithm-automated decision making lacks transparency, and the intention to discriminate is often hidden in massive data and codes. Identifying and asserting discriminatory intent is difficult for ordinary people. In addition, some of the data sets used in the process of machine learning algorithms are derived from historical data, which contain the prejudice and discrimination prevalent in the real society. Even if the developer or user of the algorithm has no subjective intention of discrimination, the algorithm is very likely to follow the existing discrimination in human society. That is to say, algorithmic discrimination may occur from time to time without discrimination intention. Therefore, in the case of algorithmic discrimination, due to the concealment of the intention of discrimination or the fact that the defendant has no intention of discrimination at all, the disparate treatment standard faces the dilemma of application.

Compared with the disparate treatment standard, the disparate impact standard has greater advantages in the identification of algorithmic discrimination. The disparate impact standard is result-oriented and does not need to prove the existence of discriminatory intentions. It directly considers the results generated by the algorithm-automated decision making to determine whether it constitutes algorithmic discrimination.[19] However, it is worth noting that it emphasizes the disproportionate impact on specific classes protected by law. Therefore, it mainly focuses on structural and systematic discrimination caused by political, economic and social factors. It aims to achieve group justice, pays attention to equality in statistical sense, and neglects individual justice.

ISSN: 1813-4890

When applying the disparate impact standard in the case of algorithmic discrimination, special attention should be paid to the measurement of liability exemption and the allocation of the burden of proof. In practice, the court's review of "commercial necessity" is generally more relaxed. In the case of employment decisions, courts do not require employers to demonstrate that their hiring requirements are necessary to carry out the job, only that defendants need to show that their policies or practices are relevant to business objectives. The burden of proof then shifts to the plaintiff, who needs to provide an alternative approach that achieves the defendant's business goals while reducing the disparate impact in order to determine that the defendant's conduct constitutes discrimination. However, algorithm-automated decision making uses thousands of data and complex algorithm models, and there is no feasible way for the plaintiff to understand the operation mechanism of algorithm-automated decision making, which may lead to the inability of the disparate impact standard to be applied in practice. Therefore, when determining algorithmic discrimination based on the disparate impact standard, how to judge commercial necessity needs to be carefully considered.

4. Enlightenment on the Construction of Anti Algorithm Discrimination System in China

As the application of artificial intelligence in China continues to expand, our lives are gradually surrounded by algorithms, and a series of problems brought by algorithms cannot be ignored. How to balance the innovative development of algorithm technology and the protection of citizens' equal rights is a common problem faced by all countries in the world.Whether it is to solve the problem of algorithm discrimination through the technical improvement of the algorithm model or the legal regulation of the algorithm, the first basic problem to be faced is the identification of algorithm discrimination. China still lacks anti-discrimination laws, and there is no clear legal standard for the identification of discrimination. Therefore, this chapter will draw on the experience of anti-discrimination law in the United States and put forward suggestions on the review standards of algorithmic discrimination in China and the formulation of anti-algorithmic discrimination laws based on national conditions.

4.1. Status Quo of Anti-discrimination Legislation and Judicial Practice in China

In China's current legal system, the legal provisions against discrimination are mainly composed of the following parts:

First, the Constitution. Article 33 sets out the principle of equality. Articles 4 and 48 respectively stipulate the relevant content of equality of all ethnic groups and gender equality, and Articles 34 and 42 respectively stipulate the equality of the right to vote and stand for election and the equality of employment.

Second, international conventions ratified by China. International conventions ratified by the Standing Committee of the National People's Congress, such as the Convention on the Elimination of All Forms of Discrimination against Women, the International Convention on the Elimination of All Forms of Racial Discrimination, and the 1958 Convention on the Elimination of Discrimination in Employment and Occupation, stipulate anti-discrimination, including acts of discrimination, consequences of discrimination, types of discrimination, exceptions to discrimination.

Third, legal provisions. Many Chinese laws, such as the Labor Law, the Employment Promotion Law, and the Law on the Protection of Minors, also emphasize that people enjoy all rights on an equal basis and are not discriminated against based on gender, ethnicity, race, religious belief, family property status,etc. In China, the concepts of equality and prohibition of discrimination have been confirmed in legislation, but not well implemented in practice. China's laws also lack clear provisions on the elements of discrimination and judicial procedures.

ISSN: 1813-4890

Although there are some deficiencies in the law, a large number of anti discrimination cases have still entered the judicial review process in recent years. Through a series of judgments of the court, we can summarize the basic ideas and methods for the court to judge whether the defendant's behavior constitutes discrimination:

First, whether there is disparate treatment is the premise for judging whether the defendant constitutes discrimination. Disparate treatment refers to the effect of classification, so that the originally disordered population is classified according to a certain standard.[20] Second, if there is a disparate treatment, the court will further judge the reason for the disparate treatment. If it is a classified reason prohibited by the law, the defendant's behavior is more likely to be finally judged as discriminatory.[21] Before the implementation of the employment promotion law, the discrimination prohibited by the Labor Law was limited to four categories: nationality, race, gender and religious belief, which could not adapt to the regional discrimination, height discrimination, hepatitis B or HIV discrimination in practice. In the trial practice at that time, some judges broke through the restrictions of the Labor Law and identified the reasons that were not prohibited by the Labor Law as the causes of discrimination. Finally, the act of disparate treatment must lead to the plaintiff's material or spiritual relatively adverse consequences, and there is a causal relationship between the classification reason and the adverse consequences.

When judging whether there is a causal relationship, the class that the plaintiff claims has been treated differently can be compared with other classes, and the similarities between the two can be removed, and the remaining differences are considered to be the causes of discrimination. However, in the judicial practice of anti-discrimination cases, most of the classes that are treated differently have no comparable classes. In this regard, the court usually turns to investigate whether the defendant has the subjective intention to discriminate, that is, to determine whether the defendant has a classification reason prohibited by the law when he makes a distinction. On the other hand, judicial practice at this stage shows that courts usually expect to find a clear chain of causality between disparate treatment and adverse consequences. On the whole, in China, the review standards for anti-discrimination cases are relatively rough, the courts are more conservative when making decisions, and the review standards for causation are relatively strict, which still cannot fully play the role of eliminating discrimination and achieving equal protection.[20]

Compared with other types of discrimination, algorithmic discrimination is extremely hidden and complex, and the application of machine learning algorithms adds more obstacles to discrimination identifying.On the one hand, judging the reason for classification is an important part of the review of discrimination cases, but in algorithmic discrimination cases, the reason for classification may be the combination of legal and illegal elements, a clear causal chain is difficult to find, and the process of labeling class attributes is also hidden in the algorithm black box, so it is difficult to determine whether its reason for classification is prohibited by law.[19] On the other hand, the results of algorithmic decision-making are not public, and usually only individuals who accept the decision can see their own results, and it is impossible to analyze the real reasons for discrimination by comparing people with each other. Therefore, it is difficult for China's current discrimination review standards to play a good role in algorithmic discrimination cases.

4.2. Constructing and improving China's anti-algorithmic discrimination legal system

4.2.1. Adjust the "disparate treatment" standard

In algorithmic discrimination cases, due to the concealment of algorithmic discrimination, the scope of application of disparate treatment standards has been greatly compressed. However, this does not mean that the standard is useless in regulating algorithmic discrimination. If the

ISSN: 1813-4890

disparate treatment standard is properly adjusted, it can still play the function of reviewing and identifying algorithmic discrimination.

On the one hand, the criteria for judging discriminatory intent should be appropriately relaxed, and the obligation of prudent review should be imposed on algorithm developers and users. Regarding the understanding of discriminatory intent, Professor Michael Gold pointed out that the legislative purpose of Section VII of the Civil Rights Act is only to prohibit conscious discrimination, such as intentional and permissive mentality.[22] A judgment made by the U.S. court based on the Fourteenth Amendment to the constitution shows that the intention of discrimination should be very clear. The plaintiff should prove that the algorithm developer made the rules to have a negative impact on a certain group, and can't judge the rule maker's intention of discrimination just because he didn't take this negative impact into account.[22] However, in addition to the programs developed or used for explicit purposes of discrimination, there may also be discrimination due to subconscious discrimination, the use of contaminated data sets or redundant coding in the development of other programs. Therefore, it can be considered to appropriately relax the criteria for judging discriminatory intent, and include the negligence of algorithm developers and users into the scope of discriminatory intent. In other words, algorithm developers and users should be subject to prudent review obligations, that is, they are not required to have a clear and direct intention of discrimination. As long as they fail to fulfill their prudent review obligations, regardless of the possibility of discrimination consequences, applying proxy variables highly related to gender, nationality, race, religious belief, registered residence, social origin and other characteristics to algorithm decision-making also constitutes algorithm discrimination.[23]

On the other hand, a data cleaning system should be established with the help of differential treatment standards, and the training data in the algorithm model should be cleaned in advance. Treating people differently according to the classification reasons prohibited by law is the direct cause of discrimination. If gender, race, nationality, religious belief, registered residence, social origin and other legally prohibited items are used as input data, even if they are not regarded as the most important features in the final algorithm model, and thus do not have a discriminatory impact on the legally protected classes, this classification still has the potential risk of discrimination against individuals. Therefore, in order to avoid the risk of discrimination, items with significant discrimination prohibited by laws such as gender, nationality, race, religious belief, registered residence and social origin should be excluded from the scope of big data collection, and they should be prohibited from being used as training data for machine learning.[11] Although forbidding the use of the above data in the algorithm model may reduce the accuracy and efficiency of algorithm decision-making, we cannot sacrifice the basic right of equality of the people simply to pursue the accuracy and efficiency of algorithm decision-making. We should pursue the accuracy and efficiency of algorithm decision-making on the basis of ensuring formal equality.[23]

4.2.2. Introduce and redefine disparate impact standard

In the age of artificial intelligence, decision-making in our daily lives increasingly relies on seemingly neutral algorithms, which allows discrimination to occur in a more insidious way. In the algorithm decision-making process, discrimination occurs in a more subtle way, which makes the identification of discrimination more difficult, and often needs to be judged and identified from the output results of the algorithm model. In this regard, the differential impact standard has a large space for application. China should introduce the differential impact standard to judge whether the defendant constitutes algorithmic discrimination by examining whether the consequences of algorithmic decision-making have adverse effects on special protected groups.

ISSN: 1813-4890

First of all, it is necessary to clarify which groups belong to the key protected objects. The standard of disparate impact originated in the United States can be applied to those cases with differential impact based on the reasons prohibited by law, but it is undoubtedly aimed at racial discrimination, the largest inequality in American society. To a certain extent, the standard can prevent the further expansion of this inequality. In the context of China, because of the serious imbalance of resource allocation, the inequality caused by the gap between the rich and the poor and social stratum differentiation may be the most serious inequality. Vulnerable classes should be taken as the main protection objects, and algorithm decisions that adversely affect vulnerable classes should be identified.[22]

As mentioned above, in the judicial practice of the United States, judges generally review "commercial necessity" more leniently, that is, the defendant only needs to show that its algorithm rules are relevant to commercial objectives to meet the exemption conditions. This may lead to the fact that the disparate impact standard cannot be applied. In view of this problem, it is necessary to reconsider the determination standard of "commercial necessity" in algorithmic discrimination cases. If the selected target variable is not sufficiently related to the business objective, the defendant cannot claim the exemption of business necessity and needs to bear the responsibility arising from the improper selection of the target variable. Once the target variable that has relevance to the business goal is determined, the next question is whether the algorithm model can predict the target variable. If a reasonable data mining tool is selected and enough information is obtained, a more accurate prediction effect may be achieved.[11]

But to be clear, the algorithm works by discovering hidden correlations in massive amounts of data and making predictions about the future based on that. Correlation is a minimum requirement for an algorithmic model, and its use in algorithmic decision making is not necessarily justified. Developers and users of algorithms must also justify the relevance of the data variables they use to specific decisions, and that the use of these data as variables is reasonable and not discriminatory.

4.2.3. Guarantee equitable access

With the development of big data, artificial intelligence and other technologies, the digital economy is profoundly changing our way of work and life. However, due to the lack of equipment or insufficient ability to access and use information, digital dividends can not benefit every member of society evenly, but marginalize some members of society, especially for vulnerable classes such as the elderly, people living in remote and poor areas and people with low education, and a widening "digital gap" lies in front of them. Algorithmic decisions need to be based on various data of users, and the lives of these vulnerable groups will be less digitized. They will be systematically left out, and will be even more disadvantaged in the process of resource allocation. Algorithm decision-making needs to be based on various data of users, and the lives of these vulnerable groups will be less digitized, they will be systematically missed, and they will be in a more disadvantaged position in the process of resource allocation. The universal application of digital technology in various fields such as government services, health care, education, and consumption will lead to the exclusion of vulnerable groups and cause substantial inequality.

In order to bridge the digital divide, it is necessary to strengthen the construction of Internet infrastructure, increase the popularity of smart devices, and ensure that members of society have equal rights to access the Internet. As of June 2020, the number of Chinese Internet users was 940 million, and the Internet penetration rate reached 67.0%. Although China's Internet industry has taken the lead in the world, it still cannot meet the needs of many people to connect to the Internet. During the epidemic, due to the lack of Internet access conditions and hardware equipment, online classes became an unbearable burden for children in remote and poor areas.

ISSN: 1813-4890

Improving Internet coverage in remote areas, narrowing the gap in the supply of Internet services, and providing preferential protection to vulnerable groups are the basis for narrowing the digital divide. In addition, it is also necessary to cultivate the IT skills of vulnerable groups and improve their ability to integrate into the Internet age. Therefore, it is necessary to ensure the fair access of digital technology to members of society in terms of improving hardware and enhancing digital skills.

While continuously promoting digital services, offline services should also be preserved and continuously optimized to ensure the universality of public services. Digital technology has brought many inconveniences to the lives of the elderly. While allowing more people to enjoy high-quality and convenient services, we should not draw a gap in front of the elderly and other groups with difficulties in using digital technology, so that they are at a disadvantage in accessing public services. At the same time, personal data should not be used as an important basis for social evaluation and public service provision. Therefore, on the one hand, personal data unrelated to specific public services should be removed, and on the other hand, personal data should be regarded as a factual description of individual behavior, only as a secondary reference for individual evaluation.[24]

5. Conclusion

Whether the way to solve the problem of algorithm discrimination is the technical improvement of algorithm model or the legal regulation of algorithm, the first basic problem to be faced is the identification of algorithm discrimination. Although the concept of prohibition of discrimination is affirmed in Chinese laws, these laws are difficult to implement in practice. The legal provisions on the constituent elements of discrimination are not clear, and the review ideas formed by judges in judicial trials are also difficult to solve the problem of how to identify algorithmic discrimination. Therefore, this paper sorts out the useful experience in the identification of discriminatory behavior in the United States, compares and analyzes the two legal standards of disparate treatment and disparate impact, and reflects on the application dilemma of the two review standards.

On this basis, the author believes that the standard of disparate treatment and the standard of disparate impact should be used for reference. On the one hand, a data cleaning system can be established with reference to the disparate treatment standard, and under the condition of appropriately relaxing the criteria for judging discriminatory intent, we can impose a prudent review obligation on algorithm developers and users according to the disparate treatment standards. On the other hand, since the defendants in a large number of algorithmic discrimination cases have no discriminatory intent or only concealed discriminatory intent, the disparate impact standard should be introduced, and algorithmic decisions that adversely affect vulnerable classes should be identified in combination with China's national conditions, and the exemption of "commercial necessity" should be carefully applied. In addition, we should ensure that members of society can have fair access to digital technology, retain and constantly optimize offline services, so as to ensure the universality of public services.

Finally, algorithm governance is a complex systematic project. To achieve the goal of eliminating algorithmic discrimination, in addition to the construction of a legal system against algorithmic discrimination, it is also necessary to further construct algorithm-related systems in combination with algorithm disclosure and personal data empowerment.[25] Due to space limitations, this article only focuses on the institutional construction of anti-algorithmic discrimination, and will not discuss other aspects of algorithmic regulation for the time being.

ISSN: 1813-4890

References

[1] Executive Office of the President, “Big Data: Seizing Opportunities, Preserving Values”, https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf, pp.51-53.

[2] Executive Office of the President, “Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Right”, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

[3] D.K. Citron & F.Pasquale, “The Scored Society: Due Process for Automated Predictions”, Washington Law Review, 2014(89), pp.1-8.

[4] X. Ding, “Algorithms and Discrimination: Algorithmic Ethics and Legal Interpretation from the U.S. Educational Affirmation Case”, Peking University Law Journal, 2017, 29(06), pp.1609-1623.

[5] J. Cui, “The crisis and response of equal rights protection under the challenge of algorithmic discrimination”, Science of Law, 2019, 37(03), pp.29-42.

[6] Y. Liu, “Research on Algorithmic Bias and Its Regulatory Approach”, Law Science Magzine, 2019, 40(06), pp.55-66.

[7] D. Hong, “On the Guidance of Constitutional Value on Algorithmic Discrimination: Based on Human Dignity”, Political Science and Law, 2020,(08), pp.27-37.

[8] W. Shen, “The Myth of the Algorithmic Transparency Principle: A Critique of the Algorithm Regulation Theory”, Global Law Review, 2019,41(06), pp.20-39.

[9] L. Zhang, “Function and Realization of the Algorithm Interpretation Rights in Business AutomatedDecisions”, Science of Law, 2018,36(03), pp.65-74.

[10] Mikella & Hurley & Julius, et al, “CREDIT SCORING IN THE ERA OF BIG DATA”, Yale Journal of Law & Technology, 2016, 148(18), pp.150-160.

[11] S. Barocas & A.D. Selbst, “Big Data's Disparate Impact”, California Law Review, 2016,104(3), pp.671-732.

[12] S.G. Mayson, “Bias In, Bias Out”, 128 Yale Law Journal 2218 (2019), University of Georgia School of Law Legal Studies Research Paper No. 2018-35.

[13] L. Sweeney, “Discrimination in Online Ad Delivery”, COMMUNICATIONS OF THE ACM, 2013, 56 (5), pp.44-54.

[14] K. Crawford, "Think Again: Big Data", Foreign Policy, 2013, pp.1–27, Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Think+Again:+Big+Data#8

[15] R.H. Sloan & R. Warner, “Algorithms and Human Freedom”, Social Science Electronic Publishing, 2019(35), pp. 10-17.

[16] W. Li, Principles of Anti-Discrimination Law, Law Press, China, 2012, pp.64.

[17] R.A. Primus, “Equal Protection and Disparate Impact: Round Three”, Harvard Law Review, 2003,117(2), pp.493-587.

[18] W. Li, Principles of Anti-Discrimination Law, Law Press, China, 2012, pp. 64.

[19] J. Cui, "The Crisis and Response of Equal Rights Protection under the Challenge of Algorithmic Discrimination", Science of Law(Journal of Northwest University of Political Science and Law), 2019,37(3), pp. 30-39.

[20] C. Li, "Judicial Protection of Equal Rights--Based on the Analysis and Summary of 116 Judgments in Anti-discrimination Litigation",ECUPL Journal, 2013(04), pp.57-68.

[21] X. Liu, "Anti-Discrimination Law Lecture Notes--Text and Cases", Law Press, China, 2016, pp. 10-11.

[22] X. Ding, “exploring the legal standard of anti discrimination and equal protection, starting from the standard of differential influence”, Peking University Law Journal, 2014,26(04), pp.1093-1094.

[23] E.Zhang, "Against Algorithmic Discrimination: Theoretical Reflection and Institutional Construction", Journal of Huazhong University of Science and Technology(Social Science Edition), 2020,34(05):60-71.

ISSN: 1813-4890

[24] Y. Gao, "rights protection of digital vulnerable groups in smart society", Jianghai journal, 2019(5). pp.168.
[25] X. Ding, "On the Legal Regulation of Algorithms", Social Sciences in China, 2020(12):138-159.