

Technium.

43/2023

2023
A new decade for social changes

Technium
Social Sciences



Risk and Legal Regulation of Algorithm Application in Insider Trading Supervision

Chen Laiyao

PhD student at the School of Law, University of International Business and Economics, majoring in Financial Law, China

1157848518@qq.com

Abstract. Insider trading is a kind of information manipulation behavior in the securities market. The insider trading has brought great damage to the securities market in recent years, so securities regulatory authorities have begun to crack down on this kind of illegal behavior. With the application of algorithms in the field of supervision, regulators can accurately identify insider trading behaviors through big data analysis and other technologies, with which the efficiency of supervision was greatly improved. However, the application of algorithms in the supervision of insider trading is prone to cause various legal risks, such as the inaccurate transformation between algorithms and regulatory regulations, the imperfection of algorithms, the infringement of private data rights, etc. The legitimate rights of the regulated objects might be endangered by above risks. Therefore, it's necessary to establish risk prevention and legal regulation to cope with the algorithms in the filed of insider trading supervision from three aspects of rule transformation, technical supervision and data security. In this way, the property rights, data rights, privacy rights of financial investors, listed companies and other stakeholders are protected.

Keywords. Insider trading; Algorithmic risk; Legal regulation; Data Security

I. Introduction

Insider trading is a serious information asymmetry behavior in the securities market, which damages the legitimate interests of investors in the financial market.¹Such behavior will cause harm to the equity of the securities market, thus to be a main type of crime being cracked down on by the national securities regulator. Although the securities supervision has implemented the insider registration management system, it is still difficult for the regulatory authorities to detect and supervise the insider trading due to its strong concealment and great temptation. Consequently, the insider trading has been repeatedly prohibited for a long time and became a major problem in the field of securities supervision. According to “The Regulatory Technology Overall Construction Plan” issued by Chinese Securities Regulatory Commission(CSRC) in 2018, the Big Data platform 3.0 version of the securities market had

¹ Mayer-Schönberger, V., & Cukier, K. , ‘Big data: A revolution that will transform how we live, work, and think’, Houghton Mifflin Harcourt(2013),at 130-135.

been built.² As the core of Big Data platform that had been applied to the supervision of insider trading, It is possible to regulate insider trading behavior through algorithm. By comparing the major information, insider information, event period and other data disclosed by listed companies, the hidden insider trading in securities market could be discriminated. As a result, timely and accurate illegal warning signals are provided for securities supervision. Due to a great deal of trading information in the securities market, the data collection and classification depend on algorithm in the automatic supervision system. With the participation of algorithms, the data of the listed company's equity structure, financial information, market profit and other aspects are converted into various machine-executable indicators, which makes it possible to obtain early warning about insider trading. Similar practices exist in foreign financial supervision, such as the US Securities and Exchange Commission using continuous algorithms to detect misconduct. The development of digital technology in the field of financial regulation is a double-edged sword with both advantages and disadvantages. It will harm the legitimate interests of regulatory subjects if being used improperly. In order to ensure the effectiveness and fairness of the insider trading regulation system and enable regulatory technology to play a greater role in market supervision, it's of vital importance that algorithm is under legal supervision.

The research on the theory and system of algorithms regulation is not a new topic in the academic world. Most Scholars prefer to study the application of the regulation technology in financial area, rather than the supervision. For example, He Haifeng et al. (2018) studied the application of regulatory technology in securities regulation; Guo Yanfang (2020) studied the application of regulatory technology in informational market manipulation behaviors. From the perspective of insider trading regulation, there are few papers related to the method research of intelligent algorithms to curb the insider trading behavior. As a technical means, algorithms has aroused a lot of discussion in practice and academia. It is frequently used in national management system over the years with great development, especially in the field of financial governance. However, there are few studies on the key of regulatory technology, self-regulation and external regulation of algorithm. In the field of regulatory science, the way of algorithmic regulation is different from other algorithmic governance of government management technology or financial management technology. It is necessary to regulate the core algorithm of regulatory science with new research methods and paths particularly in view of the algorithm characteristics of regulatory science, so as to prevent the algorithmic risks and possible financial risks in a better way. Hence, the characteristics of the algorithms of insider trading regulatory behaviors are summarized in this parper, with proposal of the regulation methods and regulatory models of the algorithms.

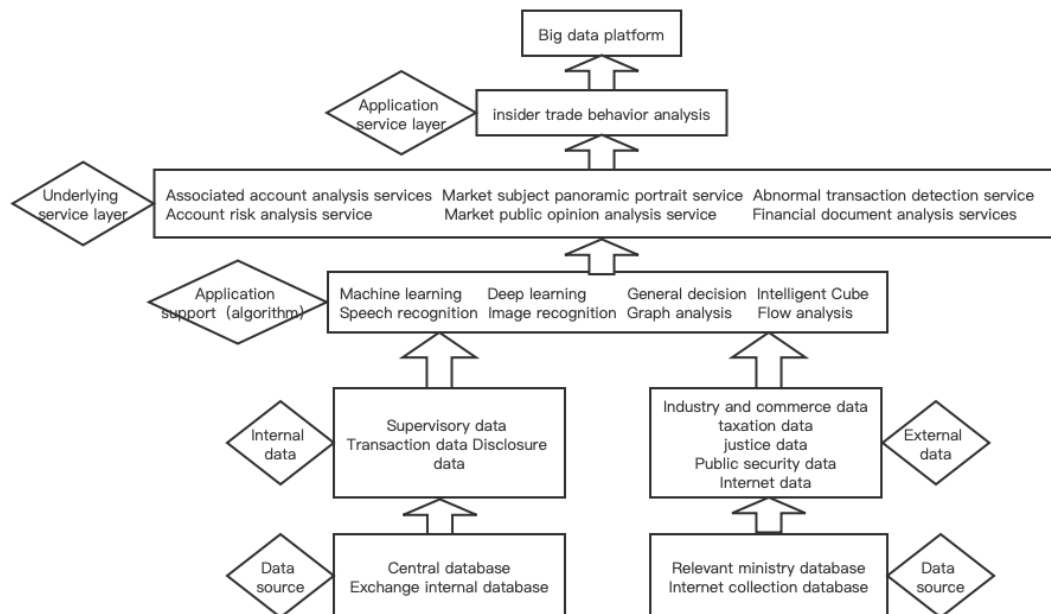
II. Application characteristics of algorithm in insider trading supervision

Insider trading is a kind of information manipulation behavior in the securities market as well as a kind of financial illegal behavior being crack down on by the CSRC and the stock exchanges. Nevertheless, the characteristics of insider trading, like large amount of money involved, strong anti-detection, good concealment and changeable forms, make it difficult for traditional supervision to curb such behavior. The development and application of intelligent

² Chinese Securities Regulatory Commission, 'China Securities Regulatory Commission Officially Issues and Implements the Overall Construction Plan of Regulatory Technology', August 31 ,2018, http://www.csrc.gov.cn/pub/newsite/zjhxwfb/xwdd/201808/t20180831_343433.html (visited on April 1, 2023)

algorithm supervision system capacitate the regulatory authorities to accurately identify insider trading and to directly locate the accounts with "abnormal trading behavior" with the method of big data analysis.³ In regulatory activities, the associated account data for trading activities are proactively identified by algorithm, to predict the trading parties and time. Moreover, regulators are encouraged to take risk prevention measures in advance based on whether the identified trading items are illegal.⁴

Figure 1 Flow chart of insider trading supervision



A. Draw data portrait

In China, the intelligent system of investor portrait on domestic stock exchanges has been updated to the third generation with the main function of static supervision of investors in the financial securities field.⁵ The data portrait of investors established by intelligent algorithm, which means the trading behavior characteristics of investors are displayed in a graphical way. It includes many dimensions like trading style, position characteristics, investment preference and historical supervision. The BIG Data platform can collect historical data of financial investors, then summarize and integrate relevant data of different dimensions of financial investors via basic algorithms including machine learning algorithm, knowledge graph algorithm, association analysis algorithm, big data algorithm and high-order algorithm model. The trading data, investment relationship data, compliance supervision data and financial income data of financial investors are taken as the basic data set, which can be distinguished as strong correlation information or weak correlation information. An investor label system is formed based on the algorithm analysis framework, and simplified display is carried out to

³ Shanghai Stock Exchange, '2014 Self-regulatory Work Report of Shanghai Stock Exchange', 8 February 2015, <http://www.sse.com.cn/disclosure/credibility/whitepaper/c/4118097.pdf> (visited on April 1, 2023)

⁴ 3. Guo Yanfang. 'Application and System Improvement of regulatory Technology -- Based on the Perspective of Information Manipulation', 03Southwest Finance 2020, at 51.

⁵ ZHANG Xin, 'The Framework, Features and Legalizing Path of Algorithmic Regulation: A Discussion on the Personal Information Protection Law (Draft)', 14(01)Business and Economic Law Review 2021, at 30-35.

facilitate supervisors to quickly grasp the abnormal trading account information.⁶ The regulatory authorities also adopt an "account association analysis model" to deeply integrate the multidimensional data of the account association.⁷ The intelligent algorithm establish account association topology by mining various association information between accounts and account groups. Under assistance of this model, the regulators can quickly identify the suspected account groups and eliminate all suspected accounts of insider trading in one shot. In information transmissional insider trading, there are usually "nepotism" and "group" insider trading involving relatives, friends, classmates, colleagues and other related groups. Through the "account correlation analysis" algorithm model, the associated groups of insider trading accounts can be identified at once, to greatly improve the efficiency of supervision.

B. Monitor trading behavior in real time

Due to the algorithm powerful data capture ability and analysis ability stock exchanges can supervise all trading information of investors' stock accounts. Most data are stored in the Intranet, including account data, trading data, behavioral data and so on. Generally speaking, the supervisory goals can't be reached by the algorithm with only these Intranet data. That's to say, the insider trading behavior won't be identified by these intranet data which can't be enough to mine characteristics of insider trading behaviors. Therefore, it is necessary to collect external data, especially internet data. After the establishment of the central regulatory information Platform, the behavior data of account holders in the external network (Internet) will be included in the supervision. Combing the behavior data with the account trading data in the internal network, and using specific algorithmic models to compare and analyze the behavior data of account holders on the internal and external networks, so as to more accurately identify insider trading behavior. In insider trading, information transmissional insider trading is the main type of illegal trading in the Chinese securities market. There are a large number of people involved in this type of insider trading. Usually, the insider information is transmitted to relatives, friends, partners, school teachers, drivers, nannies and other personnel by the legal informant. In such complex case, the securities market regulators typically rely on algorithms to identify all kinds of trading information in databases containing insider information and stock account information, such as database of listed company board of directors' information, senior executives' relatives, senior executives' stock accounts and their relatives' stock accounts, as well as enterprise employees' stock account. By analyzing the types of insider information, the insider trading algorithm model directly provides qualified suspect accounts with the combination of the company stock price fluctuations, account transaction characteristics and other information. Besides, it can automatically generate analysis reports without the need of manual identification, which is a great improvement of the supervision efficiency. Normally, manual presetting rules are not needed for algorithm. It's based on the case database of CSRC, stock exchange and other regulatory agencies, extracting data features from massive case data

⁶ Wen qicong, Smart regulation builds a strong defensive line for the stock exchange, Economic Daily News, February 2018, http://www.ce.cn/xwzx/gnsz/gdxw/201802/10/t20180210_28143113.shtml (visited on April 2, 2023)

⁷ 'Empowering Supervision with Technology -- The Intelligent Supervision System for Enterprise Portraits independently developed by Shenzhen Stock Exchange went online', Securities Times Website, October 2018, <http://money.people.com.cn/n1/2019/0802/c42877-31271375.html> (visited on April 2, 2023)

to form a behavior feedback mechanism.⁸ Then the extracted data characteristics are applied to securities trading behaviors, so as to accurately identifying the insider trading behavior through probability statistics, to provide regulators with early warnings.

C. Public opinion monitoring scenarios

Currently, the construction direction of domestic financial regulation technology is the construction of financial basic analysis capability under various financial scenarios. Using semantic analysis, text mining and other functions of algorithms can help regulators identify and monitor insider trading in financial markets by analyzing unstructured or structured data such as online public opinions, corporate financial reports and brokerage research reports.⁹ The Office of Fraud Detection and Market Intelligence (OFDMI) of the American Financial Industry Regulatory Authority (FINRA) developed a software called "Securities Watch News Analysis and Regulation (SONAR)" to combine transaction supervision and public opinion analysis with big data algorithmic analysis capabilities to quickly target insider trading behavior. By learning SONAR system from the US, The Chinese information management department of the Stock exchange (Financial Innovation Lab) and other departments also jointly developed a function of the BIG Data platform, which is called text analysis and mining function. In this function, Chinese characters are used as the smallest unit to analyze the meaning of different combinations of Chinese characters. The "word cutting" program means to cut the text information into a way that can be read by machine, so that the algorithm can automatically complete the automatic monitoring of unstructured text information such as network information and news public opinions, especially the Social networks shared resources. Through big data algorithm, the CSRC simultaneously monitors the trading data from the internal network, the behavior information of traders and news public opinions from the external network. By comparing the text information content with the trading data of the users, the CSRC makes correlation analysis to determine the existence of major news that may affect the stock price volatility and the possibility of insider trading behavior.

III. Legal risks caused by the algorithms in insider trading supervision

A. The disconnection between code writing and insider trading regulatory rules

There are also similarities in essence between algorithmic governance and legal governance. Through certain logical steps, existing factors are analyzed to obtain corresponding judgment results.¹⁰ There should be mutual transformation between computer code and regulatory rules, which regulatory rules can be identified protocolically and specific regulatory requirements can be decomposed into computer code without any rule distortion.¹¹ In the specific law enforcement scenario, algorithm is the implementation subject of insider trading supervision system. However, in practice, algorithm supervision as a law enforcement method is also with many problems to be solved. Firstly, the principle of "equity" requires that the legal

⁸ Zheng Zhihang, 'Ethical Crisis and Legal Regulation of Artificial Intelligence Algorithm', 39(01)Legal Science (Journal of Northwest University of Political Science and Law)2021, at 16.

⁹ Liao chang, 'Big Data analysis of social networks in capital markets: a regulatory application perspective', 38(01)Trading technology frontier2020, at 74.

¹⁰ Henry Brooke. 'Algorithms, Artificial Intelligence and The Law', November 12, 2019, <https://www.supremecourt.uk/docs/speech-191112.pdf>

¹¹ Zhang Song, 'Exploration on the application of Artificial Intelligence in technology financial regulation', 04China Credit Card2018, at 52.

system to be consistent with its basic principles in the use or implementation process. So the law should be presented in a flexible form during the implementation process, and all legal behaviors should not deviate from the legal principles. However, in the self-learning process of the regulation of insider trading algorithm, the algorithm structure can only be optimized by summarizing experience according to the input data. In the process of self-learning, algorithm supervision can only summarize experience and optimize the algorithm structure based on input data, rather than adjust the data processing process of the algorithm through legal principles. This can easily lead to result inconsistent with the principle of fairness and justice. Secondly, insider trading regulation rules are based on legal knowledge, thus there are certain difficulties in the conversion between them and algorithm programming. Since the identification of insider trading is mostly based on experience, for example, analyzing insider trading through the relationship between stock price fluctuations and abnormal trading behaviors. Therefore, the aim of algorithm regulation is to convert the long-term accumulated experience of regulators into machine executable code, rather than simply converting legal knowledge into code symbols. This leads to the absence of a reliable algorithm-legal writing template in the process of programming, which is tend to change the original meaning;¹² Thirdly, text files for algorithm analysis include Chinese documents and documents in other languages (overseas traders data, audit data, communication records, etc.). Whereas there are different linguistic modes among different languages, obvious differences exist in semantic logic and grammatical rules among English, Chinese and Latin, which makes it difficult to convert into a unified algorithm format. The algorithm procedures of insider trading supervision in different countries are prone to be different, which is not conducive to unify international supervision or international cooperation.

B. The technical risk of algorithm

As a means of regulating insider trading by regulatory authorities, algorithm regulation is similar legal regulation although in form; but in substance, the differences are significant. The algorithm framework of insider trading regulation is short of effective supervision, resulting in discriminatory and unfair data analysis results in its operation and use. Algorithms are designed by people and endowed with the learning ability of human-like thinking, some technical risks may arise under the influence of internal and external factors.¹³ At present, algorithmic biases and algorithmic "black box" are the main technical risk sources in algorithmic behavior. The prime reason for the algorithm bias lies in the R&D stage and deep learning function. Although the developers of the algorithm framework of insider trading regulations have rich technical knowledge, they have few experience in insider trading supervision, incomplete understanding of the regulatory rules of insider trading, and insufficient experience in judging the characteristics and attributes of insider trading, trading behavior significance, and position style preferences. Therefore, when developing the regulatory framework of insider trading algorithm, the proportion of each index parameter may be unreasonable, resulting in the unfair insider trading risk level index of listed companies or investors, which is not conducive to exerting the impact and deterrent effect of insider trading algorithm regulation. In addition, the machine learning algorithm can learn and grow through the input of regulatory data, and the algorithm model is constantly changing according to the learning, in pursuit of maximum efficiency to achieve regulatory goals. If the input data is not

¹² Zheng Ge, 'The Law of Algorithm and the Algorithm of Law', 02*Chinese Law Review* 2018, at 75.

¹³ Cui Jingzi, 'The Crisis and Countermeasures of Equality Protection under the Challenge of Algorithm Discrimination', 37(03)*Legal Science (Journal of Northwest University of Political Science and Law)*2019, at 39.

representative, the regulatory results being pursued are also biased.¹⁴ The primary cause of algorithm "black box" is that the algorithm source code is not open and the algorithm decision process is too secret. On the one hand, the complexity and professionalism of the algorithm rules make it impossible for ordinary people to understand its internal mechanism, which easily leads to doubts about the correctness of regulatory results. On the other hand, the opacity of the algorithm data processing process makes the artificial manipulation (research and development layer) and bias influence (algorithm learning layer) more hidden in the regulation of insider trading algorithm. In the process of identifying the behavior of insider trading, if the identification standard of algorithm cannot be explained to meet the regulations, the compliance and reliability of algorithm regulation will be questioned.¹⁵

C. The infringement of algorithm on private data security

"Essentially, Algorithm is a data analysis and processing mechanism."¹⁶ Insider trading has a strong concealment and dissemination in the securities market, and there is a special regulatory demand for law enforcement business activities. Therefore, in terms of regulatory methods and means, algorithm needs to be premised on the collection and analysis of a large amount of data, including personal data related to financial investors and sensitive data related to listed companies.¹⁷ In the supervision system of insider trading algorithm, some social tools, online community platforms, mobile communication and other Internet data sources are the key objects of supervision. By comparing the short videos, posts, articles, and group messages and chat records of social tools released by the supervised objects on the network platform; plus the account trading situation of the stock exchange, the account information behind the insider information disseminators can be screened out. Meanwhile, the information of listed companies is also the main supervision target of the Big Data platform. Through the supervision of the financial data, business structure, corporate governance system and other data of listed companies, the relevant characteristics of listed companies in insider trading events could be extracted; such as the characteristics of corporate finance, business development, corporate governance and other aspects are extracted, finally to predict the possibility of listed companies and their internal personnel committing illegal insider trading activities. Normally, the index of insider trading algorithm model only covers the data related to inside information, but in the Big Data platform, it still involves the data related to non-insider sources or listed companies, especially data exposure in the privacy aspect. If inappropriate data types are used as the index of the algorithm regulation model in the machine learning algorithm model, there will be some serious consequences: Firstly, excessive exposure of private entities' private data will cause infringement of individual data rights, such as privacy rights;¹⁸ secondly, the company's

¹⁴ Cui Jingzi, 'The Crisis and Countermeasures of Equality Protection under the Challenge of Algorithm Discrimination', 37(03)*Legal Science* (Journal of Northwest University of Political Science and Law)2019, at 40.

¹⁵ Yuan Kang, 'Institutional Response to the Black Box of Financial Technology Algorithms under the Concept of Social Supervision', 01*Journal of University of Science and Technology of China* (Social Sciences Edition)34, 2020, at 103.

¹⁶ Wang Huaiyong, 'Algorithmic Risk and Legal Regulation of Financial Technology', 01*Review of Political Science and Law*, 2021, at 110.

¹⁷ Zhang Xiaoshan, 'Legal Regulation of Discrimination in Artificial Intelligence Algorithms: European and American Experience and China's Path', 6*Journal of East China University of Science and Technology* (Social Science Edition)34, 2019, at 67.

¹⁸ Brenner S W, Clarke L L, 'Fourth Amendment Protection for Shared Privacy Rights in Stored Transactional Data', 14*JL & Pol'y*, 2006, at 213-214.

business data, strategic plans and other important data will be exposed, endangering the commercial activities of listed companies.

IV. Construction of algorithmic supervision framework for insider trading

A. Rule level: framework of legal principles and algorithmic framework of regulatory rules

1. Implantation of legal principles in algorithm architecture

Sometimes, words with multiple meanings are used in the legal provisions to maintain the inclusiveness and flexibility of the law. For example, in the supervision of insider trading, insider information is a scope that can be expanded and not limited to the the board of directors and senior executives. All people who may know insider information need to be contained in the supervision of insider information, including securities regulators, administrative examination and approval personnel, legal service or audit service personnel and other personnel who may know the inside information. The expanded interpretation of the insider concept is based on the long-term accumulation of the database in stock exchange, relying on algorithms to accurately identify and locate specific information in numerous databases. However, in this process, personal privacy information is inevitably involved, which infringes on personal privacy and data rights. Thus, as the executor of the legal norms of insider supervision, algorithms should be set certain internal constraints to limit its own behavior for the state to govern the securities market. Nonetheless, the situation in the securities market is complex; to make the execution of the algorithm more flexible and obtain more optimized results, the algorithmic code needs to be adjusted according to the market situation. For example, the regulatory index of the algorithm framework should be increased or decreased depending on the changes in the regulatory situation. Legal principles need to be embedded to control algorithmic behavior. In other words, when the algorithm fully utilizes the functions of deep learning and data mining, its behavior is consistent with the basic legal principles. The implanted legal principles are generally international legal principles, including the “*Personal Information Protection Law*”, the “*Data Security Law*”, the “*Securities Law*” and other relevant departmental laws. By symbolizing the relevant legal principles and writing them into the algorithm program, the algorithm supervision of insider trading in the securities market can be realized with the inherent constraints of the algorithm.

2. Transformation of insider trading regulatory rules in algorithmic mode

Insider trading regulation rules are text documents based on natural language. In practice, it is necessary to cooperate with the insider supervision experience of regulators to maximum the efficiency of insider trading regulation. The Big Data platform needs to convert legal rules and regulatory experience into algorithmic rules, using algorithmic rules to predict and identify insider trading behaviors.¹⁹ However, it is laborious and prone to failure to simply encode knowledge into algorithmic rules. Such transformation rely on the autonomous learning function of the algorithm. The regulation rules and experience of insider trading can be transformed into algorithmically executable rules by means of reverse deduction and analogical reasoning.²⁰ Reverse deduction refers to inputting specific insider supervision cases into the

¹⁹ Pedro Domingos, ‘The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World’, *basic books*, 2015, at 101-110, 232-258.

²⁰ Decision trees are a machine learning approach. It is a tree structure in which each internal node represents a judgment on an attribute, each branch represents the output of a judgment result, and finally each leaf node represents a classification result. The decision tree analyzes the data and outputs accurate predictions.

algorithm, which uses feasible induction methods to sort out the best case features (rules) and incorporate them into the main algorithm. After forming a decision tree of the rule set, new judgments and decisions can be made based on the extracted rule set for the new factual situation. Analogical reasoning measures the similarity between cases, determines the degree of similarity and applies case characteristics (rules) to similar factual situations. Based on the database of insider trading cases, the algorithm uses analogical logical reasoning methods to find similar factual situations, associates the corresponding characteristics of insider trading, and predicts the insider trading behavior in the securities market to raise the alarm.²¹

However, the defects of the two reasoning methods mentioned above lie in that the extracted or applied case features (rules) are easily exceeded the legal scope, and the case features (rules) tend to lose their original meaning in the reasoning process. In view of the limitations of the algorithm in the conversion process of inside supervision rules, legal personnel should cooperate with technical personnel to systematize and accurately express the supervision knowledge in code language, and smooth out the logical relationship of legal knowledge points. Otherwise, the algorithm will make inaccurate or inconsistent decisions.

B. Technical level: joint regulation of limited transparency principle and algorithmic interpretability

"Algorithms are a self enclosed island."²² Without professional computer knowledge, the public (regulated entities) have difficulty understanding the regulation methods and operation logic of algorithms, leading to doubts about the behavior of algorithms in "exercising" public power to regulate the securities market. Although the function of algorithmic deep learning is in the clothing of technology neutrality, its systematic self-learning and automatic decision-making are easy to cause the suspicion of unfair behavior by algorithms, which is believed to be the culprit contributing to the continuous marginalization of minority groups. At the technical level, the algorithm transparency can be improved by disclosing the algorithm source code, and the algorithm explainability can be enhanced under the such condition. The combination of the two can effectively regulate the algorithm misconduct.

1. Limitations of algorithm transparency principle

Disclosure of algorithmic source code is the most discussed requirement for algorithmic transparency.²³ Source code is the code of programs written by algorithm developers in computer language. It is a computer instruction that human can read, containing thousands of code symbols with specific meanings.²⁴ In the Big Data platform of algorithm, the source code is equivalent to the elements or modules constituting the regulatory system, which is converted into the target code (binary mode) by compiling program. Algorithm achieves the regulatory goal by reading target code to execute tasks. That is, the algorithm is the idea of the regulatory system, and the source code is the concrete expression of this idea.²⁵ Nevertheless, the disclosure of source code often involves the protection of intellectual property rights or trade

²¹ Zheng ge, 'Algorithmic law and algorithmic law', 02*Chinese Law Review*, 2018, at 82.

²² Zheng Zhihang, 'Ethical Crisis and Legal Regulation of Artificial Intelligence Algorithm', 01*Legal Science (Journal of Northwest University of Political Science and Law)*39, 2021, at 20.

²³ Wei Yuanshan, 'The Loss and Regression of Algorithm Transparency: Functional Localization and implementation Path', 01*Northern Law*15, 2021, at 152.

²⁴ Wang Qinghua, 'Multiple Dimensions of Algorithm Transparency and Algorithm Accountability', 06*Comparative Research*, 2020, at 169.

²⁵ Lou Xianying, Fu Guangrui, et al, 'Legal Issues in Digital Transformation -- Starting from the Source Code of Software', *KWM Research Institute*, January 8, 2021, <https://www.infoobs.com/article/20210108/44652.html>

secrets.²⁶ Especially in the case that the securities regulatory authorities entrust a third party to develop the insider trading supervision system, the algorithm developers independently develop the supervision system and have the rights of the source code but are reluctant to disclose. When delivering the regulatory system, algorithm developers generally only license the client's right to use the source code, without involving the delivery of the source code, which makes the disclosure of the algorithm source code in the regulatory system lack the rights basis. In addition, open source code can easily bring algorithmic security issues,²⁷ which can easily endanger the financial safety and regulatory effectiveness of the securities market. Therefore, the third party algorithm review can be used as a compromise.

Algorithm review refers to establishing a neutral algorithm review agency composed of professional technical and legal personnel to review whether the algorithm contains discriminatory content and whether the algorithm process is fair and transparent, and publish the algorithm review results to the public. Algorithm review is conducted by professionals to review algorithm risks, rather than disclosure to the general public without professional knowledge. It can not only avoid the risk of business secret disclosure and algorithm security risks caused by the exposure of algorithm source code, but also play a reasonable role in regulating algorithm security in the supervision system. What's more, the algorithm review requires regular inspections of the supervision system, including the review of the original algorithm model and subsequent modifications and adjustments of the algorithm. So as to ensure the algorithm model to be under administrative supervision, which is also helpful to solve the dispute over the responsibility attribution of the machine learning algorithm. The algorithm review institution or department shall act as an independent administrative unit, and bear the responsibility for the algorithm filing and the fairness of the algorithm review results, which helps to improve the public's trust in the review results. Therefore, in the scenario of entrusted research and development of regulatory system, the regulatory department should clarify the acquisition of the algorithm source code authority at the protocol level, particularly in the "target source code" (black box),²⁸ and submit the source code to the algorithm review institution or department for regular inspection to prevent the technical risks of the algorithm.

2. Enhancement of algorithm interpretability

The interpretability of algorithm is a concrete manifestation of its transparency principle. In the legal system, the interpretability of algorithm is manifested as the algorithm interpretation right enjoyed by the subject to algorithm supervision. This right is determined by the legal exercise in different countries, including the General Data Protection Regulation of the European Union, the Information Commissioner's Office of the United Kingdom and the Alan Turing Institute, jointly issued the "*Artificial Intelligence Decision Instructions Guide*", and China's Personal Information Protection Law, etc., which indicates that the international community attaches great importance to the interpretability of algorithms. However, the issue of algorithm bias and algorithm "black box" has always been controversial. People have questioned whether the algorithm can make a fair decision, which leads to public doubts about the fairness of regulatory recommendations made by the Big Data platform. Consequently, the public hopes to open the "black box" by explaining the algorithm, understand the internal

²⁶ Joshua New and Daniel Castro, 'Algorithmic Accountability', Center for Data Innovation, 8 May 2018.

²⁷ Feng Changyang, 'Research on the Open Code Policy System and Content of Major Countries and International Organizations', *06Library*, 2019, at 46.

²⁸ The target code of the black box focuses only on the decision-making results of the algorithm, not the decision-making process.

operation process of the algorithm decision-making, and achieve the transparency in the algorithm decision. The construction of algorithm interpretability can be realized from two aspects: one is the explanation of algorithm decision based on case. The supervised object can request regulatory department to explain the decision made by algorithm of Big Data platform, which means to describe the causal relationship between data input and result output. The other is a functional explanation of the system for the public. The regulatory authorities take the initiative to disclose the whole process of algorithmic decision-making, and bring the operational process, logic, data types, decision indicators and other factors of the algorithm of the regulatory system into the scope of algorithm interpretation.²⁹

C. Data security: algorithmic constraints on data desensitization and data usage restrictions

In the securities market, the algorithm supervision system needs to rely on the massive database of the stock exchange, but the database covers a large amount of information of all parties involved in the securities market transaction, including personal privacy or the business activities information of enterprises, which poses significant security risks to the privacy and business information of the market participants. Therefore, in the process of insider trading supervision, data collection, storage and analysis must be desensitized to prevent the improper use or disclosure of sensitive data.

1. Hierarchical protection and desensitization algorithm for sensitive data

The premise of sensitive data protection is to define the type and scope of sensitive data, then classify the protection level of sensitive data based on the "value embodiment" and "sensitivity degree" of the regulatory data set by analyzing the subject, content, and source of the regulatory data set.³⁰ The data input into the BIg Data platform comes from a wide range of sources, numerous subjects and complex contents. There are data actively submitted by regulated entities, such as the data related to insiders; also the data obtained by the securities regulatory authorities through other channels, such as the public opinion monitoring data. The algorithm regulation extracts the required regulatory information through data mining, and predicts the possibility of insider trading risk occurring in trading accounts based on the extracted information. However, the data samples inputted by the algorithm may involve the privacy information (communication content, interpersonal relationship, property information) of the regulated subject or important sensitive data related to business secrets.³¹ Hence, for sensitive data involving important privacy information corporate property security, and commercial activity security, strict data collection and warning lines should be established to prohibit the collection and analysis of such data; Other sensitive data shall be classified and protected according to different subjects, contents and sources, as well as differences of data value and sensitivity. The collection and analysis of data shall be restricted according to the level of data protection. The algorithm model used by the regulatory system should protect the data accordingly, especially in the selection of the identification indicators for the algorithm model. Strictly abiding by the warning line, and avoiding the selection of sensitive data types involving personal privacy, corporate trade secrets, intellectual property and other important

²⁹ Li Cheng, 'Legal Governance of Artificial Intelligence Discrimination', 02*Chinese Law*, 2021, at 143.

³⁰ Jiang Dongxing and GAO Ruonan, Wang Haoyu, 'Research on Big Data Governance Scheme of Securities and Futures Industry Regulation', 03*Big DATA*5, 2019, at 32.

³¹ Zhang Xiaoshan, 'Legal Regulation of Discrimination in Artificial Intelligence Algorithms: European and American Experience and China's Path', 06*Journal of East China University of Science and Technology (Social Science Edition)*34, 2019, at 65.

information are all taken as important indicators.³² If the use of some sensitive data is essential for the regulatory algorithm model, such as interpersonal relationship data of insiders, undisclosed data of listed companies, desensitization algorithm can be developed to meet the regulatory needs.

2. Restrictions on data application scenarios of the algorithm

Massive database and machine learning algorithm are the central structure of securities Big Data platform. In the development and use of algorithm model, massive database is an irreplaceable factor. Even so, in some application scenarios involving data, it is necessary to set some restrictions on the development and use of algorithms to ensure the data security of securities market. Firstly, in the algorithm development scenarios, the regulatory department entrusts a third-party company to develop the algorithm model of insider trading regulation, and provides real securities data to exercise and test the algorithm. However, in order to protect the data security, the data cannot be directly handed over to the third-party company for independent use. In stead, the data can be used under the control of the regulatory department, or be desensitized before providing it to the research and development party. Secondly, in the data portrait scenarios, it is necessary to standardize the algorithm identification index set by the development party. Not only to prohibit the use of sensitive data involving personal privacy and company secrets as the algorithm model identification index, but also to match the corresponding desensitization algorithm according to the type and level of sensitive data. Thirdly, in the transaction monitoring scenarios, the identification and selection of behavioral data of financial investors and listed companies should strictly abide by the data rights warning line, so as not to infringe the data, personal privacy, property and other rights of financial investors, listed companies and other stakeholders.

³² Wu Jiaojun, Guo Waner, 'Legalized Governance of Algorithm Black Box in the Era of Artificial Intelligence', *01 Science and Law (Chinese and English)*, 2021, at 25.

References

1. Mayer-Schönberger, V., & Cukier, K., 'Big data: A revolution that will transform how we live, work, and think', Houghton Mifflin Harcourt(2013),at 130-135.
2. Chinese Securities Regulatory Commission, 'China Securities Regulatory Commission Officially Issues and Implements the Overall Construction Plan of Regulatory Technology', August 31, 2018, http://www.csrc.gov.cn/pub/newsite/zjhxwfb/xwdd/201808/t20180831_343433.html (visited on April 1, 2023)
3. Shanghai Stock Exchange, '2014 Self-regulatory Work Report of Shanghai Stock Exchange', 8 February 2015, <http://www.sse.com.cn/disclosure/credibility/whitepaper/c/4118097.pdf> (visited on April 1, 2023)
4. Guo Yanfang. 'Application and System Improvement of regulatory Technology -- Based on the Perspective of Information Manipulation' ,03Southwest Finance 2020,at 51.
5. ZHANG Xin,'The Framework, Features and Legalizing Path of Algorithmic Regulation: A Discussion on the Personal Information Protection Law (Draft)',14(01)Business and Economic Law Review2021,at 30-35.
6. Wen qiong, Smart regulation builds a strong defensive line for the stock exchange, Economic Daily News,February 2018,http://www.ce.cn/xwzx/gnsz/gdxw/201802/10/t20180210_28143113.shtml(visited on April 2, 2023)
7. 'Empowering Supervision with Technology -- The Intelligent Supervision System for Enterprise Portraits independently developed by Shenzhen Stock Exchange went online', Securities Times Website, October 2018,<http://money.people.com.cn/n1/2019/0802/c42877-31271375.html>(visited on April 2, 2023)
8. Zheng Zhihang, 'Ethical Crisis and Legal Regulation of Artificial Intelligence Algorithm', 39(01)Legal Science (Journal of Northwest University of Political Science and Law)2021,at 16.
9. Liao chang, 'Big Data analysis of social networks in capital markets: a regulatory application perspective',38(01)Trading technology frontier2020,at 74.
10. Henry Brooke. 'Algorithms, Artificial Intelligence and The Law', November 12, 2019, <https://www.supremecourt.uk/docs/speech-191112.pdf>
11. Zhang Song, 'Exploration on the application of Artificial Intelligence in technology financial regulation', 04China Credit Card2018, at 52.
12. Zheng Ge, 'The Law of Algorithm and the Algorithm of Law', 02Chinese Law Review 2018, at 75.
13. Cui Jingzi, 'The Crisis and Countermeasures of Equality Protection under the Challenge of Algorithm Discrimination', 37(03)Legal Science (Journal of Northwest University of Political Science and Law)2019, at 39.
14. Cui Jingzi, 'The Crisis and Countermeasures of Equality Protection under the Challenge of Algorithm Discrimination', 37(03)Legal Science (Journal of Northwest University of Political Science and Law)2019, at 40.
15. Yuan Kang, 'Institutional Response to the Black Box of Financial Technology Algorithms under the Concept of Social Supervision', 01Journal of University of Science and Technology of China (Social Sciences Edition)34, 2020, at 103.
16. Wang Huaiyong, 'Algorithmic Risk and Legal Regulation of Financial Technology' , 01Review of Political Science and Law, 2021,at 110.

17. Zhang Xiaoshan, 'Legal Regulation of Discrimination in Artificial Intelligence Algorithms: European and American Experience and China's Path', *Journal of East China University of Science and Technology (Social Science Edition)*34, 2019, at 67.
18. Brenner S W, Clarke L L, 'Fourth Amendment Protection for Shared Privacy Rights in Stored Transactional Data', *14JL & Pol'y*, 2006, at 213-214.
19. Pedro Domingos, 'The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World', basic books, 2015, at 101-110, 232-258.
20. Decision trees are a machine learning approach. It is a tree structure in which each internal node represents a judgment on an attribute, each branch represents the output of a judgment result, and finally each leaf node represents a classification result. The decision tree analyzes the data and outputs accurate predictions.
21. Zheng ge, 'Algorithmic law and algorithmic law', *02Chinese Law Review*, 2018, at 82.
22. Zheng Zhihang, 'Ethical Crisis and Legal Regulation of Artificial Intelligence Algorithm', *01Legal Science (Journal of Northwest University of Political Science and Law)*39, 2021, at 20.
23. Wei Yuanshan, 'The Loss and Regression of Algorithm Transparency: Functional Localization and implementation Path', *01Northern Law*15, 2021, at 152.
24. Wang Qinghua, 'Multiple Dimensions of Algorithm Transparency and Algorithm Accountability', *06Comparative Research*, 2020, at 169.
25. Lou Xianying, Fu Guangrui, et al, 'Legal Issues in Digital Transformation -- Starting from the Source Code of Software', KWM Research Institute, January 8, 2021, <https://www.infoobs.com/article/20210108/44652.html>
26. Joshua New and Daniel Castro, 'Algorithmic Accountability', Center for Data Innovation, 8 May 2018.
27. Feng Changyang, 'Research on the Open Code Policy System and Content of Major Countries and International Organizations', *06Library*, 2019, at 46.
28. The target code of the black box focuses only on the decision-making results of the algorithm, not the decision-making process.
29. Li Cheng, 'Legal Governance of Artificial Intelligence Discrimination', *02Chinese Law*, 2021, at 143.
30. Jiang Dongxing and GAO Ruonan, Wang Haoyu, 'Research on Big Data Governance Scheme of Securities and Futures Industry Regulation', *03Big DATA*5, 2019, at 32.
31. Zhang Xiaoshan, 'Legal Regulation of Discrimination in Artificial Intelligence Algorithms: European and American Experience and China's Path', *06Journal of East China University of Science and Technology (Social Science Edition)*34, 2019, at 65.
32. Wu Jiaojun, Guo Waner, 'Legalized Governance of Algorithm Black Box in the Era of Artificial Intelligence', *01Science and Law (Chinese and English)*, 2021, at 25.