CSCanada
Cross-Cultural Communication
Vol. 16, No. 1, 2020, pp. 65-75
DOI:10.3968/11612

ISSN 1712-8358[Print]
ISSN 1923-6700[Online]
www.cscanada.net
www.cscanada.org

Legal Regulation of Algorithms From the Perspective of Interpretability

CAI Lin[a],*; JI Xueke[a]

[a] Department of Law, Northwestern Polytechnical University, Xi'an, Shaanxi, China.
*Corresponding author.

Supported by the National Social Science Fund Project "Research on the Legal Regulation of Algorithms under the Chinese Discourse System" (Project Number: 19XFX001).

Received 16 December 2019; accepted 27 February 2020
Published online 26 March 2020

Abstract

The human life in the age of artificial intelligence has undergone tremendous changes. Algorithm technology is widely developed and applied as one of the core technologies of artificial intelligence. However, a series of problems such as algorithm discrimination, algorithm killing, and "information cocoon" caused by unexplainable algorithms represented by artificial neural networks needed to be solved urgently,which forms a risk society. The algorithmic order is gradually "offside" into a new social order, which challenges the existing legal order. Because the existing legal order upholds the neutral value of technology tools and does not pay attention to the legal regulations of technology itself, it cannot make ethical prejudgment of unexplainable algorithms to prevent and control social risks. With the deepening of the "intelligence" of algorithm technology, social risk is expanding. The field of algorithm technology creates interpretable algorithms to respond to social risks. However, due to the lack of legal value and institutional design support, the technological advantages of interpretable algorithms to prevent and control algorithm black boxes, "Offside order" and coping with a risky society cannot be confirmed and guided by law. Therefore, it is an effective way to solve the risks in the age of artificial intelligence by taking the technical critical theory of risk society as the value basis and taking the interpretability of algorithms as a necessary condition for algorithm regulation.

Key words: Algorithm; Interpretability; Risk society

Cai, L., & Ji, X. K. (2020). Legal Regulation of Algorithms From the Perspective of Interpretability. *Cross-Cultural Communication, 16*(1), 65-75. Available from: http//www.cscanada.net/index.php/ccc/article/view/11612
DOI: http://dx.doi.org/10.3968/11612

1. LACK OF ALGORITHM INTERPRETABILITY REGULATIONS

After entering the 21st century, algorithm technology has been widely concerned for its accuracy and efficiency, and has entered all aspects of society (Guo, 2017). For example, apply algorithmic technology to comprehensively analyze the student's campus card consumption records, family conditions, and other conditions to achieve implicit funding for poor students (Zhang, 2018, pp.65-74); Many companies in the insurance, finance and other industries apply artificial intelligence systems to replace manpower (Si and Cao, 2017); use algorithms to count votes (Krull, et al, 2019). Scholars predict that "artificial intelligence will subvert human's existing lifestyle and usher in a brand new high-tech era." (Wu, 2017, pp.76-78).

However, the sprout and development of any new technology will inevitably be accompanied by profit and loss. In recent years, algorithmic accidents have occurred frequently and hidden dangers have frequently appeared. Algorithmic technology risks in the artificial intelligence society have also followed. In 2016 China International High-tech Achievements Exhibition, the robot "Xiaopang" suddenly lost control and smashed the glass platform, causing damage to people and property (Gao and Zhang, 2018); In the same year, a traffic accident occurred in California due to an error in the autonomous driving system (He, 2019); The COMPASS algorithm was controversial because of racial discrimination in criminal risk assessment (Zhang, 2019, pp.16-26); Google uses algorithms to perform black box operations to eliminate

competition (Liu, 2019, pp.55-66); Internet consumer platforms such as taxis, movies, and hotels use algorithm technology to analyze users' personal data and implement differentiated pricing and precision marketing (Zou & Liu, 2018). All of the above are contrary to the original intention of human development of algorithm technology, that is, "the original human work was left to the machine." (Chu, 2014) With the development of technology, human beings have to face a series of technical risks such as algorithm infringement, algorithm discrimination, algorithm killing, algorithm black box, and "information cocoon" while enjoying the intelligent dividend (Jiang and Li, 2019). Algorithm problems such as these are rooted in the unexplainable algorithms.

Unexplainable algorithms can be divided into two types: First, subjectively unexplainable, that is, black box operations caused by no obligation to interpret. It usually manifests as the "rule of the elite", using algorithmic technology's dominant position through trade secrets and other methods, leading to the opacity of algorithmic operation rules. COMPASS technology is of this type, which has caused controversy in US lawsuits. Most of these unexplainable algorithms belong to weak artificial intelligence algorithm technology. Second, it is objectively unexplainable, that is, the final decision rules of the algorithm itself are changing with deep learning, and technically lacks predictability. This kind of unexplainable algorithm is derived from deep learning technology, [1]that typified by artificial neural network, which belongs to strong artificial intelligence algorithm technology. For example, the deep learning algorithm Deep Mind developed by Google is typical of strong artificial intelligence algorithm technology. The subjective unexplainable algorithm can be solved by adjusting the existing trade secret system and the information disclosure obligation corresponding to the public right to know. Therefore, the social risks that laws need to prevent and control mainly come from objective unexplainable algorithms.

Objective unexplainable algorithms cannot achieve the transparency and predictability of technology because of the scope beyond interpretability. Machine learning of objective unexplainable algorithms is the "uncontrollability of machine learning results" (Zheng, 2018, pp.66-85), and "the results are unpredictable for unsupervised and reinforcement learning" (Zheng, 2018, pp.66-85). Its operation process and decision-making principles are in a black box. As a result, the post-event responsibility of technical operations and the prevention and control of legal risks do not have a realistic technical basis. Therefore, the law cannot prejudge the legality and ethics of the technology through the operation principle of such algorithmic technology.

In the field of algorithm technology, technical solutions have been implemented for the unexplainable problems of deep learning algorithms such as artificial neural networks. The emergence of new interpretable algorithm technologies such as The Interpretable Classification Rule Mining algorithm (Zheng, 2018, pp.66-85),,2New Ordered Predictor Selection algorithm,3and Interactive Three-dimensional Model Technology4,has enabled deep learning algorithms to have both deep learning and interpretability, and has brought the possibility of strong artificial intelligence algorithms to achieve interpretability .Such algorithms are built on the basis of objective unexplainable algorithm technology to overcome the dilemma of interpretation. The obvious difference can be called "Interpretable algorithm".

However, due to the lack of institutional orientation, such technology research and development has not been able to become the direction of the development of strong artificial intelligence algorithm technology. It can be seen that the social risks caused by the unexplained solution algorithm cannot be completely dependent on the technology's response to the risks. In particular, the field of artificial intelligence algorithms has not yet formed a unified industry ethics and industry standards. Therefore, uniform mandatory rules are needed to guide the development direction of algorithm technology. Through the value of law, reverse generation of industry ethics, and guide the benign development of strong artificial intelligence algorithm technology to reduce the potential risks caused by its unexplainable and "black box" autonomous decision-making.

[1] Shixia Liu,Xiting Wang, Mengchen Liu,Jun Zhu,Towards better analysis of machine learning models: A visual analytics perspective：Although machine learning models are widely used in many applications, they often fail to explain their decisions and actions to users. Without a clear understanding, it may be hard for users to incorporate their knowledge into the learning process and achieve a better learning performance.

[2] The Interpretable Classification Rule Mining (ICRM) algorithm, is designed to maximize the comprehensibility of the classifier by minimizing the number of rules and the number of conditions. The experiments show that the proposal obtains good results, improving significantly the interpretability measures over the rest of the algorithms, while achieving competitive accuracy. This is a significant advantage over other algorithms as it allows to obtain an accurate and very comprehensible classifier quickly.

[3] OPS is a recognized method to select variables in multivariate regression . It shows high ability to improve the prediction of models after the selection of a few and important variables. The new OPS approaches outperformed the first OPS version and the other variable selection methods. Results showed that in addition to greater predictive capacity, the accuracy in the selection of expected variables is highly superior with the new OPS approaches. Overall, the new OPS provided the best set of selected variables to build more predictive and interpretative regression models.

[4] Shixia Liu,Xiting Wang, Mengchen Liu,Jun Zhu,Towards better analysis of machine learning models: A visual analytics perspective：Machine learning models are seamlessly integrated with state-of-the-art interactive visualization techniques capable of translating models into understandable and useful explanations for an expert. The strategy is to pursue a variety of visual analytics techniques in order to help experts understand, diagnose, and refine a machine learning model.

With the gradual popularization of deep learning algorithms and showing the trend of mainstream technology in the future algorithm society, this potential risk will push the algorithm society to a deeper risk society. "While bringing endless expectations and joy, algorithms will also gradually challenge our existing laws, ethics and order " (Tencent Research Institute, etc, 2017, p.322) . To cope with this new and risky society, science and technology legislation should set standards for value judgment and ethical definition for algorithmic society in advance. At present, domestic and foreign artificial intelligence technology legislation has its own emphasis. The "Artificial Intelligence Standards Development Guide" issued by the National Institute of Standards and Technology lists the principles of technology moving towards kindness, reducing technical prejudice, and flexible formulation of artificial intelligence standards to balance artificial intelligence tool performance and technical credibility; the "General Data Protection Regulations" promulgated by the European Union provides for the protection of personal privacy and data interests by algorithmic automated decision-making…It can be seen that the issue of algorithm interpretability has not attracted much attention from legislators worldwide. "The autonomy and opacity of intelligent algorithm decision making makes it impossible for humans to see the specific process of algorithm decision making, so that the algorithm becomes an emerging force that allocates social resources and forms a de facto technical power." (Zhang, 2019, pp. 63-75) "Algorithmic power" gradually expands under the impetus of technology, runs wild outside the legal system, gradually establishes a set of algorithmic order instead of legal order, and eventually leads to human beings in the "elite rule" controlled by the owner of algorithmic technology. In this trend, scholars at home and abroad have gradually realized that the idea of technology neutrality is no longer applicable, and have sought new technology theories. Some scholars believe that "algorithms are gradually separated from purely instrumental roles, and the creation of rights, obligations and responsibilities with its own characteristics and connotations is the future direction of legal development" (Zhang, 2018, pp.65-74); Others believe that "smart 'new generation robots' have relative autonomy and a higher level of artificial intelligence, and in order to deal with the risks of the artificial intelligence society, they should formulate a robot ethics charter in time and conduct special robot legislation in due course." (Wu, 2017, pp.128-136)

At the same time, concerns about the "rule of the elite" followed, and algorithmic autonomous operation rules and decision rules entered the field of legal research. Algorithm interpretation has been moved from the field of algorithm technology to the field of algorithm regulation, and it has become one of the research paths to solve the social risk of algorithms. Facing the major algorithm problem of algorithm black box, Chinese scholars have proposed a legal regulation path for algorithm interpretation. For example, when dealing with the risks of business automation decision-making, set the right to interpret algorithms to "make up for the deficiencies of traditional power systems in responding to the technological development of the artificial intelligence era" (Zhang, 2018, pp.65-74); Taking algorithmic interpretation as the entry point to "imitate Europe and the United States to start from two aspects of data regulation and algorithm detection and supervision, and create an algorithm security committee based on China's current situation, and establish a liability mechanism for algorithm infringement through legislation" (Sun, 2019, pp.108-117); The algorithm is explained for the purpose of protecting data security and the rights of relief subjects. It is proposed that "when a regulatory agency lacks the necessary resources or information, a meta-regulation model should be adopted, that is, through positive and negative incentives, to promote the data controller's own self-regulation of the problem Respond " (Cheng, 2019, pp.48-55). Although the current research involves the interpretability of the algorithm from different angles, it is only used as a path to solve a certain type of problem of the algorithm, and it does not dig deeper into its fundamental meaning to construct a systematic legal regulation scheme. All algorithm problems are rooted in unexplainability. Only through interpretation can the algorithm achieve transparent operation. At this time, legal intervention and technical regulation have a basis of rationality and legitimacy.

2. TECHNICAL BASIS OF INTERPRETABLE ALGORITHM LEGISLATION-ALGORITHM EXPLANATION

The risks caused by the development of science and technology will eventually return to the technical field and seek solutions (Luo, 2003, pp.1-7). As Hobbs said: "... I always thought that everything depends on science, because in the end only science can determine the relative value of our different social goals within the limits of the level it has developed ..." (Hobbes, 1985, p.72). In recent years, a variety of new algorithm technologies developed in the field of computer science and technology can solve the problem of unexplainability of deep learning algorithms in a recorded or visualized manner, providing a technical basis for algorithm legal regulation. This type of new algorithm technology is similar to traditional deep learning algorithms in that they have strong artificial intelligence characteristics, that is, the ability to simulate the human brain to perform autonomous learning through high-speed operations. However, the new algorithm has many characteristics that are different from traditional artificial neural network algorithms, namely recordability,

interpretability, predictability, and transparency. The core difference between the two types of algorithm technology is that the new algorithm has interpretable features. In order to distinguish it from traditional deep learning algorithms, this new type of algorithm technology is collectively referred to as interpretable algorithms in this paper.

The development of interpretable algorithm technology is still in its infancy. There are four types of existing technology implementation paths at home and abroad, including new visualization technologies, latent feature interpretability, new ordered predictor selectionalgorithms, and interpretability classification rule mining algorithm.Traditional visualization technology is only based on point-based visualization. It cannot reproduce the three-dimensional structure of the artificial neural network algorithm. It cannot make its deep learning process transparent, so it cannot implement the function of interpreting algorithms. The new visualization technology uses a novel algorithm transparency technology, which breaks through the limitations of traditional visualization model technology and can explain traditional artificial neural network algorithms. For example, the interactive three-dimensional model technology reproduces the network topology structure by reordering the matrix, realizes the interactive visualization of the technology, and generates an interpretable model, so that the neural network algorithm can be interpreted (Liu, et al, 2017). The research results released by Google in 2018 on the subject of "interpretable infrastructure" (Olah, 2018). The core of his research is the nuclear magnetic resonance resonance imaging algorithm technology of artificial neural network. This visual solution technology makes the traditional artificial neural network algorithm transparent by simplifying the algorithm information and so on, which is convenient for the public to understand and understand (Zhang, 2019, pp.16-26). The latent feature interpretability method is a new interpretable algorithm developed by FICO in 2019.This algorithm technology converts the driving characteristics of each hidden node specification in the traditional artificial neural network algorithm to convert it into an interpretable neural network model. In this process, if the number of inputs of a single invisible node reaches the upper threshold of the node's interpretability, the interpretation of the potential node needs to take another way, that is, iteratively activate the hidden node by applying analytical techniques and construct a new artificial neural network algorithm. The new ordered predictor selection algorithm adopts the method of establishing a multivariate regression model and selects a set of best selection variables from the centralized data. This method is superior to the traditional feature selection method and can be more predictable and interpretable Regression model (Roque, et al, 2019); The interpretable classification rule mining algorithm is an EP method suitable for classification problems. It follows an iterative rule learning model and uses a single rule representation. Therefore, there are no paired rules with the same antecedent and different results. This classifier algorithm achieves the maximum degree of interpretability of the algorithm while pursuing accurate results by obtaining the minimum conditions and number of rules (Cano, 2013). The four new types of algorithm technologies, whether starting from the construction of interpretable basic models or from the transparency of transformation algorithms, can improve the deep learning capabilities of algorithm technologies while making them interpretable. Based on the essential characteristics of interpretable algorithms-interpretability, the regulation of laws has a technical basis.

Different from interpretable algorithms, traditional artificial neural network algorithms cannot be interpreted, predicted or transparent due to the existence of algorithm black boxes, so they do not have the technical conditions for legal regulation.Generally, the interpretability and accuracy of algorithm technology show an inverse correlation. The higher the accuracy of algorithm technology, the lower the interpretability of the algorithm (Ma, 2006). The deep learning algorithms represented by traditional artificial neural network algorithms have the highest accuracy, but they are often not interpretable. This type of algorithm processes the data separately through the training set and the test set, and finds rules for deep learning during the running of a large amount of training data input and output. This operating principle has led to the "black box" phenomenon of the traditional artificial neural network algorithm in the decision-making process. Even the inventor of the algorithm cannot predict and master its operating rules and calculation results (Zong and Chow, 2020). Therefore, traditional artificial neural network algorithms are naturally unexplainable. If traditional deep learning algorithms are included in the legal category, the research and development and application of such algorithm technologies will be in the blind zone of legal monitoring due to their unexplainability. And there will be no basis for legal supervision and accountability after the fact, the legal regulation at this time will lose its proper meaning. In view of the fact that algorithm industry practices and technical ethics have not yet been formed, such unexplainable algorithms do not have any targeted technical regulation at this stage. Therefore, in order to achieve the purpose of risk prevention and control, traditional deep learning algorithm technology without interpretability should be included in the legal forbidden area.

Technical regulation and risk prevention and control are surely important, but the law cannot blindly pursue the interpretability of algorithm technology and abandon the pursuit of higher accuracy. It cannot inhibit the progress of algorithms and limit the technical potential in order to achieve risk prevention and control. The original

intention of technology legislation will also cause a country's algorithm technology to lag behind in the global industrial chain of artificial intelligence. Take France 's recently issued ban on the use of artificial intelligence tools to analyze judicial opinions as an example. This ban makes legal artificial intelligence tools lose their original meaning, technology research and development no longer has application and promotion value, and it cannot achieve the purpose of technology to benefit human society. Therefore, the law's attitude towards emerging technologies should not be blindly suppressed, nor should it be allowed to allow technology to be researched and abused arbitrarily, and the world of algorithms should not and could not become a place outside the law. The law should seek a balance between technological development and risk prevention and control. The optimal path is to set access standards for algorithms in the era of artificial intelligence to prevent risks from occurring or to provide effective relief methods when they come. Interpretable algorithm technology has the technical advantages of both deep learning and interpretability, which makes the access law regulation method have realistic possibility. Therefore, the institutional arrangement for legal risk prevention and control should take interpretability as the threshold for the development and application of algorithm technology.

The interpretable algorithm not only meets the requirements of scientific and technological development, but also conforms to the concept of good science and technology. The promotion and application of this technology has necessity and practical possibility. The recordability and interpretability characteristics of interpretable algorithm technology enable the law to regulate and account for the technology through pre-regulation and after-the-fact relief, so the application of this emerging technology will not bring new technical risks and technological risks. Algorithm interpretability legislation can effectively prevent the emergence of a series of algorithm problems such as algorithm black box, discrimination, differentiated pricing, and standardize the technology research and development and application promotion in the field of algorithms. This technology lays the foundation for the legalization of algorithms, which can not only realize algorithms to promote artificial intelligence society, but also effectively prevent scientific and technological risks, and achieve the purpose of technology for the benefit of humankind.

Although interpretable algorithms have many technical advantages, they have not yet become the trend of algorithm market research invention and promotion. The reason is that the research and development of this technology is still at the conscious stage of the enterprise, and there is no legal mandatory requirement, and most companies do not apply it for the consideration of research and development costs and commercial interests. However, from the perspective of consumers, their ability to master information resources and science and technology has a natural disadvantage compared to enterprises. The existence of unexplainable algorithm technology has made the gap between the two more disparate. Therefore, in order to realize social public welfare and maximize the protection of consumers' legitimate rights and interests, the law must use interpretability as the entry threshold for algorithm technology, popularize and interpret interpretable algorithm technology, prevent the emergence of elite governance in society, and eventually plunged into a situation where technology controls humans. On the other hand, law has a guiding role in technological development. After the legislative setting of algorithm interpretability standards, the field of algorithms will take interpretability as the direction of technology research and development, and will promote the iteration of interpretable algorithm technology in reverse, which greatly saves resources and after-the-fact relief costs. This will gradually form a complete and orderly algorithm technology industry chain, and promote the development of algorithm technology in a standardized and orderly direction.

3. INTERPRETABILITY AND REALIZATION OF THE VALUE DEMAND OF ALGORITHM REGULATION LEGISLATION

The legislative regulation of the algorithm is an inevitable choice to deal with the risk society in the era of artificial intelligence. The interpretability contains the ethical connotation of preventing and controlling risks and protecting public welfare. It is the basis for solving social risks such as the algorithm black box. It has a natural fit with the value requirements of algorithm regulation. Technology development and risk society are inseparable. To a certain extent, it can be considered that technology has created risks in contemporary society (Perrow, 1994). Risks such as threats to human subjects and violations of public welfare caused by algorithm black boxes are contrary to the basic ethics of human society. As the emerging technology of algorithms has not yet formed technical ethics and industry rules, it is necessary to adopt technical critical values to legislate to prevent and control risks, delineate ethical boundaries, and use the interpretability of algorithms to manage many algorithm problems rooted in the algorithm black box, and guide the healthy development of algorithm technology.

3.1 Interpretability and Governance of Risk Society

"Antibiotics problem", "Ditch oil problem", "Terrorism", "Sanlu milk powder storm" (Liu, 2011) …risks occur frequently, and various events indicate that human society has ushered in an era of risk. As the founder of the theory of risk society, Baker proposed that society itself

is in a state of confrontation, which is a contradiction between the premise and result of modern society and the risks it contains (Wang, 2009). The key content of this concept is "risk". To some extent, risk belongs to post-rationality, which is the product of overcoming the instrumental rational order (Yang, 2003). On this basis, Anthony Giddens made an in-depth explanation of the risk society. In his view, many social conflicts cannot be simply reduced to order contradictions, but should be defined as risk issues beyond human control (Wang, 2009). All in all, modernity is a large experiment intertwined with risk and certainty that is led by humans but to some extent beyond human control (Baker, Giddens, and Rush, 2001). The artificial intelligence society is a risk society. Algorithmic technology, while promoting social innovation, has also bred the risks of a smart society. Algorithm discrimination, algorithm black box, information cocoon room, algorithm threats human subject status and so on, many risks are accompanied by technology. The long-term lack of technical regulation system will allow the infinite expansion of algorithm power, resulting in frequent risks and crises in artificial intelligence society. The original legal system cannot cope with the new situation of the development of algorithm technology, and the new technical regulation legal system has not yet been established, and risk prevention and control in artificial intelligence society will have no legal basis. Therefore, the risk social governance in the context of artificial intelligence society is a major topic of the times that human beings must face. The exploration and establishment of a technical regulatory legal system is necessary and urgent.

The premise of risk social governance is to choose the value concept of technology regulation that is suitable for it. Similar to the development and changes of industrial society, the value concept of human regulatory technology has also undergone a long-term evolution process. From the Industrial Revolution to the middle of the 20th century, the dominant concept of technological regulation has always been the theory of technology as a tool (Liu, 2011, pp.68-78), the technology under this theory is neutral, that is, the technical method has nothing to do with the value goal it helps to achieve. The so-called value is only related to people (Yang, 2003), the value of technology originates from the application, so the regulation of technology should not be less than the technology itself, but should be targeted at the application behavior and application subject of the technology. Once the behavior of the application technology violates the law, the application is regulated, and it does not go back to the technology .In the process of development design and promotion, technology inventors do not need to take responsibility. However, this theory is obviously inherently flawed. In the case of gene-edited babies, the research and development of this technology has run counter to human ethics (Wang, 2019, pp. 134-144). If only the application of technology is regulated but the research and development orientation is not regulated, it will greatly impact the existing legal order and cause a series of ethical issues. The concept of strict technical secrecy and technology neutrality under the theory of technology as a tool obviously contradicts the purpose of risk prevention and control of risk social governance, and cannot be used as the basic value paradigm of technology regulation in the age of risk. With the increasing penetration of technology into human life and its pervasiveness, the disadvantages caused by the development of science and technology have become increasingly apparent, and human fear and anxiety about the development of technology have increased day by day. Against this background, the technical pessimism advocated by the theory of technological entities is rising (Tilly, 2000). This theory stands on the opposite side of the theory of technology as a tool, and claims that technology is not neutral, but is rooted in other cultural categories and is closely connected with various fields (Wang, 2006). The theory of technological entity opposes technological progress and advocates a return to the natural state of innocence, which seriously deviates from the tide of technological development and social progress and will inevitably withdraw from the historical stage (Liu, 2011, pp.68-78). The governance of a risk society stems from many hidden dangers in the highly developed technology, and always adheres to the attitude of actively responding to technological risks. The technical pessimistic attitude held by the technology entity theory runs counter to the concept of risk social governance and cannot solve the technical regulation problem in the risk society. After a brief development, the technical entity theory has been replaced by the technical critique theory. The technical critique theory lies between technology supremacy and technology pessimism. It neither advocates technology neutrality without regulating the technology itself, nor does it hold a negative pessimism Evading technological risks, this theory advocates taking appropriate regulatory measures while actively facing technological progress, in line with the value pursuit of risk society (Liu, 2011, pp. 68-78). Therefore, it can be used as the value concept of technical regulation in risk social governance.

The starting point of risk social governance is not limited to case balance, but to focus on the root cause of the problem and the prevention and control of technological risks from a macro perspective to protect social welfare. Algorithmic problems can be attributed to the risks and social conflict brought about by a kind of technology. Contemporary society should consider algorithmic technology in the context of risky social governance when advancing legislation and technology. Adhere to the attitude of technological prevention and control, and adopt technology critique. As the value concept of algorithm legal regulation. In order to achieve effective prevention and control of technical risks, the law must require algorithms to be interpretable, and

algorithmic transparency can only be achieved through algorithmic interpretation, which can predict risks, implement supervision and strengthen accountability. From the perspective of the protection of private rights, the interpretability of the algorithm is the basis for guaranteeing the right to know, and being informed means being able to choose, by explaining the source of the right to know the risk, the path of risk prevention and control, and the right to choose relief. At the same time, the interpretability of the algorithm provides an objective basis for litigation by the counterparty. Because in the case of clear and sufficient evidence, the risk of losing the counterparty is relatively small, otherwise it has no confidence in winning the case and seeks legal ways to avoid risks and safeguard its rights .From the perspective of government regulation, algorithmic transparency can only be achieved through algorithmic interpretation, and the government can objectively and impartially evaluate its potential risks and predict the social conflicts it may cause in order to achieve legal and reasonable supervision. Therefore, the interpretability of algorithms is the foundation of risk prevention and control, both for relatives and for government regulation. The legal inclusion of interpretability into legislative standards can not only protect the innovation and progress of algorithm technology, but also put algorithms under the framework of legal regulations to effectively implement risk prevention and control. This institutional arrangement is in line with the value of technical criticism. It can be seen that the value concept pursued by interpretable legislation tends to be consistent with the value requirements of risk social governance contained in algorithm legal regulation. Therefore, interpretable legislation is the way to solve the root of algorithm black box and algorithm legal regulation (Jiang and Li, 2019).

3.2 Interpretability and Technical Ethics Regulation

Laws are lagging, and legislators cannot foresee new things and potential risks in a risk society. Generally speaking, emerging technologies are not within the scope of legal regulation, and often need to form technical ethics and industry rules through long-term human practice. However, this type of technology ethics does not have coercive force, and the implementation of ethical rules depends on subjective consciousness and industry self-discipline. Laws only come into being after a certain stage of ethical development, delimiting research and development boundaries for technology, and enabling the free development of technical ethical rules within the legal framework. However, the technical regulation in the context of artificial intelligence has its particularity. The speed of updating and iterative application of algorithm technology and the breadth of application are quite different from the slow development and promotion of traditional technology. Although the artificial intelligence society has not yet arrived, the current cases of algorithms that are contrary to human ethics frequently occur. Algorithm problems such as algorithm discrimination, black boxes, differentiated pricing, data security, etc, not only infringe on personal privacy, but also involve social welfare. Regulation of algorithm technology needs urgent solution. China's current law does not have an institutional arrangement to regulate algorithms. The huge difference between algorithm technology and traditional technology determines that it cannot apply traditional ethical rules. However, new technical ethics have not yet been formed, and the urgency of technical regulation has determined that we cannot wait for ethics to be generated through long practice accumulation. Therefore, the law must first intervene to define the framework for the development of algorithmic technology, to prevent the emergence of legal but unreasonable technology, to build the boundaries of ethical order through the basic values of the legal system, and to generate ethics in the reverse to avoid technological research and development out of ethical tracks.

The technical ethical value contained in algorithmic legal regulation points to the basic ethical concepts of human society, and specifically includes the basic connotations of maintaining human subject status and maintaining fairness and justice. First of all, we must clarify the subjective status of human beings. Regardless of the advanced level of technology, they are always in the status of objects. They must serve the development of human beings, but not anti-objects, gradually enslaved and controlled human life. Second, because algorithms are highly technical and professional, there are inherent hidden dangers of justice. Those who master resources and technologies can easily use their own advantages such as technical barriers to fool and control others, which is contrary to the concept of fairness and justice. Many of the current algorithmic problems often develop arbitrarily because they deviate from basic ethical values. For example, the ad targeting algorithm created by Google was tested by Carnegie Mellon University's Amit Datta using the "Ad Phisher" software for its algorithm bias in the job recommendation process. The deviation of the algorithm in the job recommendation process is largely due to the gender discrimination in the algorithm.[5] The Google Photos algorithm automatically tagged two black people's photos as gorillas, and rights holders claimed that the algorithm was racially discriminatory and caused disputes. The US COMPAS algorithm is widely used in judicial trials, but according to a ProPublica survey, the COMPAS recidivism assessment system has a tendency to discriminate against blacks in the process of assessing the risk of recidivism. Among non-recidivism offenders, blacks are twice as likely to be whites when assessing recidivism rates, which

[5] Anonymous: "Sex Discrimination" Highly Paid Position Recommended by Google ",http: / /finance. cnr. cn / gundong/20150713/t20150713 _ 519186018. shtml

can lead to more severe penalties for blacks.[6] If the law has technical ethical regulations on the algorithm before research and development and application, it can minimize the occurrence of technical risks.

There is a natural fusion between the value pursuit of interpretability and the technical ethics concepts pursued by algorithm legal regulation. In the age of artificial intelligence, information is the most valuable resource, and those who master the information often have the initiative. The main bodies of algorithm inventors and applicators are companies that have technical information or expertise in a certain professional field, while the counterparts are ordinary consumers. The gap between the two types of subjects 'ability to obtain information and technical resources is very wide, and their risk tolerance is also different. Although the counterparty has the right to remedy after the infringement consequences occur, according to the litigation principle of "who advocates, who gives evidence", the premise of exercising the right of remedy is that the counterparty knows the tort, the consequences of the tort and has the ability to provide evidence. Algorithm interpretability is the technical guarantee for the relative to enjoy the right to know. The algorithm can realize the transparency of the operation process and decision results through interpretation, so as to ensure the effective implementation of the right to know and the right to choose. The law treats algorithm interpretability as a technical regulation threshold, which is consistent with the justice of Ge Dewen (1982), the legislative setting is centered on the central purpose of making up the weak and filling the gap between the subjects, so as to realize the value pursuit of fairness and justice. Algorithm interpretability is a necessary prerequisite for technical supervision and accountability. Unexplainable algorithms cannot be recognized and understood due to technical barriers and professional restrictions, even algorithm designers cannot solve the problem of algorithm black boxes, technical supervision and accountability are even more lawless. In the long run, it will inevitably lead to the borderless expansion and overflow of algorithmic power, subvert the existing legal order, and then threaten the human subject status. The interpretable system design comprehensively considers balancing the subject's ability gap and the practical needs of regulatory algorithm technology. It is in line with the technical ethics of justice and is consistent with the value requirements of algorithm legal regulation. Therefore, the system construction of algorithm interpretable legislation has a natural value basis.

[6] By Sam Corbett — Davies, Emma Pierson, Avi Feller and Sharad Goel. A computer program used for bail and sentencing decisions was labeled biased against blacks.It's actually not that clear. https://www. Washingtonpost.com/news/monkey—cage /wp /2016 /10 /17 /can—an— algorithm—be—racist—our—analysis—is—more—cautious—than—propublicas/ noredirect = on

4. INSTITUTIONAL DESIGN OF ALGORITHM INTERPRETABLE LEGISLATION

Algorithm problems such as black boxes, hacking, and discrimination are becoming more and more frequent, but legal regulations are still lacking, and technical risks lack effective countermeasures. Algorithm interpretable legislation has emerged as the times require. The development of interpretable algorithms has laid a technical foundation for the construction of algorithm interpretation systems, and the value orientation of technical criticalism in responding to risk society has provided theoretical support for algorithm interpretation systems. At present, the algorithm interpretability legislation has the necessary prerequisites, and the introduction of the algorithm interpretation system is imperative. To establish a brand-new system, the preparation of legislation needs to be carried out from the following five aspects: legislative value, legislative principles, interpretability obligations, regulatory systems of interpretability, and liability for breach of interpretability obligations.

4.1 Legislative Value

The current law's technology legislative model is still a dual division model under the guidance of technology tool theory: legal regulation separates technology invention and application, it adopts an extreme attitude of prohibition or laissez-faire on technology invention, and is only ex post accountability for technology application behavior. In the future, the direction of technology legislation should be changed to a technology invention-promotion-application chain-based legislative model in the perspective of a risk society, taking technology criticalism as the value concept. Future technology legislation should take risk prevention and control as the value orientation, algorithm interpretation as the entry threshold for algorithm technology invention, algorithm transparency as the accountability guarantee for promotion and application, constructing a legal system of algorithm interpretability legislation.

4.1.1 Algorithm Access Principle

The core of the construction of algorithm legal system in the era of artificial intelligence is that the algorithm must be interpretable. Interpretability is the entry threshold for algorithm technology to enter the market. Algorithms in the era of weak artificial intelligence are interpretable. Algorithms in the era of strong artificial intelligence can be interpreted through the application of technologies such as recording and visualization. In principle, the application of unexplainable algorithm technology is not allowed to be utilized. Once the unexplainable algorithm is applied in the market, the administrative department will conduct post-examination and accountability of the

algorithm technology according to the algorithm risk review system. Algorithms and technologies involving personal rights and property rights should be ordered to withdraw from the market completely, and their continued application is strictly prohibited. For the application in the commercial field, following the principle of efficiency and the principle of autonomy, the algorithm inventor must fulfill a comprehensive and sufficient obligation to explain. If the other party expressly expresses willingness to continue to apply, the algorithm technology is allowed to continue in the field. Application, but the premise of continued application is that the company must improve the algorithm design, open ports to allow algorithm interpretation technology access, and accept periodic review by the administrative department. If the algorithm inventor refuses to set up the port, the administrative department or the court can forcibly set the port access interpretation algorithm according to law.

4.1.2 Algorithm Transparency Principle

The algorithm interpretation system contains the two necessary connotations of algorithm interpretation and algorithm transparency. Among them, algorithm interpretation is the threshold, which is the basis for administrative supervision and accountability. Algorithm transparency is the consequence, which is the choice of the administrative or judicial authority after balancing the interests of all parties. The standard for algorithm interpretation system is both a market entry threshold and a basis for algorithm transparency. When damage occurs, the law requires that the algorithm be interpretable as a precondition does not necessarily lead to consequences of algorithm transparency. The law balances between protecting business secrets and algorithmic transparency. After the algorithm enters the market after review, it means that the inventor of the algorithm implicitly acknowledges the transparent results of the algorithm in the future. Related departments can directly transparentize the algorithm without the consent of stakeholders in the regulatory or judicial process.

4.1.3 The Principle of Fairness

Based on the theory of balance of interests, the law cannot overbalance the user of the algorithm, it must also consider the business interests of the enterprise. The algorithm inventor can apply to the administrative department for non-disclosure. The administrative department and industry associations such as the Computer Association and the Computer Society will intervene to comprehensively review whether the reasons for the algorithm application's refusal to be publicized are established. Sign a confidentiality agreement under the supervision of the government, implement targeted disclosure and take relevant confidentiality measures. In litigation procedures, appraisers and expert assistants can make professional assessments and judgments as a guarantee of algorithmic technical confidentiality. The confidentiality clauses stipulated in the current "Contract Law"[7] reflect the value orientation that the law focuses on protecting user rights while not sacrificing commercial interests. The construction of a future algorithm interpretation system echoes the current regulations and constitutes a strict and complete legal system.

4.2 Interpretability Obligations

Interpretability, as an entry threshold for the algorithm market, should be incorporated into the scope of administrative law as an administrative license. The algorithm inventor has the obligation to disclose information to the special supervision department, that is, to regularly disclose the algorithm decision principle and reference factors and their coefficients in the decision process. In order to protect the business secrets and technical interests of the company, this disclosure obligation does not involve the content of the source code, as long as its decision-making process can be explained, and it is verified that the algorithm conforms to technical ethics and does not infringe on public welfare. If the algorithm inventor violates this obligation, the algorithm is not allowed to enter the market. If he violates his obligation to enter the market without permission, he shall bear corresponding legal liabilities. In the process of signing a technology contract with an algorithm inventor, algorithm users must do their due diligence to check whether the other party has an administrative license certificate, otherwise the contract is invalid and their trading behavior is not protected by law.

4.3 Regulatory Systems of Interpretability

A statutory obligation must have a complete supervision system in order to be realistic and feasible. The supervision system includes both ex-ante supervision and ex-post supervision. The ex-ante supervision of algorithm interpretability legislation focuses on examining whether technology meets the entry threshold for interpretability. The implementation of this work requires the establishment of a special supervision department under the Ministry of Industry and Information Technology. Algorithm inventors must report their research and development progress to the special regulatory department and accept spot checks and supervision by the special regulatory department. The content of their review is limited to whether the algorithm decision process can be interpreted and whether there is a decision basis that violates laws or ethics. After a preliminary review, the special department will report the algorithm that meets the requirements of interpretability to the Ministry of Industry and Information Technology for review and approval. The

[7] Article 43 of the "Contract Law of the People's Republic of China": "The trade secrets that the parties have learned during the process of entering into a contract shall not be disclosed or used improperly, whether or not the contract is concluded. Shall be liable for damages."

algorithm that does not meet the requirements will not be licensed and will cooperate with the technical association to inform its improvement standards. In addition to the prior supervision, the special supervision department must regularly check and record the algorithms on the market. Once it finds that an unexplainable algorithm has entered the market, it must start a post-mortem supervision procedure to order the application to withdraw the algorithm from the market, and carry out administrative penalties such as fines, warnings, and compensation for losses caused to third parties. In serious cases, even criminal liability must be assumed.

4.4 Liability for Breach of Interpretability Obligations

Because the interpretable obligation belongs to the scope of administrative law, when the algorithm inventor uses the unexplainable algorithm without the administrative permission, he or she must bear administrative responsibilities such as fines and confiscation of illegal income. If the right holder of the algorithm does not exit the market for rectification within three months after the administrative penalty, the special supervision department has the right to forcibly terminate all its business activities and register the algorithm on the technical blacklist, restricting it from obtaining administrative licenses within five years .It is strictly forbidden to enter the market during five years, and the offender is forcibly cancelled.

After the algorithm enters the market, its owner will have a civil legal relationship with the equal subject. The legal relationship of algorithms includes both the legal relationship of the technical contract signed by the applier of the algorithm and the inventor and the legal relationship of the contract resulting from the use of the algorithm product by the consumer. In the former technical contract legal relationship, after the unexplainable algorithm was ordered to exit the market, the contract was invalid from the beginning. If the user has fulfilled a reasonable review obligation, the inventor must bear the responsibility for negligence of the contract and compensate the user for economic losses; otherwise one party does not need to bear responsibility for the other party. If the legitimate rights and interests of a third party are violated due to the breach of the interpretability obligation by the algorithmic subject, the algorithmic subject shall bear civil liabilities such as stopping the infringement and compensating for losses. Specifically, algorithm inventors should assume responsibility according to the principle of no-fault liability, while algorithm users should assume responsibility for the fault. Due to the rapid development of science and technology, it also brings more risk factors, the fault is more difficult to prove, and the conditions for judging the causal relationship between the behavior and the damage result are insufficient. According to the Tort Liability Law, The fault condition's imputation conditions are too strict to achieve relief, which is contrary to the legislative purpose and value orientation of the law (Wang, 2008). In algorithm cases, subjective faults of algorithm inventors cannot be inferred, and evidence proving the existence of causality cannot be obtained. Therefore, in the process of designing the system around the responsibility of the inventor, the algorithm interpretation system should adopt the resolution method without fault liability to achieve the right relief for the victims (Wang, 2001, p.34.).

After the algorithm violates the obligation of interpretability to enter the market, the reference factors and content included in its decision-making process are in a black box. This process may violate the privacy of the right holder, and the results of algorithm may contain content that violates human rights or harms the public interest, such as unfair preferences and discrimination. In this case, the algorithm inventor and the algorithm applicator should be convicted and bear criminal responsibility separately. The algorithm inventor should be inferred as an intentional crime, the subjective aspects of the algorithm applicator should be determined by the situation. At the time of conspiracy, they were presumed to be intentional and punished by common crime.

CONCLUSION

Science and technology legislation is a good way to deal with algorithm problems in the age of artificial intelligence and effectively prevent risks. The algorithm interpretation system starts from the root problem of algorithm black box, and uses interpretability as a threshold to achieve legal regulation of algorithm technology. Just as Habermas said that "right is a social construct", the risk society under the background of artificial intelligence is quite different. The power construct under the previous legal framework was no longer adapted to the development of modern society. Therefore, under the new legal system, the regulation system needs to be constructed to adapt to social trends. The rights and obligations system constructed by the algorithm interpretation system is in line with the technological status quo of the modern risk society and balances the needs of various stakeholders. In the future, it will promote the development of algorithm technology and promote the continuous upgrade of China's algorithm technology in the global artificial intelligence industry chain.

REFERENCES

Baker, Giddens, & Rush (2001). *Reflexive modernization* (p.76. W. W. Zhao, Trans.). Commercial Press.

Cano, A., Zafra, A., & Ventura, S. (2013). An interpretable classification rule mining algorithm. *Information Sciences, 240*(08), 1-20.

Cheng, Y. (2019). Data protection and algorithm regulation

under the meta-regulation mode——Taking the European Union's "General Data Protection Regulations" as the research sample. *Law Science (Journal of Northwest University of Political Science and Law), 37*(04), 48-55.

Chu, Q. W. (2014). *Artificial intelligence from the perspective of philosophy.* Wuhan University of Technology.

Gao, Q. Q., & Zhang, P. (2018). On the multidimensional challenges of artificial intelligence to future laws. *Journal of Huazhong University of Science and Technology (Social Science Edition),* 32(01), 86-96.

Ge Dewen, W. (1982). *Theory of political justice* (Vol. 1, M. L. He, Trans.) Beijing: Commercial Press.

Guo, Y. D. (2017). *Philosophical thinking on artificial intelligence*. Harbin University of Science and Technology.

Hobbes (1985). *Leviathan* (p.72. S. F. Li & T. B. Li, Trans.) Commercial Press.

Jiang, Y., & Li, Y. J. (2019). Breaking algorithm black box: functional proof and application path of algorithm interpretation power—Taking social credit system construction as the scene [J/OL]. *Journal of Fujian Normal University (Philosophy and Social Sciences Edition)*, (04), 84-92+102+171-172[2019-08-12]. http://kns.cnki.net/kcms/detail/35.1016.C.20190722.1523.018.html.

Krull, J. A., Huey, J., Barlockas, S., Felton, E. W., et al (2019). Accountable Algorithm. *Research of Local Legislation, 4*(04), 102-150.

Liu, S. X., Wang, X. T., Liu, M. C., & Zhu, J. (2017). Towards better analysis of machine learning models: A visual analytics perspective. *Visual Informatics, 01*(03), 48-56.

Liu, T. G. (2011). The paradigm shift of technical regulation foundation in risk society. *Modern Law Science, 33*(04), 68-78.

Liu, T. G. (2011). The paradigm shift of technical regulation foundation in risk society. *Modern Law Science, 33*(04), 68-78.

Liu, Y. H. (2019) .Research on algorithmic bias and its regulation path. *Journal of Law*, *40*(06), 55-66.

Luo, Y. Z. (2003). Strategic thinking on perfecting china's science and technology legal system. *Science & Technology*, (01), 1-7.

Ma, M. (2006). *Research on data-driven fuzzy system modeling method*. Jilin University.

Olah, C. (2018). *The building blocks of interpretability.* https://distill.pub/2018/building-blocks/, accessed on March 15, 2018.

Perrow, C. (1994). Accidents in high -risk system. *Technology Studies, 1*, 23.

Roque, J. V., Cardoso, W., Peternelli, L. A., & Teófilo, R. T. (2019). Comprehensive new approaches for variable selection using ordered predictors selection. *Analytica Chimica Acta, 1075*(10), 57-70.

Si, X., & Cao, J. F. (2017). On the civil liability of artificial intelligence: taking autonomous vehicles and intelligent robots as the entry point. *Science of Law (Journal of Northwest University of Political Science and Law)*, *35*(05), 166-173.

Sun, J. L. (2019). Research on legal regulation of algorithmic automated decision risks. *Research on Rule of Law*, (04), 108-117.

Tencent Research Institute, Internet Law Research Center of China Academy of Information and Communication Technology, etc. (2017). *Artificial intelligence-the national artificial intelligence strategic action grasp*. Beijing: Renmin University of China Press.

Tilly (2000). *History of western philosophy* (p.428. L. Ge, Trans.). Beijing: Commercial Press.

Wang, K. (2019). Legal responsibility in human experiment of "Gene editing baby" —— A Hermeneutic analysis based on China's current legal framework. *Journal of Chongqing University (Social Science Edition), 25*(05), 134-144.

Wang, L. M. (2008). Construction of China's tort liability system—Thinking about relief law. *China Law Science*, (04), 3-15.

Wang, N. D. (2009). A review of the theory of "Reflexive Modernity". *Foreign Theoretical Trends*, (07), 98-102.

Wang, Y. H. (2006). Phenomenological reflections on the essence of technology and its possible future: Heidegger's philosophical questioning of technology in his later period. *Journal of Peking University Graduate Studies,* (1), 71.

Wang, Z. W. (2001). *Tort law* (Vol.1, p.34). China University of Political Science and Law Press.

Wu, H. D. (2017). Institutional arrangement and legal regulation in the age of artificial intelligence. *Social Science Abstracts*, (12), 76-78.

Wu, H. D. (2017). The institutional arrangement and legal regulation in the age of artificial intelligence. *Science of Law (Journal of Northwest University of Political Science and Law), 35*(05), 128-136.

Yang, Q. F. (2003). *Technology as the purpose*. Fudan University.

Zhang, L. H. (2019). Iteration and innovation of algorithm regulation. *Forum on Law*, *34*(02), 16-26.

Zhang, L. H. (2018). Research on algorithmic interpretation power of business automation decision. *Science of Law (Journal of Northwest University of Political Science and Law)*, *36*(03), 65-74.

Zhang, L. H. (2019). Iteration and innovation of algorithm regulation. *Law Forum, 34*(02), 16-26.

Zhang, L. H. (2019). The rise, alienation and legal regulation of algorithmic power. *Research on Law Merchants, 36*(04), 63-75.

Zheng, G. (2018). Algorithm law and legal algorithm. *China Law Review*, (02), 66-85.

Zong, W., Chow, Y. W., & Susilo, W. (2020). Interactive three-dimensional visualization of network intrusion detection data for machine learning. *Future Generation Computer Systems, 102*(01), 292-306.

Zou, K. L., & Liu, J. M. (2018). The dilemma and outlet of the legal regulations of big data "Killing Familiarity" —— Only from the perspective of the "Consumer Rights Protection Law". *Price Theory and Practice*, (08), 47-50.