Advances in Social Science, Education and Humanities Research, volume 385
3rd International Conference on Education, Economics and Management Research (ICEEMR 2019)

Research on the Role of Algorithm Transparency in Algorithm Accountability

Weimin Ouyang
Shanghai University of Political Science and Law
Shanghai, China
oywm@shupl.edu.cn

***Abstract*—Algorithm transparency as an algorithm accountability method has been widely praised by the community. Algorithm transparency is considered to be an effective way to crack the algorithm black box. The algorithm transparency refers to the disclosure of the source code of the algorithm system and the data used in the algorithm system. This paper first introduces the concept of algorithm and algorithm governance, and then critically analyzes the side effects of algorithm transparency, and finally discusses the dialectical relationship between algorithm transparency and algorithm trust, claims that establishing algorithm trust is the foundation of algorithm accountability, and points out that algorithm transparency can't rebuild the trust of algorithm. At the same time, it describes the basic methods to establish the trust of the algorithm.**

***Keywords*—algorithm transparency; algorithm accountability; algorithm trust; algorithm governance**

I. Introduction

With the advent of the third wave of artificial intelligence, artificial intelligence has penetrated into all aspects of society. More and more affairs, such as criminal justice, food safety, social services and transportation, are decided by algorithms. Algorithms, which are almost everywhere, have actually become an important power in today's society, which determines to a large extent what we see, what we hear, what we know to be true or false, and who we interact with [1]. As we all know, power needs to be monitored, but the power of algorithms is hard to monitor at the moment. Around the world, whether Facebook, Twitter, Google in the United States or Baidu, Alibaba and Tencent in China, almost all Internet companies regard algorithms as important corporate secrets and are unwilling to open the black for the outside world to understand. As commercial companies, it is understandable that these Internet giants protect their algorithms as core secrets. However, the public does not think so. Since the algorithm has affected our work and life so extensively and deeply, it can no longer be presented in the form of a black box on the basis of "trade secrets"[2]. The public should have the right to know and supervise the algorithm. Artificial intelligence is facing a strong call for fairness, morality and security. In response to public concerns, the New York City Council took the lead in passing the Algorithmic Accountability Act in December 2017 and set up a working group to study how to monitor and evaluate the algorithms used by the public sector. On May 25, 2018, the European Union formally implemented the General Data Protection Regulation (GDPR), which requires companies to explain all automatic decisions to consumers. Critics like Elon Musk, who has attracted much attention, call on policymakers to take more measures to regulate AI and relieve distrust of AI.

To this end, a response called algorithm transparency is to be proposed to require companies to disclose the source code and data used in their AI systems. This strategy has been widely praised by all sectors of society [3, 4]. On the surface, algorithm transparency is indeed an effective countermeasure to crack the algorithm black box. However, from the perspective of actual operation, algorithm transparency is not realistic and has a little practical effect, which often just brings some psychological sense of control. In the rest of this paper, we first introduce the concepts of algorithm and algorithm governance, and then comprehensively analyze the side effects of algorithm transparency. Finally, we discuss the dialectical relationship between algorithm transparency and algorithm trust.

II. Algorithmic Governance and Its Social Needs

An algorithm is a set of instructions designed to accomplish a specific task. Many organizations use algorithms to make decisions and allocate resources based on large data sets. Usually, any AI algorithm includes collecting data from one or more data sources, applying machine learning, statistical analysis and other techniques to discover the correlation between features and results, and using this technology to generate a previously unknown or not accurately described pattern [5].

Algorithms are attractive because they are considered to have the natural attributes of value neutrality and objective justice. Algorithms accept data and deliver results. However, the algorithm is not "neutral" in fact. The design, purpose and data usage of the algorithm are all the subjective choices of the design developers, and their subjective bias will naturally be embedded in the algorithm system. The validity and accuracy of data will also affect the accuracy of decision-making and prediction of the algorithm. It is precise because of design bias, data defects, and the "black box" characteristics of the algorithm that the algorithm bias, algorithm discrimination and algorithm manipulation often

Funded by China National Institute for SCO International Exchange and Judicial Cooperation.

occur. Under the background of the continuous development of artificial intelligence and big data, algorithms will have spawned unprecedented social new phenomena, such as algorithm politics, algorithm economy, algorithm war, algorithm news, algorithm ethics, etc. Among them, some of the consequences caused by the algorithm have caused alarm, such as a small number of "Internet water army" using software to spread a large number of false information, misleading public opinion, interference with social stability. The results caused by some algorithms are hard to be perceived in a short time. For example, the algorithm can acquire and adjust the price in time to block the price competition behavior and form price cheating and price monopoly.

This kind of algorithm problem is very covert and destructive, and it must be effectively treated. Algorithmic governance has become a subject that must be faced. Algorithm governance refers to the process in which the government, social organizations, enterprises, public institutions, communities, individuals and other subjects guide and regulate the design, development and application of algorithms through equal cooperation, dialogue, consultation, communication and other means, and ultimately realize the maximization of public interests [6]. The public's concern about algorithm governance includes but is not limited to the following:

- Is the algorithm politically incorrect? Does it involve racial discrimination or gender discrimination?
- Will algorithm optimization grant different privileges to different stakeholders?
- Does the algorithm have functions that should be disclosed but are not disclosed?
- Whether the algorithm is really executing according to its design objectives, and how to detect the results of the algorithm?
- Is the algorithm fair? How do you test the fairness of an algorithm?
- How does the algorithm interpret its results?
- Does the algorithm increase or decrease the organization's capability?
- Is algorithm governance detrimental to critical evaluation and feedback, thus leading to the reinforcement of result bias?
- Will the algorithm weaken the decision-making ability and participation consciousness of the organization and undermine the social recognition of the organization?

III. ALGORITHM TRANSPARENCY AND ITS SIDE EFFECTS

As the penetration of artificial intelligence to society is more and more extensive, algorithms play an increasingly important role in human society. Artificial intelligence is changing the original rules and practices of human society. While human beings give AI the power of choice and even decision-making, enjoying the convenience, they are worried about the negative consequences of algorithm bias, algorithm discrimination and algorithm invasion of privacy and social relations, and express the fear that they and the human world will be dominated and ruled by AI [7]. Human beings have an innate fear of the unknown. For human beings, algorithms are like "black boxes," with no outside knowledge of how the AI algorithm works between its input data and its output. For this reason, the concept of algorithm transparency has been proposed, which requires companies to disclose the models, source code and input data used by algorithms. The idea has now become a social consensus and is regarded as a perfect prescription for solving the black box problem of algorithms.

Algorithmic transparency as an accountability approach has gained broad support from governments to private companies. The idea is based on the simple idea that when people see something, they can naturally supervise it. The implicit assumption of this idea is that when one can see inside the algorithm system, one can gain trust in the algorithm by understanding the inside of the algorithm, and improve the trust in the algorithm by modifying or changing the algorithm in case of problems. The more you know about algorithms, the more you can understand them and trust them. Algorithm transparency is actually based on the claim that transparency of the source code of the algorithm and the data that the algorithm is processing is the most effective way to evaluate the fairness of the algorithm. However, this idea may be oversimplified by a lack of knowledge of algorithms. From the viewpoint of computer science, even if the algorithm is transparent and can be examined, its effect on the potential consequences will be insignificant. When the algorithm is very complex, "black box" not only means that it cannot be observed but also means that even if the algorithm explains to us, we cannot actually understand it without corresponding professional knowledge.

In this section, we will discuss the side effects of algorithmic transparency, pointing out that not only is transparency, not a good way to hold algorithms accountable, but it can also cause more problems.

A. Algorithm transparency leads to Privacy Disclosure

Algorithm transparency requires that the data processed by the algorithm should be disclosed. It is expected that data openness can be used to evaluate the data, discover and eliminate structural imbalance, embedded bias and injustice of the data, so as to ensure the data is "clean". However, it is clear that when the data set is open, privacy will be revealed, and full transparency will cause great harm.

In the world of big data, the social software you used to use is the most likely to know your family and your social relationship, the payment software you used to use is the most likely to know your financial situation and the platform you used to buy online is the one that knows your shopping habits and ability. If these data are combined, a complete and accurate data profile about you will be formed. Once such data is leaked, you will have no personal privacy and may even be used for illegal business transactions. Facebook admitted that Cambridge Analytica, a data analysis company that helped Trump win the US presidential election in 2016, had illegally obtained information from 50 million Facebook users. The information of tens of millions of Facebook users has been leaked, causing a global panic. If the Facebook system is transparent and all user data is open, the consequences would be even worse. Users would almost certainly give up using Facebook. The data transparency of the algorithm system will leak the privacy, damage the data security, destroy the trust of the algorithm, and then lead to the disaster of the algorithm system.

Data transparency must be cautious, and it is not advisable to be completely open to the public. Data should only be open to trusted authorities or organizations with certain supervision authority. Whether the public has the right to access the data set of algorithm system should be carefully considered. Of course, the items involving individuals should be open to that person.

B. Algorithm Transparency is Detrimental to the Competitiveness of Enterprises

So far, the general consensus of the society is to take an algorithm as the core competitiveness of the company, which is usually protected by patents or trade secrets, so as to maintain its competitiveness. Because the patent has a certain time limit, after the protection period, it must be disclosed with free charge, and there is no time limit for business secrets. In theory, it is indefinite. Most companies will protect the algorithm by trade secrets, so as to obtain and maintain the competitive advantage against competitors to the maximum extent. Algorithm transparency may be good for competitors, but it may not be good for the general public, regulators, etc. When too much information makes it impossible to identify useful information, transparency will obscure the truth. The algorithm itself is in the process of evolution, and the opened algorithm is often no longer the one in current use, which may have little effect on rebuilding trust in algorithms.

Intellectual property is an important mechanism to protect a company's core competitiveness and promote its development, but it is not the only mechanism. In contrast to intellectual property protection, there are many individuals and organizations that adopt the way of knowledge sharing, such as open-source software. On November 15, 2018, ArcSoft, the world's leading provider of visual artificial intelligence technology, announced the free open algorithm and free AI development platform, aiming to build an open ecological platform, form an ecosystem and grow together. The algorithm goes from empowering a single enterprise to empowering an entire ecosystem.

C. Algorithm Transparency is Easy to Game the System

It is not a wise choice to disclose algorithm details to the public. The disclosure of the details of the internal working mechanism of the algorithm may have unexpected consequences, which may enable some people or organizations to take advantage of the characteristics of the system to cheat, that is, to game the algorithm system. Algorithm transparency does not necessarily bring trust. Trust is two-way. Some developers don't want to open algorithms, not because of trade secrets, but because they don't want to be maliciously used by some people.

Paper checking is an important technical means to prevent paper plagiarism, which has been widely used. However, with the use of paper checking system, people found its loopholes, that is, the paper checking is only limited to the form of the text, not the meaning of the text. As a result, a variety of methods to reduce the repetition rate have emerged， such as replacing key words, changing sentence patterns, changing word order, changing synonyms, paraphrasing, deleting repeated parts. These methods are proposed in the case that the algorithm has not been disclosed, which has a great adverse impact on the paper checking system. If all the details of the algorithm are disclosed due to the algorithm transparency, the paper checking system will exist in name only.

D. Algorithm is Inherently Opaque

The algorithm is highly professional, and often contains the professional knowledge of an application field, which makes great difficulties for non-professionals to understand. Programming is a highly specialized technical skill, which is difficult for ordinary people to acquire and understand. This kind of algorithm opacity can be called the opacity caused by technical illiteracy.

The second kind of algorithm opacity is due to the technical limitations of the algorithm. Algorithms are getting bigger and more complex, often beyond the comprehension of professional programmers. In this case, it is useless to make the algorithm transparent for public scrutiny.

The third kind of algorithm opacity is due to the fact that some machine learning algorithms are essentially opaque. For example, the learning results of deep learning algorithms cannot be explained, and the parameters of neural networks can only be explained by weight in mathematics. However, deep learning systems often have millions or even billions of parameters, which makes it difficult for developers to explain a complex algorithm in an interpretable way to mark a complex neural network is used to label, not to mention to explain its results. Therefore, deep learning algorithm is essentially a "black box algorithm", and the output results cannot be properly explained. Even experts in the field of machine learning can’t understand the deep learning process, and making the algorithm transparent to the public has no substantive significance except formal significance.

The fourth kind of algorithm opacity is due to the time limitation of algorithm transparency. The development of the algorithm is not accomplished in an action, which needs to be constantly developed and improved. The algorithm is usually in the process of evolution. The disclosure of the current algorithm does not mean that it can accurately understand the future algorithm, nor does it mean that it can predict whether the future algorithm exists and what problems exist.

IV. Algorithm Transparency and Algorithm Trust

The purpose of algorithm transparency is to make the public can observe and review the algorithm through disclosing the internal situation of the algorithm system, so as to eliminate the doubts, anxieties and even fears of the algorithm, and then enhance the trust of the algorithm. Because of its abstractness and complexity, the concept of trust has different definitions in different disciplines, such as sociology, psychology, marketing, economics, management, and so on [8]. So far, there is no unified definition of trust. In social science, trust is regarded as a dependency. Trustworthy individuals or groups mean that they seek to practice policies, codes of ethics, laws and their previous commitments. Supporters of algorithmic decision making argue that humans place too much trust in other humans, although in fact, some algorithms have surpassed human experts. Human beings can and are accepting other human mistakes, but they hold higher standards for algorithms, and it is difficult to accept algorithm mistakes. Some studies have shown that after finding evidence that the algorithm may be wrong, subjects are significantly less likely to use the algorithm to output results, even if the algorithm is still more accurate than their own or even human experts' answers. From this perspective, the lack of human trust in algorithms is irrational. However, as we pointed out earlier, algorithms are just as biased as humans because they embed subjective values.

People who are entrusted with trust have a direct relationship with the attribution of decision-making responsibility. One way to avoid responsibility is to avoid talking about who is ultimately responsible. Trusted people have a direct relationship with the attribution of decision-making responsibility. One way to evade responsibility is to remain silent or ambiguous about who should ultimately be responsible.

V. Conclusion

The idea of transparency and openness is taken for granted because it is generally believed that transparency and openness will destroy confidentiality, which is usually considered to hide motivation or intention, and often implies deception. Transparency is often seen as a panacea for many problems related to corruption, fraud, discrimination, etc. In fact, although transparency will destroy confidentiality, it will not reduce false information such as deliberate deception that destroys the trust relationship. If we want to gain trust, we need to reduce deception, not secrecy. There is no logical relationship between transparency and trust. Transparency can even destroy trust. For example, due to the professional complexity of the algorithm, even if the algorithm is open and transparent, the non-professional social public can't understand it, let alone supervise it. Instead, it will be used by professional groups to shape their own authority, and even be corroded by interest groups, so as to serve their interests by misinterpreting the public information. Algorithm scientists play an important role in building trust in artificial intelligence.

Algorithm scientists are responsible for developing data analysis and metrics, so they play an important role in building trust mechanisms. We should establish a set of design norms that are in line with laws and ethics, to ensure that it is impossible for algorithm scientists to bypass these design norms, and guide them to do the right things correctly through the correct organizational technology and technical measures.

References

[1] J.Burrell, “How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms,” Big Data & Society 3 (1), 2016, pp.1-12.

[2] Frank Pasquale, the Black Box Society, Harvard University Press. Cambridge, Massachusetts, 2015.

[3] Paul B.Laat, Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability? Philosophy & Technology, November 2017, pp.1-17.

[4] N. Diakopoulos, “Accountability in Algorithmic Decision Making,” Communications of the ACM, February 2016, Vol 59, No. 2. pp. 56-62.

[5] Fatml, “Principles for Accountable Algorithms and a Social Impact Statement for Algorithms,” 2017. Retrieved from https://www.fatml.org/resources/principles-for- accountable-algorithms.

[6] M. Ziewitz, “Governing algorithms: Myth, mess, and methods,” Science, Technology, & Human Values, 2016, 41(1), pp.3-16.

[7] C. Ducuing, L.Oneto, R.Canepa, Fairness and Accountability of machine learning Models in Railway Market: Is Applicable Railway Laws Up to Regulate Them? ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence, and Machine Learning. Bruges Belgium, 24-26 April 2019.

[8] IEEE Standards Association, “The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems,” 2018. Retrieved from https://www.acm.org/publications/authors/reference- formatting