

Review Article

On the Transparency of Artificial Intelligence System

Yanyong Du

School of Marxism, Shanghai Jiaotong University, Shanghai 200240, China

ABSTRACT

In order to improve the effectiveness of the management of artificial intelligence system, there is a growing demand for improving the transparency of artificial intelligence system from all parts of society. Improving the transparency of artificial intelligence system is conducive to relevant personnel better assuming their responsibilities and protecting the public's right to know. Therefore, the principle of transparency appears most frequently in all kinds of ethical principles and ethical guidelines of artificial intelligence, but there are some differences in the definition of its connotation by different subjects. The transparency of artificial intelligence system is reflected in many aspects like algorithm interpretation, data transparency and function transparency. We need to fully understand the limit of artificial intelligence transparency from the perspective of the characteristics of intelligence, the current situation of artificial intelligence technology and the feasibility of technical governance. For the construction path of artificial intelligence system transparency, there are many ways, such as technical approach, ethical and legal regulation and cultural approach.

Keywords: Artificial Intelligence; Transparency; Limits; Ethical Principles

ARTICLE INFO

Received: Dec 6, 2021
Accepted: Jan 10, 2022
Available online: Jan 17, 2022

*CORRESPONDING AUTHOR

Yanyong Du
yanyong817@sina.com;

CITATION

Du Y. On the transparency of artificial intelligence system. Journal of Autonomous Intelligence 2022; 5(1): 13-22. doi: 10.32629/jai.v5i1.486

COPYRIGHT

Copyright © 2022 by author(s) and Frontier Scientific Publishing. This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).
<https://creativecommons.org/licenses/by-nc/4.0/>

1. Introduction

With the rapid development of AI technology and industry, AI products are more and more integrated into people's daily work and life. Artificial intelligence products not only bring great convenience to mankind, but also cause a series of social and ethical problems, which makes both academia and the public generally concerned about how to reasonably standardize the R & D and application of artificial intelligence. In the relevant literature, the requirements and calls for "transparency" of artificial intelligence system can be seen everywhere, but there are different opinions on the theoretical connotation and implementation approach of this concept. On the basis of synthesizing the existing research results, this paper first briefly discusses the importance of transparency of artificial intelligence system in theory, and expounds the connotation of the concept of transparency. Then it focuses on the limitations of transparency requirements of artificial intelligence system. Finally, it puts forward some basic methods to construct transparency, so as to form a more comprehensive and objective understanding of the transparency of artificial intelligence system, and to attract more attention and make in-depth discussion.

2. Why is Transparency Important

Transparency is a key factor for the success of modern society, whether in political, business and professional fields, or in social, cultural and religious life^[1]. It is generally believed that transparency can

promote individual and collective behavior to strive to achieve meaningful goals, such as better accountability of public officials, more effective business behavior, etc., which can benefit the whole society^[2]. Generally speaking, in the era of intelligence, the importance of transparency of artificial intelligence system is reflected at least in the following aspects.

First, openness and transparency is the basic means and important goal of public management and social governance. With the development and progress of society, the transparency of all aspects of society is gradually increasing as a whole. From the perspective of government management, the government behavior in agricultural society is mysterious. The government behavior in industrial society is gradually transparent. In post-industrial society, the transparent administration of the government will evolve into the role consciousness of the government^[3]. In recent years, governance theory has been highly valued by the academic community, and transparency is usually regarded as the basic standard and requirement of good governance. Yu Keping believes that good governance has five basic elements: Legitimacy, transparency, responsibility, the rule of law and response. Transparency refers to the openness of political information, and the higher the degree of transparency, the higher the degree of good governance^[4]. There is no doubt that transparency has been generally valued in all areas of social life. For example, the principle of transparency is a basic principle in the field of news communication, and has been paid more and more attention by the press. It requires journalists to tell the audience what they know and don't know as truthful and open as possible, so as to make the audience accept the news report and show the journalists' respect for the audience^[5].

Second, transparency is closely related to responsibility. In recent decades, the concept of responsibility ethics has had a great impact on the academia and aroused the resonance of the general public. The ethics of responsibility appropriately reflects the spiritual needs and temperament of contemporary society in the face of the great challenges of the technological era, and it adapts to the

spirit of the times^[6]. Transparency can promote stakeholders to obtain corresponding information, so as to achieve mutual supervision and clarify their respective responsibilities. Many scholars believe that from the perspective of enterprise management and organization theory, the more transparent an organization or enterprise is, the better it can assume responsibility. In other words, transparency strengthens the responsibility of the enterprise and makes a close relationship between transparency and responsibility. Of course, transparency cannot be directly connected with responsibility causally. People often refer to the famous saying of US Justice Louis Brandeis that "Sunshine is the best disinfectant", which is essentially an intuitive expectation. In fact, the relationship between transparency and responsibility is not so simple, and transparency is not easy to be directly and causally linked with responsibility. Moreover, as explained in the third part below, the function and realization of transparency are limited, but we can't deny the positive role of transparency. To a certain extent, a certain degree of transparency can indeed enable management departments and managers at all levels to better clarify and assume responsibilities, and accept the supervision from all parties.

Third, transparency and the right to know. Academia generally believes that transparency is closely related to the public's right to know. From the perspective of law, the right to know refers to "Citizens have the right to know all information held and kept by the government and related to the exercise of their power, unless there are exceptions in the law"^[7]. Joseph Stiglitz, the winner of the Nobel Prize in Economics, believes that with transparency and openness becoming a hot topic among the public, "information" has also become the focus of economics and public management. Information plays an important role in free choice. Substantive participation in the democratic process requires participants to be informed of sufficient relevant information. Although there are certain limits to information disclosure, the improvement of government transparency has its inherent value. The public has the basic right to know, and the government should further improve transparency^[8]. In other

words, improving transparency can enable the public to obtain more information in a reasonable and legal way, which is the basic means to protect their right to know.

Fourth, people's general expectations and requirements for the transparency of artificial intelligence system. As we all know, artificial intelligence systems are more widely used. From smart phones to driverless vehicles, from personal shopping recommendations to government major decisions, from job screening to intelligent medical treatment, from daily life to modern battlefield, artificial intelligence products are everywhere, and human dependence on them is deepening day by day. While enjoying the convenience brought by science and technology, people are also generally worried about the abuse or misuse of artificial intelligence system, because some negative effects have been produced, and there is a great possibility of other negative effects. Therefore, in order to avoid the possible negative effects of artificial intelligence, people generally require it to have a certain degree of transparency. It is easy to find that scholars who discuss the ethics and governance of artificial intelligence generally pay attention to the transparency of artificial intelligence system. Jurist Thomas Wischmeyer believes that the fuzziness of artificial intelligence itself has become an important political issue. Opening the black box is essential for checking user privacy protection, detecting bias and discrimination, and preventing all kinds of potential harm^[9]. Nicholas Diakopoulos stressed that the transparency of artificial intelligence should at least be used to increase the availability of information, so that people can understand who is using information, what information designers must consider to spread, and how to spread among different audiences^[10].

In this historical context, since 2016, many international organizations, government departments and academic groups have issued AI ethical principles and guidelines. Anna Jobin and other people collected and compared 84 documents written in English, German, French and other different languages with ethical guidelines and ethical principles included. They ranked the relevant ethical principles according to the frequency of occurrence. It was

found that 73 documents contained content related to "transparency", which appeared most frequently in various ethical principles^[11].

3. Transparency as an Ethical Principle (Guide)

Although transparency as an ethical principle and ethical guide has been highly valued, there are obvious differences in the definition of transparency by different organizations and groups. In general, various documents mainly define the transparency related to artificial intelligence from the following perspectives.

First, it emphasizes the interpretability of the decision-making process and mechanism (mainly algorithm) of artificial intelligence system, and the predictability and explicability of decision-making consequences, which is the most common definition of the transparency of artificial intelligence system. First of all, pay attention to the intelligibility and interpretability of artificial intelligence decision-making process. In December 2017, the Institute of Electrical and Electronics Engineers (IEEE) released the report on "Ethical Design". Among the five ethical principles, the fourth is transparency, which points out that transparency means that we can find out how and why the artificial intelligence system makes a specific decision. At the same time, IEEE emphasizes that transparency has different connotations for different stakeholders. For users, transparency means that it provides users with a simple way to understand what the AI system is doing and why. Therefore, users can understand the capabilities and limitations of the artificial intelligence system they use or interact with, so as to reduce risks and harm. Secondly, it emphasizes the interpretability of artificial intelligence decision-making and its impact. Asiloma AI principle proposed in early 2017 interprets transparency from two aspects: Failure transparency and judicial transparency. Failure transparency means that if an AI system causes damage, there should be a way to find out the cause. Judicial transparency means that in judicial decisions, whenever the system is independently developed, a convincing explanation should be provided and audited by a competent

person.

Second, it emphasizes the disclosure of the identity and function of the artificial intelligence system. In September 2010, the Engineering and Physical Sciences Research Council of the UK released the *Principles of Robotics*, emphasizing the artificial characteristics of robots, believing that we can not deceive vulnerable groups with robots, and the mechanical characteristics of robots should be transparent. In other words, anyone who owns or interacts with robots should be able to realize what is real and what is artificial. In July 2018, the ethical principle of artificial intelligence proposed by Deutsche Telekom defined “transparency” as “When it is artificial intelligence that receives customers, we will never hide it. Moreover, we are transparent about how we apply customer data.”

Third, it emphasizes the comprehensiveness, openness and timely update of data. In May 2018, IBM released the *IBM Trust and Transparency Principles*, which pointed out that in order to make the public trust artificial intelligence, it must be transparent. The technology company must know who is training the artificial intelligence system, what data is used in the training process, and the most important is which data is used for algorithm recommendation. If we use artificial intelligence to assist in making major decisions, it must be interpretable. In June 2018, the Ministry of Digital, Culture, Media and Sports of the British government issued the *Data Ethics Framework*, proposing seven data ethics principles, of which the sixth is the principle of transparency and responsibility, requiring individuals and institutions involved in public domain data to be open and transparent about tools, data, algorithms, user needs and other information (except for special reasons such as fraud prevention and anti-terrorism), and the relevant data should be accurate and up-to-date.

Fourth, it emphasizes transparency through means such as supervision and evaluation. Ten suggestions are put forward in the 2017 *AI Now* report released by New York University. The first one emphasizes that the “black box” artificial intelligence and algorithm system cannot be applied in high-risk fields such as criminal justice, medical

health, social welfare and education. In other words, pre-trained models that have not been reviewed and verified cannot be used. The artificial intelligence system should be licensed by a third-party organization, and its application should at least be subject to public supervision, testing, evaluation, etc. In May 2018, a coalition of human rights and technology groups in Toronto issued the *Toronto Declaration*, calling on governments and technology companies to avoid discrimination and protect human rights in machine learning systems, emphasizing maximum transparency to the public in the application of machine learning systems, and pointing out the need to strengthen supervision and strengthen independent institutions to analyze and evaluate them.

We note that the relevant reports issued in recent two years usually emphasize the above aspects. In April 2019, China’s national AI standardization group released the *AI Ethical Risk Analysis Report*, which stressed that the development of AI in China should follow the principle of human fundamental interests and responsibility, and emphasized the principle of transparency in the principle of responsibility. It pointed out that the realization of this principle depends on the interpretability, verifiability and predictability of AI algorithms, and emphasized that the transparency of data sources is also very important.

In addition, the definition of transparency of artificial intelligence system in domestic and foreign business and academic circles is basically consistent with the connotation of transparency as an ethical principle. For example, legal researcher Guo Rui believes that the principle of transparency requires human beings to know how and why AI makes specific decisions. The realization of the principle of transparency depends on the interpretability, verifiability and predictability of AI algorithms^[12]. Shen Xiangyang, global executive vice president of Microsoft, and Brad Smith, President of Microsoft, believe that the transparency of artificial intelligence system lies in its easy understanding. Microsoft tries to replace those methods that are too complex and difficult with practical norms, various methods that is easily understandable, algorithms or

models, so as to effectively improve the transparency of artificial intelligence system^[13]. Computer scientists at Bath University of the UK believe that transparency refers to the mechanism of artificial intelligence system reporting its behavior, reliability, perception and goal, which can be obtained at any time^[14]. It can be seen that although scholars in different fields have different emphases on the connotation of the concept of transparency, but the coincidence degree is very high, which shows the importance of artificial intelligence transparency.

4. Limits of Transparency of Artificial Intelligence System

Although the transparency of artificial intelligence system is of undoubted importance, it is equally important that we must see its limitation and not exaggerate the function and necessity of transparency.

First, the contradiction between transparency and shielding: the degree of transparency is limited. In the information society, people think that getting more information can increase the transparency of society and realize the rational management of social problems. This is actually an illusion. In fact, the more information we get, the more we obscure the objects we want to know, resulting in less understanding and trust and making the society as opaque as before. Some scholars believe that the simple pursuit of information disclosure may even produce a kind of “the tyranny of light”, which leads to extreme skepticism and enhances uncertainty^[15]. Similarly, Strason calls the sheer pursuit of transparency “the tyranny of transparency”. He believes that people have different interest demands in social life. The requirement for transparency seems reasonable and moral, but it is obviously arbitrary, that is, it is always right to make invisible visible^[16]. From the perspective of legal principles, as a generally mandatory legal principle, the transparency principle of artificial intelligence algorithm may conflict with legal interests such as national security, social order and private subject rights, and does not have the universal feasibility to match the basic legal principles^[17]. In other words, data and information related to personal privacy, trade secrets, national security,

etc. should be an exception to the principle of transparency. Therefore, the transparency of artificial intelligence system should have different performance in different fields, and can not be generalized.

Second, uncertainty and autonomy: the ontological shackles of intelligent features on transparency. People share different opinions about the topic of essential characteristics of intelligence, but the characteristics of intelligence are at least reflected in autonomy, interactivity and emergence. Therefore, scientists and technicians attach great importance to the independent decision-making and judgment of artificial intelligence system under certain conditions, and emphasize the synergy between artificial intelligence and human and environment. Therefore, the autonomy and interaction with the environment of the artificial intelligence system determine that the decisions made by the artificial intelligence system do not have high precision and predictability. The related concept is the so-called Polanyi’s Paradox put forward by Michael Polanyi in 1966: “We know more than we can say... Drivers’ skills cannot be completely replaced by driving theory^[18].” Will Knight also believes that at present, scientists and technicians have made a lot of efforts to explain the interpretability of artificial intelligence system, but just as it is impossible to explain human behavior in detail, it may be impossible to explain all behaviors of artificial intelligence. Even if some people think that he has given a reasonable explanation for the artificial intelligence system, it may not be complete. Perhaps only part of the essence of intelligence can be explained rationally, and some parts belong to instinct and subconscious^[19].

Third, the technology black box that cannot be opened: the barrier of the current technology situation. Experts in the field of artificial intelligence generally recognize that the algorithm system of artificial intelligence is indeed a black box to some extent. Zhou Zhihua believes that the typical deep learning model is a deep neural network. The model with more parameters and larger capacity can complete more complex learning tasks. The simple way to improve the capacity is to increase the number of hidden layers. Moreover, from the perspective of

increasing the complexity of the model, increasing the number of hidden layers is more effective than increasing the number of hidden neurons, so the number of hidden layers of the deep neural network increases gradually. Neural network is a “black box model” which is difficult to explain, although some work has tried to improve the interpretability of neural network^[20]. It should be emphasized that even if the AI system is a black box, it will not hinder people from applying AI products. At present, smart phones and personal computers have entered thousands of households. Most users do not understand their working principles, but this does not affect people’s application of them.

Fourth, focus on the process or result: the feasibility of transparent regulation. It is generally believed that the clearer the understanding of AI decision-making process and technical characteristics, the higher its transparency and the more feasible its regulation. For example, Ryan Calo believes that we should treat robots and artificial intelligence from the perspective of law and policy on the basis of recognizing their technical characteristics. He stressed that the clearer we understand how technology affects society, the more favorable we can be in the process of integrating technology into society^[21]. On the contrary, Jack Balkin believes that it is useless to discuss the legal regulation based on the key characteristics of robots and artificial intelligence. The reasons are that the technical characteristics depend on how people apply technology in life and social relations, and people’s application of technology will change constantly, and new ways will be found to use technology for good or evil. He suggested taking a more pragmatic stand and focusing on some important specific problems caused by the application of artificial intelligence and robots^[22]. It can be seen from the debate among jurists that different scholars have different views on the key points of transparency regulation of artificial intelligence system. Carlo emphasizes the intelligibility of AI decision-making process and technical characteristics, while Balkin emphasizes results and functions. From the perspective of operability, Balkin's stand is more reasonable, which also reminds us that we should pay attention to under-

standing transparency from the perspective of system results and functions.

Fifth, trust or doubt: the public’s awareness of artificial intelligence and the necessity of transparency. Although evidence-based algorithms can usually make more accurate predictions than humans, sometimes when people choose between humans and algorithms, they still choose human predictions. Berkeley Dietvorst and others call this phenomenon “algorithm aversion”. Their research shows that when people find that the algorithm has made the same mistakes as humans, people are more likely to lose confidence in the algorithm than human foretellers^[23]. The research results of Dietvorst and others mean that people are usually more harsh on technical products such as artificial intelligence, and more tolerant to themselves and their peers. However, for the general public of non-professionals, people prefer to accept suggestions from algorithms rather than others. Jennifer Logg and others call this effect “algorithm appreciation”. Logue and others believe that people’s attitude towards the algorithm is not only related to the accuracy of the algorithm, but also closely related to the field of application. For problems related to personal preferences, people are indeed more willing to accept the suggestions of close friends. But in areas with clear standards for accuracy, such as investment decision-making, sports game prediction, etc., people are more willing to accept the suggestions given by algorithms^[24]. People require the artificial intelligence system to improve its transparency, fundamental purpose of which is to enhance its credibility. However, in fact, even if the transparency of the decision-making process of the artificial intelligence system is not high, as long as its prediction is accurate and its function is reliable, the public is still willing to believe and accept them under some conditions. Of course, we are not here to deny the importance of transparency. We are trying to emphasize that from the perspective of public cognition, it is better to focus on the functional stability, accuracy and reliability of artificial intelligence system than on the interpretability of technology. There is no doubt that the progress in these aspects will greatly strengthen the public’s trust in artificial intelligence

system.

5. Construction Path of Artificial Intelligence System Transparency

It can be seen from the foregoing that the transparency of artificial intelligence system involves multiple interrelated levels, and there need to be construction approaches available. We need to comprehensively build the transparency of artificial intelligence system based on reality, necessity and feasibility for different subjects.

First, internal technology: an all-round technological progress from algorithm decision-making to result analysis. In recent years, academic circles have made great progress in the interpretation technology of artificial intelligence system. For example, Chinese scholars try to understand deep learning from the perspective of geometry and put forward AE-OT (Autoencoder-Optimal Transportation) model, which can improve the preciseness and transparency of the theory and has attracted extensive attention in the academia^[25]. As another example, back propagation technology starts with the output results of deep neural network and projects the predictions of lower levels respectively until the input layer. This technology has achieved preliminary results in image recognition and other fields^[26].

Efforts to increase the transparency of artificial intelligence system from the perspective of technical details should be encouraged, but other research approaches also deserve our attention. For example, Lilian Edwards and others advocate a subject centered interpretation approach, that is, pay attention to the preset and interaction of interaction between subjects in some specific fields, and explain AI through external interaction model rather than from within the system^[27]. Other scholars put forward the idea of interpretation of opposite facts to improve the interpretability of the algorithm, that is, to tell people what different results the algorithm will give in the case of opposite facts. This interpretation approach does not try to open the black box of the algorithm, but provides easily understandable and practical information to make the decision-making of the artificial intelligence system understandable, so as to

provide a clear goal and direction for future efforts^[28]. Generally speaking, although the interpretability of artificial intelligence system has made great progress, it is still an emerging field, and various research approaches vary greatly. There is still a long way to go to improve the interpretability of artificial intelligence system.

Second, technical specifications: setting standards and strengthening tests. Technical standards have important basic, leading and normative functions for R & D of scientific and technological products, which can not only promote the healthy and rapid development of artificial intelligence industry, but also play a positive role in promoting the transparency of artificial intelligence system. In the implementation suggestions on the principle of transparency in the report on ethical design (2nd Edition) released by IEEE in 2019, it is pointed out that new standards should be developed to express the measurable and detectable level of transparency, so that the artificial intelligence system can be objectively evaluated. The IEEE standards working team has specially set up P7001 project, which aims to set specific technical standards for the transparency of autonomous systems and clarify the level of transparency. Our government also attaches great importance to the standardization of artificial intelligence. In 2017, the *Development Plan of New Generation Artificial Intelligence* issued by the State Council stressed the need to “strengthen the research on artificial intelligence standard framework system”. In August 2020, the national standardization management committee formulated the guide for the construction of national new generation artificial intelligence standard system, and put forward the framework and construction contents of China’s artificial intelligence standard system, etc.

At the same time, artificial intelligence system testing is an important means to ensure the reliability and stability of artificial intelligence functions. Therefore, we must strengthen the testing work and improve the construction of testing standard system. China has formed a series of national standards for software testing, which has played an important role in promoting the industrial application of standards and guiding the development of the software indus-

try^[29]. We should establish and improve relevant test standards for artificial intelligence system, so as to promote the transparency and reliability of artificial intelligence system.

Third, social constraints: ethical and legal regulation. As discussed in the second part of this paper, transparency has been widely recognized as an ethical principle. Next, we need to explore how to translate the ethical principle into different behavioral agents^[30]. We believe that the most important approach of ethical governance may be to clarify the responsibilities of relevant subjects and improve the corresponding responsibility mechanism. For example, for developers, they should first disclose the objectives of artificial intelligence system R & D. The goal determines the function. If the artificial intelligence system has clear goals and stable and reliable functions, it is obviously easy to get public recognition and support. Some scholars advocate regulating algorithm decision-making through algorithm “interpretation right”. As the algorithm controller allocates social resources based on massive data operation and gradually grasps a new algorithm power, we need to enhance the transparency of algorithm decision-making, so as to improve its comprehensibility and accountability. The power of algorithm interpretation is to give individuals the right to interpret algorithm decision-making, so as to realize the appropriate restriction on the abuse of algorithm power^[31]. The research results of the legal circle on the legal regulation of artificial intelligence deserve the high attention of the ethical circle. At the same time, some important ways to ensure transparency should be clearly regulated with legal means. For example, in 2018, the EU issued the *General Data Protection Regulation*, which emphasizes the transparency of data collection and use, and focuses on protecting the privacy and the right to know about data subjects.

Fourth, environmental construction: construct the social culture in the era of artificial intelligence. When Karl Popper talks about the source of human knowledge, we believe that the most important source of knowledge is tradition, because most of our knowledge is obtained through examples, others’ descriptions, reading, learning how to criticize,

learning how to accommodate and accept criticism, and how to respect truth^[32]. It can be seen that cultural tradition and social atmosphere play an important role in the rational cognition of the transparency of artificial intelligence system of people from all walks of life. Therefore, we need to make various efforts to jointly construct the social culture of the intelligent era.

For mass media, it is necessary to guide the public to fully understand the necessity and limitations of transparency of artificial intelligence system. Although people usually hope to ensure the fairness of artificial intelligence system as much as possible through artificial intelligence transparency and other means, any technology will have defects and deficiencies, and there is no absolute fair and transparent artificial intelligence. The public needs to pay more attention to the functional reliability and security of artificial intelligence system, reasonably protect personal privacy, and actively advocate the right to know and other legitimate rights and interests. It should be emphasized that scientific and technological workers should actively assume the responsibility of creating a cultural atmosphere in the intelligent era. Scientific and technological workers are obliged to explain and popularize all kinds of knowledge of artificial intelligence system to the public. The efforts of the science and technology community to realize the transparency of artificial intelligence not only help to eliminate the public’s doubts and fears about artificial intelligence, but also help to promote the public’s trust in it, so that they would be more willing to accept artificial intelligence products.

6. Conclusion: Build a Trustworthy Artificial Intelligence System

Human’s trust in artificial intelligence system is the basic prerequisite for maintaining social order in the era of intelligence. Ensuring and realizing the transparency of artificial intelligence system to a certain extent is an important means to realize the public’s sense of trust. Therefore, transparency is the most emphasized basic principle in the ethical principles of artificial intelligence, and all sectors of society have high expectations for the transparency

of artificial intelligence system. However, the transparency of artificial intelligence system is multifaceted, from the transparency of data to the interpretability of algorithm, from the transparency of function to the user's right to know and so on. We need to build trustworthy, humanized and warm artificial intelligence products through transparency and other means. At the same time, we should recognize the limitation of the transparency of artificial intelligence system, have a rational understanding of its function, recognize its necessity and do not exaggerate its role. Transparency is the basic means to achieve harmony between people and intelligent products, and it is not the ultimate goal we pursue. Generally speaking, the research on the transparency of artificial intelligence system is still in the pioneering stage. We look forward to richer research results which effectively promote relevant research.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Oliver R. What is transparency. New York: McGraw-Hill; 2004.
2. Hansen H, Christensen L, Flyverbom M. Introduction: Logics of transparency in late modernity: Paradoxes, mediation and governance. *European Journal of Social Theory* 2015; 18(2): 117–131. doi: 10.1177/1368431014555254.
3. Liu ZY, Lin L. Transparent government: The history and logic view on the transformation of a government pattern (in Chinese). *Journal of Sichuan University (Philosophy and Social Sciences edition)*, 2009; (1): 21–28.
4. Yu KP. Governance and good governance (in Chinese). Beijing: Social Sciences Academic Press; 2000. p. 9–10.
5. Kovach B, Rosenstiel T. The elements of journalism: What news people should know and the public should expect. Beijing: Peking University; 2014. p. 110–115.
6. Gan SP. The frontiers of applied ethics (in Chinese). Guiyang: Guizhou University; 2019. p. 86.
7. Zhang JS. The right to know and its guarantee (in Chinese). *China Legal Science* 2008; (4): 12.
8. Stiglitz J. On liberty, the right to know, and public discourse: The role of transparency in public life. *Global Law Review* 2002; (3): 263–273.
9. Wischmeyer T, Rademacher T. Regulating artificial intelligence. Switzerland: Springer, 2020. p. 76.
10. Dubber MD, Pasquale F, Das S. The Oxford handbook of ethics of AI. Oxford: Oxford University Press; 2020. p. 200.
11. Jobin A, Ienca M, Vayena E. The global landscape of AI ethical guidelines. *Nature Machine Intelligence* 2019; (9): 389–399. doi: 10.1038/s42256-019-0088-2.
12. Guo R. The ethics and governance of artificial intelligence (in Chinese). Beijing: Lawpress; 2020. p. 38.
13. Shum H, Smith B. The future computed (in Chinese). Beijing: Peking University; 2018. p. 39.
14. Theodorou A, Wortham R, Bryson J. Designing and implementing transparency for real time inspection of autonomous robots. *Connection Science* 2017; 29(3): 230–241. doi: 10.1080/09540091.2017.1310182.
15. Tsoukas H. The tyranny of light. *Future* 1997; (9): 827–843.
16. Strathern M. The tyranny of transparency. *British Educational Research Journal* 2000; 26(3): 309–321.
17. Shen WW. The myth of the algorithmic transparency principle. *Global Law Review* 2019; 40(6): 20–39.
18. Vardi M. The moral imperative of artificial intelligence. *Communications of the ACM* 2016; 59(5): 5.
19. Knight W. The dark secret at the heart of AI. *MIT Technology Review* 2017; (3): 55–63.
20. Zhou ZH. Machine learning (in Chinese). Beijing: Tsinghua University; 2016. p. 113–115.
21. Calo R. Robotics and the lessons of cyberlaw. *California Law Review* 2015; 103(3): 513–563. doi: 10.2139/ssrn.2402972.
22. Balkin J. The path of robotics law. *California Law Review Circuit* 2015; (2): 45–60.
23. Dietvorst B, Simmons J, Massey C. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 2015; 144(1): 114–126. doi: doi:

- 10.1037/xge0000033.
24. Logg J, Minson J, Moore D. Algorithm appreciation: People prefer algorithmic to human judgement. *Organizational Behavior and Human Decision Processes* 2019; 151(1): 90–103. doi: 10.1016/j.obhdp.2018.12.005.
 25. Lei N, An D, Guo Y, *et al.* A geometric understanding of deep learning. *Engineering* 2020; (3): 361–374. doi: 10.48550/arXiv.1805.10451.
 26. Montavon G, Samek W, Muller K. Methods of interpreting and understanding deep neural networks. *Digital Signal Processing* 2018; 73(1): 1–15. doi: 10.1016/j.dsp.2017.10.011.
 27. Edwards L, Veale M. Slave to the algorithm? Why a right to an explanation is probably not the remedy you are looking for. *Duke Law & Technology Review* 2017; (1): 18–84.
 28. Wachter S, Mittelstadt B, Russell C. Counterfactual explanations without opening the black box. *Harvard Journal of Law & Technology* 2018; (2): 841–887.
 29. Zhao Y, Hu Y, Gong JY, *et al.* Research on domestic standardization of software quality and software testing (in Chinese). *Standard Science* 2021; (4): 25–31.
 30. Wu H, Du YY. Artificial intelligence ethical governance: Translating principles into practices. *Studies in Dialectics of Nature* 2021; 37(4):49–54. doi:10.19484/j.cnki.1000-8934.2021.04.009.
 31. Xie ZS. Regulating algorithmic decision: Focusing on the right to explanation of algorithm. *Modern Law Science* (in Chinese). *Modern Law* 2020; 42(1): 179–193.
 32. Popper K. *Conjectures and refutations*. Hangzhou: China Academy of Art Press; 2003. p. 36.