Reasonable Construction and Legal Regulation of Interpretable Mechanism of Educational Artificial Intelligence Algorithms

Luji Liu[1](✉) and Xietao Cheng[2]

[1] College of International Education, Shandong Jiaotong University, Jinan, China
justiceliu@126.com

[2] College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China

Abstract. The problem of algorithmic "black box" has a negative impact on the deep application of AI in education. The current educational AI still has many problems in data generation and access, algorithm prediction and decision making, and human-computer interaction interface, etc. The establishment of algorithm explainable mechanism is the key to solve the current problems of safety, fairness, and trust of educational AI. Although different stakeholders such as AI developers, users and managers have different concerns about interpretable dimensions, they should reach a consensus on the construction of algorithm interpretable mechanism and follow the basic principles: trustworthiness, fairness, interactivity and causality. The development of science and technology has provided the necessary technical support and tools for the establishment of AI explainable mechanism, and effective algorithm regulation needs technical support and legal system to work in both directions. In the future, it is necessary to establish a legal accountability mechanism with full traceability and further optimize the intellectual property protection system of algorithms.

Keywords: The Artificial Intelligence in education · Algorithm · Interpretable mechanism of algorithms

1 Introduction

In recent years, with the rapid development of artificial intelligence technologies such as machine learning and cloud computing, a new round of scientific and technological revolution is in the ascendant. The wide application of artificial intelligence technology in the field of education has provided new opportunities and platforms for the intelligent transformation of education and teaching. The intelligent education model based on Deep Learning, Cross-border integration, Human-machine Cooperation and other technologies is taking shape. With the application of new artificial intelligence technology, machine learning algorithms begin to give up the requirement of interpretability when designing, and emphasize on improving the processing performance of algorithm generalization [1]. This has led to the "black box" of algorithms, data bias, algorithm

Z. Zhan et al. (Eds.): SEAA 2022, ASSEHR 675, pp. 233–240, 2023.
https://doi.org/10.2991/978-2-494069-05-3_30

unfairness and other problems in the application of artificial intelligence. The application of artificial intelligence in the field of education has had a negative impact and trust crisis. In 2021, UNESCO issued the recommendation on the ethics of artificial intelligence, which put forward the ethical requirements of "transparency and interpretability" of artificial intelligence. Education is a special activity of human beings. Human subjectivity and autonomy are the distinctive characteristics and objective requirements of educational activities. Therefore, the educational field has higher requirements for the interpretability of artificial intelligence. The lack of transparency and interpretability has caused the effectiveness of intelligent education system to be widely questioned by users [2]. Therefore, the establishment of algorithm interpretable mechanism has become the key to solve the problems of security, fairness and trust in current educational artificial intelligence.

2 Necessity Analysis of Establishing Interpretable Mechanism of Educational AI Algorithm

With the development of modern educational technology, artificial intelligence has been widely used in education. However, users' satisfaction with educational AI is not high. At present, educational AI products have many problems in data generation and access, algorithm prediction and decision-making, and human-computer interaction interface.

2.1 Main Application Forms of AI in Education

The main application forms mainly include the following: first, intelligent tutoring system (ITS): with the help of artificial intelligence technology, it plays an important role in helping learners acquire knowledge and skills without the guidance of human instructors. It integrates the teaching resources and research results of national excellent teachers and teaching experts through the Internet computer artificial simulation technology to conduct scientific analysis and intelligent judgment, and provide immediate and effective teaching content; The second is personalized learning system, which is a learning assistant system based on the deep integration of technology and education, which reflects the differences of students' personality and promotes the development of students' personality; The third is educational robot, which is a comprehensive application of artificial intelligence, speech recognition and bionic technology in education. It is mainly divided into teaching activity robot and education service robot according to its purpose [3].

2.2 Users' Satisfaction with AI in Education

In view of the current users' evaluation of educational artificial intelligence products, the results of a recent questionnaire conducted by the author show that the satisfaction of users at the three levels of students, teachers and school administrators is not high. Among them, the satisfaction of teachers is the lowest, only 36.5%, and the satisfaction of school administrators is relatively high, 66.7%. To some extent, the survey results reflect that there is still much room for improvement in the development of educational artificial intelligence products. How to further establish a trust relationship with users is

Table 1. Satisfaction survey of educational AI products.

User group	Satisfied	Basically satisfied	Dissatisfied
Student	45%	18.5%	36.5%
Teacher	36.5%	15%	48.5%
School	66.7%	16.3%	17%

the key to the problem. In the future, it is necessary to further improve the transparency of the algorithm of educational artificial intelligence products, enhance the interpretability of the algorithm, and then enhance the users' trust in the product (Table 1).

2.3 Main Problems of Current AI in Education

While AI is widely used in the field of education, it also brings many negative problems, which affect the effectiveness of the application of AI in education. The main change is that in terms of data generation and preservation, the data as the course content should have higher authenticity requirements; In terms of data access and control, the "digital divide" has affected the supply of education and the risk of privacy disclosure; In the aspect of algorithm prediction and decision-making, the lack of interpretability of push algorithm results in the information cocoon of the educated; In terms of human-computer interface and intelligent devices, the abuse of intelligent devices infringes on the privacy of users and impairs the autonomy of teachers and students. The algorithm black box problem has caused users' trust crisis in educational AI products.

2.4 The Interpretability of Algorithms is Particularly Important for AI in Education

The field of education is an important part of the overall social system which is inseparable from other fields. The artificial intelligence technology used in the education system is generally based on the general artificial intelligence underlying technology and follows the same algorithm prediction and decision-making process. At the same time, educational artificial intelligence is in the special application field of educational system, which shows special requirements. First, the data as knowledge has higher requirements for authenticity, and second, the fairness of education requires the non discrimination of artificial intelligence algorithm, Third, based on the protection of minors and other special groups, AI education is required to pay more attention to the protection of minors' privacy during data collection [4].

The education AI algorithm relies on the comprehensive data generated by users in many aspects for data support, and recommends customized education products for users through big data mining and Deep Learning; The generation mechanism of this seemingly customized learning plan or scheme is beyond the direct understanding of the general public. It has no ability to judge and analyse, and only has the option to choose whether to use it or not. The user's cognition is gradually surrounded by algorithms.

However, this learning recommendation algorithm makes the user unable to distinguish whether it is reliable or not [5]. For a long time, the recommendation algorithm may create an "information cocoon" of bias for users. Bad learning habits and wrong learning content may not be corrected by themselves, which is contrary to the essential function of education.

The report "short read: artistic intelligence and school education" released by the Australian Ministry of education points out that artificial intelligence education is faced with the dilemma of "it is difficult to understand the accurate logic of machine decision-making", "For the humanities such as education, it is the core principle to explain why such a decision is made by using intelligent systems for teaching and management" [6]. Therefore, in order to ensure the effective application and sustainable development of artificial intelligence in the field of education, the interpretability of algorithms is a problem that must be solved in the research and development of intelligent education.

3 Principles and Methods for Establishing Interpretable Mechanism of Educational AI Algorithm

The goal of establishing algorithm interpretable mechanism is to ensure the high performance of intelligent decision-making and give a reasonable interpretable model, so that human users can understand, trust and effectively manage artificial intelligence.

3.1 The Meaning of AI Algorithm Interpretability

Interpretability refers to the specific process and method of artificial intelligence system design, so that ordinary users can understand and trace the causal relationship between the input and output of algorithm module and the basic mechanism of system operation. From a technical point of view, algorithm interpretability is the technical premise for implementing the transparency principle, and it is a design principle to ensure that technology serves human beings. In the second generation artificial intelligence technology, especially the deep learning algorithm model, the learning results, technical principles and processes of the deep learning model are more complex, which often forms a "technical black box" that is difficult for ordinary people to understand. The requirement of algorithm interpretability is the response to this problem. The establishment of algorithm interpretability mechanism has become an objective requirement for the healthy development of artificial intelligence. Therefore, the ethical position that must be adhered to under the concept of technology serving people is that there is no absolute unexplainability of human intelligence, and unexplainable artificial intelligence should be prohibited from being used for human beings.

At present, the interpretability of algorithms in related research is mainly expressed by interpretation, explanation and understanding; There are some differences in the connotations of the three: "interpretation" means mapping abstract concepts to fields that human beings can understand; "Explanation" is a feature set of interpretable domain, which is used to explain the decision-making process of a given instance; "Understanding" refers to the functional interpretation of the model.

3.2 Principles for Constructing Interpretable Mechanism of Educational AI Algorithm

Although different stakeholders such as AI developers, users and managers are educated about different interpretable dimensions of interest, consensus should be reached on the construction of interpretable mechanism for algorithms, and basic principles should be followed so as to ensure the realization of the goal of interpretable mechanism for algorithms, that is, to eliminate the opacity and inequity caused by algorithm black boxes, and to build a reasonable interpretable model while ensuring the performance of algorithm decision making. Generally speaking, the construction of interpretable mechanism of algorithms should follow the following principles [7]:

Firstly, the principle of trustworthiness is that when people face an algorithmic decision, they can accept and trust the results of model operations. The algorithm has excellent performance and stability, and can be interpreted as trusted by the general public in a technical way. The principle of trustworthiness is especially important for educating AI algorithms and is the first principle for constructing an interpretable mechanism of algorithms.

Secondly, the principle of fairness, which is from a social ethical point of view, prevents prejudices in the algorithm model, and provides explanations that can withstand the public's consideration of the fairness of algorithm decision-making. Fairness is an important value pursuit of education. The principle of fairness should become an important principle in constructing explanatory mechanism of educational AI algorithm.

Thirdly, the principle of interactivity means that users can exert influence on the algorithm model and have the right to optimize and adjust independently. The majority of users of educational AI products are non-computer technicians. Users participate in the design and operation of the decision-making model and put forward optimization suggestions according to the using needs. The optimized algorithm can improve the user experience and promote the establishment of trust relationship with users.

Finally, the causality principle refers to that the algorithm interpretation mechanism should help the general public understand the causal relationship between data input and output, and make a general understanding of the correlation between data.

3.3 Technical Tools for Constructing Interpretable Mechanism of Educational AI Algorithm

Although in the implementation of interpretable technology, educational AI can design and develop algorithms according to its own system requirements, and can also choose open-source interpretable AI tools. Choosing mature and open source technology tools is an effective means to improve the efficiency of educational system and educational product design and development [8] (Table 2).

Table 2. Main AI algorithm interpretation tools.

Tools	Local or global	Language type	Tool features
LIME	Partial interpretation	Python, R	Suitable for tabular and image, poor stability
SHAP	Local/Global	Python	Calculating the marginal contribution
PDP	Global	Python	Intuitive and easy to realize
AIX360	Local/Global	Python	Provide multiple interpretable methods

4 Legal Guarantee for Establishing Interpretable Mechanism of Educational AI Algorithm

The establishment of a scientific and effective algorithm interpretable mechanism is inseparable from the support of technology. Of course, it also needs the protection of the law. The law establishes a system and investigates the responsibility with the force of the state to promote the healthy development of interpretable and reliable educational AI.

4.1 Top-Level Design of Legal Regulation

The establishment of algorithm interpretable mechanism needs both technical support and legal regulation. The design and development of algorithms often pursue the optimal solution in mathematics, pay attention to efficiency first and ignore the value of law and ethics. Therefore, it is necessary to carry out legal regulation, embed human values and ethics into the algorithm, and strengthen the control of legal regulation over algorithm power. Of course, in the process of regulating the algorithm through the law, we also need to rely on the information and technical advantages of the algorithm platform, programmers and artificial intelligence experts to achieve cooperative governance. Therefore, algorithm platform enterprises, programmers and artificial intelligence experts are not only the objects of government regulation, but also the participants, decision makers and executors of government regulation. The formulation and implementation of national laws and regulations on algorithm regulation are inseparable from their active participation [9].

The legal regulation of the algorithm should realize the transformation from terminal regulation to whole process regulation. The competent department should upgrade the regulation idea and carry out the whole process regulation of "prevention in advance, supervision in process and relief afterwards". First, from the beginning of the design of algorithm technology, the law and technology are linked from the source, weakening the separation of algorithm technology and law. Secondly, in the process of the operation of algorithm power, we should strengthen the supervision and accountability of the operation of power, improve the transparency of the operation of algorithm power, and establish a regulatory system combining supervision and review. Finally, in view of the external negative effects brought by the algorithm, the specific legal relief after the event is carried out to make the interpretable right of the algorithm related to the accountability mechanism of the algorithm [10].

4.2 To Establish Legal Accountability System for Educational AI Application

The establishment of legal accountability system is an important measure to ensure the construction of interpretable mechanism of educational AI. If the algorithm black box in the application of artificial intelligence education is not handled properly in advance, then an accountability system should be established afterwards to investigate the legal responsibility of the responsible subject for the negative results of the operation of educational AI. In reality, when teachers, students and other groups are adversely affected by intelligent decision-making systems, they often passively open their privacy and protect their rights to data managers. Due to the complexity of artificial intelligence systems, it is difficult for users to determine the responsible subjects. Therefore, the establishment of the accountability system requires a full process information and data traceability system, and a complete data tracking and recording scheme to ensure that the algorithm process is transparent and that the responsible subjects can be quickly found and determined.

4.3 To Modify the Intellectual Property Protection System of Relevant Algorithms

From a technical point of view, the algorithm black box is not a real black box, but rather a complex and similar black box. In the final analysis, the problem of explicability is the relationship between input and output. As long as each factor and link is explained clearly, the algorithm will become well known to the public. However, in the era of market economy, the core structure of algorithm is undoubtedly an important business secret of enterprises, and its content is protected by intellectual property law. Therefore, the establishment of algorithm interpretable mechanism needs to balance the relationship between algorithm intellectual property protection and user interpretable right. The content of the algorithm is hierarchical. The structure of the algorithm model, model training methods and training data for the development enterprise, which core technologies as trade secrets need to be protected, and which necessary information as the user's interpretable right must be made public, which needs to be defined in law. The intellectual property protection legal system of the algorithm needs to modify and improve the disclosure rules, and clarify what to protect, what to make public, and what procedures to make public.

5 Conclusions

Only focusing on the breakthrough in AI technology cannot achieve a qualitative breakthrough in the development of AI. In the future, we should not only improve the innovation ability in the field of AI, but also strengthen the construction of ethical and credible AI. This is especially true in the field of educational AI. The establishment of AI interpretable mechanism needs the support of technological progress and the protection of legal system. In the future, it is necessary to improve the top-level design of legal regulation, establish a full traceable accountability mechanism, adjust disclosure rules and optimize algorithm intellectual property protection. Interpretable artificial intelligence is the key to open the "black box" of intelligent education. The prediction and decision-making provided by intelligent systems and tools can really promote the development of fair, just, high-quality and efficient future education only if they are understood and trusted by teachers and students.

References

1. Chen, K., & Meng, X. (2020). The interpretability of machine learning. *Journal of Computer Research and Development, 57*(09), 1971–1986.
2. Sun, B. (2022). Interpretable AI: The key to the black box of future intelligent education. *Journal of China Education Informatization, 28*(04), 3–4.
3. Yang, K., & Dong, X. (2022). The application status and prospect of AI technology in education. *Journal of Science and Technology Innovation and Application, 12*(12), 189–192. https://doi.org/10.19981/j.CN23-1581/G3.2022.12.044
4. Miao, F. (2022). The analysis and governance of AI Ethics–the education interpretation of AI ethical issues proposal. *Journal of China Electrified Education, 06*, 22–36.
5. Yu, C., & Liu, F. (2022). Research on the ethical risks and Countermeasures of AI education application. *Journal of Robot Industry, 02*, 32–37. https://doi.org/10.19609/j.cnki.cn10-1324/tp.2022.02.005
6. Hall, T., Cao, M., Ming, Z., & Yuan, L. (2022). Educational perspective that can explain artificial intelligence: Thinking Based on ethics and literacy. *Journal of China Education Informatization, 04*, 5–13.
7. Kong, X., & Tang, X. (2021). A survey of AI decision interpretability. *Journal of Theory and Practice of Systems Engineering, 41*(02), 524–536.
8. Wang, P., Tian, X., & Sun, Q. (2021). Interpretable educational AI research: System framework, application value and case analysis. *Journal of Distance Education, 39*(06), 20–29. https://doi.org/10.15881/j.cnki.cn33-1304/g4.2021.06.003
9. Zheng, Z. (2021). The ethical crisis and legal regulation of AI algorithm. *Journal of Social Sciences Abstract, 04*, 74–76.
10. Zhao, Y., & Chen, L. (2021). Algorithm power alienation and legal regulation. *Journal of Yunnan Social Sciences, 05*, 123–132.