International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Privacy Risk and Correction Mechanism of Higher Education MOOC Based on Blockchain Technology in the Era of Big Data

Chengfei Wei[1], Fangjian Shang[2]

[1]Lecturer, Academy of Marxism, Nanjing University of Finance and Economics, Nanjing, China, P.h.D, Jindal Institute of Behavioural Sciences, O.P. Jindal Global University, Haryana, India
[2]Research Assistant, Nanjing University of Finance and Economics, Nanjing, China

Abstract:
The emergence of Massive Online Open Course (MOOC) has sparked a digital tsunami globally, bringing a huge impact on higher education. The open curriculum format of MOOC has achieved the optimal transmission of top teaching resources and personalized teaching based on different groups, which is different from traditional teaching models. At the same time, in the process of implementing educational activities, the MOOC platform based on big data mining will bring a certain degree of privacy risk due to improper data collection, storage, or processing. The traditional MOOC have not achieved as much as they should. The main problem is that fully centralized online education cannot solve the trust problem. The emergence of blockchain technology solves the problem of lack of trust, but the open and transparent mechanism of blockchain also introduces a new problem, that is, the privacy of user data can not be reasonably protected.

Keywords: Big Data; MOOC; Blockchain; Privacy Risk; Correction Mechanism

I. Introduction

The emergence of MOOC has brought about the widespread dissemination of knowledge. More learners can choose online courses independently, learning according to their own needs anytime and anywhere, avoiding a long and complex application process, and participating in the learning process between different learners through MOOC, reflecting educational innovation and creating an interactive learning environment. In addition, it cannot be ignored that the MOOC system may “reveal some new facts about individuals” in the process of mining sample information based on big data analysis, which may trigger privacy risks. With the development of blockchain education, the MOOC platform that uses blockchain technology as the underlying capability has gradually been accepted and recognized, which can endorse the reliability of MOOC course data, and also improve the authenticity of students’ learning data to a certain extent. However, at present, the MOOC platform based on blockchain technology still lacks sufficient understanding of the privacy protection of personal data. Although it makes use of the open and transparent characteristics of blockchain technology, it lacks privacy protection and access control of data. It is urgent to promote the sustainable development of MOOC by studying the technological advantages, and privacy risks and exploring relevant corrective measures of MOOC in the era of big data.

II. The Technological Advantages of MOOC

2.1 The MOOC mode achieves optimal transmission of teaching resources

In the context of Big data, the rapid development of sophisticated technologies such as data storage and artificial intelligence has promoted the MOOC model breaking the bottleneck of traditional teaching. The MOOC model gathers massive teaching information from top universities around the world such as Harvard University, Massachusetts Institute of Technology, and University of Oxford, and has the characteristics of high speed, low cost, high value, and multiple types, which significantly increase the coverage of virtual teaching content.[1] The MOOC teaching platform is no longer limited to traditional professional categories and auxiliary classroom teaching but is more closely related to the special needs of learners, emphasizing the comprehensiveness, universality, and sustainability of knowledge. The MOOC model based on the globally interconnected system is aimed at all netizens, with strong openness beyond traditional localized online teaching, creating a virtual classroom where learners are not limited by time and space and can freely participate in learning through any chain network channel. With the wide audience, the individual's right to education will also be expanded to the extreme. For example, the Coursera Platform has gathered over 1.6 million learners around the world since its launch.

2.2 The MOOC mode realizes personalized teaching

sThe large-scale MOOC system stores learners' login traces, educational records, examination scores, and other information. In fact, all trajectory data of any MOOC platform participant will be thoroughly collected. The system provider establishes the personalized teaching model by deeply mining these data, analyzing learning situations, summarizing course requirements, and other important factors that lead to decision-making, and combining the development characteristics of learners themselves. The low-cost individualized teaching model is conducive to helping learners better grasp progress and acquire knowledge and is a green way to cultivate versatile talents. Learners can arrange the progress according to their own learning situation, and choose their favorite courses according to their own hobbies. The system will optimize the learning experience and automatically adjust the learning plan to meet the learners' own needs. This also means that learners can participate in every aspect of the MOOC, greatly improving the initiative of learning.

2.3 The natural advantages of MOOC Sociality

The 'read/write' nature of social media and the internet has become a part of online education. In the traditional classroom, each learner is a separate individual, requiring only to do his own and ignoring cooperative learning with others. MOOC has expanded globally and integrated social tools, which can better match students' social learning tendencies. The open MOOC environment allows learners of any age, location, socio-economic or educational background to participate, providing learners with various ways to communicate and learn from each other, improving their divergent thinking, and expanding the development of academic networks. In addition to video lectures and quizzes, learners are able to discuss open-ended issues during MOOC learning, which is a constructive way for learners to process their knowledge and collaborate with others. In addition, MOOC uses peer rating or crowdsourcing to grade, and by reviewing, comprehending, and reflecting on peers' homework, learners can benefit from it and become aware of their own strengths and weaknesses.

III. Privacy Risks of Data Mining on the MOOC Platform

3.1 Manifestations of privacy risks in MOOC

3.1.1 Direct infringement of privacy caused by improper data collection and storage in MOOC

Big data can collect insights and breakthroughs related to data analysis, such as the side effects of drugs and the wave of quantitative criminalization. Its data set can be concentrated in a very small range (such as health research) or a large range (such as consumer marketing). In MOOC, if you choose to study a course, you must first create a MOOC account. For example, to register for edX, you need to create a username and password and provide the following information: your email, name, country, gender, year of birth, highest education, and reason for registration. In order to complete the registration, you must agree to edX's terms of service and privacy guidelines and these are not special requirements for website registration. But once you start the MOOC you choose, a new set of counting points will be created during the participation process, which will be collected by MOOC. [2] In addition, when you access a module (virtual equivalent unit of the class), how often do you access the module, how long does it take to complete the test, the test score, the number of times you watch videos, and whether you stop watching the course, the edX log will record the whole process. [3] In summary, edX collects approximately 20 billion bytes of user data for each course (equivalent to millions of physical pages of information). [4] In terms of exam invigilation, learners are required to allow remote access to computers and have cameras and microphones, so that invigilators can maintain continuous contact with learners during the exam. Due to the uncertainty of whether they can still access students' computers after the exam and the inability to guarantee what invigilators can see on learners' computers during the exam, this not only puts unnecessary financial pressure, it also seriously infringes the privacy rights of learners. Moreover, a large amount of personal privacy information of learners is centrally recorded in educational archives. Big data mining in MOOC involves the collection of a large number of user education files, which are kept by educational institutions, educational associations, or third-party agencies. If the well-intentioned insiders of these institutions fail to comply with the security policy, they may inadvertently cause data breaches.

3.1.2 Association analysis of scattered data in MOOC reveals learners' privacy

The privacy problems caused by big data mining are because it involves a large amount of personal information collection and because the MOOC system may "reveal some new facts about individuals" in association with mining sample learning information based on big data processing technology, which may trigger privacy risks. Big data analysis can be predicted and inferred to form a new anonymous data set of personal authentication information, which causes a basic privacy protection problem and obtains an unexpected inference. [5] For example, suppose you are a father with a daughter who is in middle school. According to her shopping habits, big data finds out that your daughter may be pregnant through mining and analysis. Therefore, the supermarket sends your daughter preferential information and advertisements related to pregnancy. This activity seems harmless, but as a father, you can only complain by calling. However, the data analysis shows that the conclusion of "pregnancy" has nothing to do with the identity of the daughter and adolescent, and may only be based on the purchase of other products such as odorless lotion and vitamin supplements related to pregnant shoppers. Similarly, on the surface, the data collected by MOOC appears to be chaotic, but when the data accumulates to a certain extent, group behavior will exhibit order and pattern. Unlike traditional teaching, MOOC can study learners' learning trajectories by recording mouse clicks and discovering different reactions of different learners to different knowledge points, such as how much time it takes, which knowledge points need to

be repeated or emphasized, and which statement or learning tool is the most effective. By opening up to the world and allowing the most learners to learn and use it, the MOOC can collect the most data and study the behavioral patterns of learners from various countries around the world. Although this information does not reveal the identity of MOOC users, the prediction of big data is surprising. Moreover, the operations in MOOC are all carried out through the network, and there are very few unencrypted data. Assuming that during the process of MOOC collecting data, MOOC users' data is accidentally or intentionally leaked, even if it is only one of the many MOOC education records, information related to MOOC users can be known, leading to privacy leakage. For example, when you log into a Coursera account and visit websites, these websites can list your course registration records.

3.1.3 Poor supervision during the outsourcing or transfer of MOOC data infringes learners' privacy

"MOOC" is synonymous with "outsourcing". Curriculum development, curriculum management, curriculum content, curriculum communication mode, teaching design, result evaluation, information services, learning channels, funding estimation, etc. are all delivered to third-party providers (that is, the MOOC platform is responsible for these originally undertaken by the university). In fact, the provider of "MOOC" already faces the challenge of sustainable development. On the one hand, the courses in MOOC are open and have the attribute of being free. On the other hand, the maintenance, updating, and development of MOOC face enormous cost pressure. If there is no feasible profit model, the provider of MOOC is likely to be overwhelmed by operating costs. There have been too many similar cases in the history of the Internet, and such cases will continue. Therefore, the providers of MOOC must find practical and feasible profit models, and customize courses for other educational institutions (such as vocational colleges) through data mining in MOOC. These educational institutions pay corresponding fees, which is a good profit model. However, in the process of data outsourcing transfer, if individual departments in the organization do not authorize information operation according to their job roles, and do not follow the principle of "minimum privilege" for authorization management, the authority owned by employees will substantially exceed the requirements of their job responsibilities. For user-sensitive information operations (such as location queries, user call record queries, accounts with sensitive business permissions, etc.), there is no regular audit conducted, and problems cannot be identified in a timely manner. Some organizations even sell the educational files of MOOC users to other companies, websites, or researchers for considerable economic benefits, which leads to privacy leakage and infringes on the privacy of participants.

3.2 The causes of privacy risks

3.2.1 The irrefutable big data forecast

A common phenomenon predicted by big data is the discovery or development of personal verification information. For example, for decades, many countries have prematurely assigned students to different learning tracks, usually including the following three categories: Vocational education for underperforming students; Ordinary courses for general students; Pre-university courses for outstanding students. The system predicts that the success probability of a particular student may be based on the performance of 1 million other students, and will provide a direct customized education to that student. Customized education actually makes it more difficult for students who attempt or have the ability to break specific tracks, making them victims of probability prediction rather than their own abilities. This is based on two points: (1) Most people do not realize what big data will bring and how much it will

bring; (2) The regulatory solution to this transparency problem is that the data has been anonymized rather than requiring individuals to be explicitly informed of its collection. From this perspective, privacy has been optimized before data usage. Most discussions on protecting privacy seem to be accepted, and importantly, it feels like others can easily know too much about us, even though the results of this knowledge may never seriously affect our lives.

3.2.2 The irresistible path to wealth

The analysis results of MOOC will have enormous commercial and social value and have significant value in improving teaching standards, predicting student employment directions, selling auxiliary learning tools, and even judging the economic situation and industrial development prospects of the entire society. Personal data, whether created by individuals or data agents, is becoming increasingly detailed and can be transferred, predicted, and profitable. Therefore, after mastering a large amount of data, providers will outsource or transfer it to third parties to obtain economic benefits. The New York Daily News reported on inBloom's project in 2013 (with $100 million in support from the Bill & Melinda Gates Foundation and Carnegie Corporation, signing agreements with nine states and establishing student data repositories): "This is an unprecedented move by the education department to transfer students' personal information to newly established private companies for the establishment of a national database to serve enterprises that sign contracts with public schools." [6] For a long time, education institutions have always been vertically integrated into management based on data and its flow methods, and the institutions that produce and collect data are often the ones that analyze the data. The school not only provides performance and other feedback data, and also serves as an entity for storing and using this information. The school uses this data for decision-making and communicates the decisions to students, parents, future employers, and other schools. From admission, providing guidance, evaluating performance, and awarding certificates, all handled by educational institutions. For example, various platforms are starting to make money through paid certificates. Coursera announced a special course, and students must complete a series of paid courses before obtaining a certificate.

3.2.3 Lack of relevant policies and legal norms

Most countries have enacted some form of privacy protection law to prevent the comprehensive collection and long-term storage of private information. Generally speaking, these laws require users of data to disclose the objects of data collection and the purpose of the data and to obtain the consent of the other party before using it, but this method has little effect. These laws only stop at the prohibition of privacy infringement and have not formulated detailed punishment measures for privacy infringement sanctions. For example, the provider and operator of the MOOC platform make users agree to release the information that will not disclose their personal identity or agree to data anonymization of the disclosed information by clicking on the contract, so as to determine their property rights to MOOC data and reduce their security responsibility. But the actual effect is not well, the ID of users who are protected by privacy rights actually does not have any privacy protection actions. Secondly, the enforcement efforts are insufficient. The value of big data mainly lies in the fact that the data can be reused, and the purpose of data reuse is hardly considered when collecting data initially. So obtaining informed consent before data collection is often not possible. Many benefits of big data in education cannot be realized if the spirit of legal protection is strictly implemented. Therefore, privacy infringement during the operation of MOOC is often unpunished, resulting in MOOC platform providers, outsourcing contractors, and data transfer recipients all daring to engage in rampant infringement.

IV. MOOC Mining Correction Mechanism

4.1 Using technology, contracts, and other means to protect user privacy

The reasons for the privacy risks of MOOC are multifaceted, with the biggest driving force being technological advancements. On the other hand, utilizing technological advancements can protect the privacy of MOOC users. The existing methods for protecting educational records often overlook the indirect identifier of data, the data that cannot clearly identify individual data points but can be combined with other data to reveal personal identity. In fact, through technology, educational records can be data anonymization, thus protecting the right to privacy functionally. On the other hand, strengthening the regulation of personal data usage through technology also plays a supervisory role in protecting privacy rights. For example, carefully review the implementation of the big data system by relevant institutions, or act as an expert consultant to the regulatory authorities to test the use of Big data.

In addition, online click contracts are used in the MOOC platform. In essence, click contracts are format contracts. By clicking contracts, MOOC platform providers and operators make MOOC users agree to release information that will not disclose their personal identity or agree to data anonymization the disclosed information, so as to determine their own property rights to data and reduce their security responsibility. Therefore, it is necessary to strengthen the formulation of contracts, clarify the responsibilities and benefits of MOOC users, providers, and third parties through contract terms, and agree on a reasonable accident compensation mechanism to protect privacy.

4.2 Strengthen the self-discipline awareness of MOOC providers and third-party data users, and form industry self-discipline

It can be seen from the behaviors of providers and operators who participate in MOOC to reduce their privacy protection responsibilities through service agreements and other ways that they should strengthen the industrial standards for protecting privacy in the Internet, standardize the behaviors of network platforms and network practitioners, formulate and publicize privacy protection measures, clearly inform users of their data security commitments and risk tips, and adopt relevant technical measures to protect users' personal information security. At present, by strengthening moral and legal education, the overall quality of all participants in MOOC can be effectively improved, enabling industry self-discipline to play a greater role than before. At the same time, the public's understanding of the common self-discipline methods in the industry is gradually deepening. In this context, it will inevitably encourage more and more MOOC-related institutions to use technical measures to protect users' privacy rights. Many countries have experience indicating that reasonable industry self-discipline standards, legitimacy review by regulatory agencies, open operations, and free communication will further enhance the effectiveness of industry self-discipline and promote its greater role. Therefore, it is necessary to strengthen the self-discipline and industry self-management of MOOC, strictly register the personal information and relevant information storage system of MOOC users, adopt technical measures to protect personal information security, establish a self-discipline organization to protect the right to Internet privacy and a certification body to protect the right to Internet privacy. At the same time, various institutions participating in MOOC should supervise each other and continuously promote the formation of a positive atmosphere of conscious compliance and mutual supervision in the MOOC industry.

4.3 Improve legal norms related to privacy rights and clarify the status of privacy rights in the law

At present, many countries have formulated laws to protect personal privacy. For example, the United States, the European Union, and other countries have also improved privacy legislation specifically for the characteristics of the big data era. For example, the Family Education Rights and Privacy Act in the United States protects students' educational records. It stipulates that FERPA protects the disclosure of citizen information collected by schools and government agencies only for specific and legitimate purposes. [7] Except for specific circumstances stipulated by laws and regulations, any educational institution must obtain permission from the parents of students or students over the age of 18 to publicly disclose their educational records. Therefore, in the process of protecting the right to Internet privacy, we should treat the right to Internet privacy as an independent personality right, separate the right to privacy from the right to reputation, and add legal provisions related to the right to privacy, so as to truly protect the legitimate interests of citizens. In addition, administrative rules should be improved, network management should be strengthened, and rules for educational institutions at all levels should be strictly formulated. Educational institutions should provide users of MOOC with the opportunity to make mistakes, and only with the written consent of the user or their guardian can they have the right to use their educational records. If educational institutions, MOOC providers, operators, etc. disclose their educational records without their own authorization, administrative rules should also impose corresponding penalties. Of course, there are exceptions to legal requirements that require public records.

V. Conclusion

Since the concept of MOOC was proposed in 2008, MOOC has developed rapidly around the world and brought us a new educational experience, achieving the optimization of educational resources and personalized teaching, and to some extent promoting educational equity. In this process, whether it is a provider of MOOC, traditional universities, or other educational institutions, they have all made contributions to the development of MOOC. However, in the implementation process, due to improper data collection and processing, ineffective third-party data supervision, the lack of corresponding legal norms, and the characteristics of Big data itself, the privacy of MOOC users under MOOC cultivation mode was violated. In this regard, we should actively take appropriate corrective mechanisms to protect users' privacy rights. Firstly, sufficient support should be provided in terms of technical means, and users' educational records should be fully protected; Secondly, strict supervision should be imposed on MOOC platform providers and others to encourage them to form industry self-discipline and consciously protect user privacy; Once again, improve the legislation on privacy rights in order to clarify the status of privacy rights, and provide reasonable supervision and corresponding measures for relevant educational institutions, in order to achieve the rectification, punishment, and warning effects of privacy violations. Finally, the MOOC platform should also provide a readable privacy policy, listing specific goals for data disclosure in a specific form, or providing legal norms on how to use data.

Acknowledgement

This study is supported by the 2023 Ministry of Education Industry-University Collaborative Education Project "Ethics Course of Blockchain Technology and Application" [Grant No. 230801791163401] and the 2023 Nanjing University of Finance and Economics Teaching Reform Project "Evidence Acquisition and Innovation Research of Ideological and Political Education in Colleges and Universities from the Evidence-Based Perspective of Numerical Wisdom" [Grant No. JGX2023011].

REFERENCES

1. Chen Jianlin. Research on Foreign language Teaching in MOOCs in the era of Big Data: Challenges and opportunities [J]. Foreign Language audio-visual Teaching, 2015.1, p3-4
2. See Register, EDX, https://courses.edx.org/register (last visited July 14, 2023).
3. See Jon Daries, The HarvardX-MITx Person-Course Dataset AY2013, HARVARD DATAVERSE NETWORK 3–4 (May 27, 2014), http://dx.doi.org/10.7910/DVN/26147. (last visited July 14, 2023).
4. See A. D. Ho et al., HarvardX and MITx: The First Year of Open Online Courses 5 (HarvardX and MITx, Working Paper No. 1, 2014), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2381263. (last visited July 14, 2023).
5. See Kate Crawford & Jason Schultz, Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms, 55 B.C. L. REV. 93, 98 (2014).
6. Corinne Lestch and Ben Chapman, New York Parents Furious at Program, inBloom, That Compiles Private Student Information for Companies That Contract with It to Create Teaching Tools," New York Daily News, March 13, 2013, http://www.nydailynews.com/new-york/student-data-compiling-system-outrages-article-1.1287990. (last visited July 14, 2023).
7. See Legislative History of Major FERPA Provisions, US DEP'T OF EDUC. 2 (June 2002), http://www2.ed.gov/policy/gen/guid/fpco/pdf/ferpaleghistory.pdf. (last visited July 14, 2023).