

Research on Machine Learning Copyright

Shuheng Wang

Hebei University, Baoding, 071000, China

Abstract: A new round of scientific and technological revolution and industrial transformation led by artificial intelligence is on the rise. Driven by new theories and technologies, artificial intelligence presents new features such as deep learning, cross-border integration, human-machine collaboration, swarm intelligence openness, and autonomous control, and is having a significant and far-reaching impact on economic development, social progress, and global governance. Artificial intelligence technology has made a leap forward, and the ensuing legal issues have aroused discussions in the academic and practical circles, mainly focusing on the copyright law of artificial intelligence products, including the legal attribute of the generated results and its protection path, as well as exploring the new path of machine learning infringement risk and its compliance governance.

Keywords: Generative AI, artificial intelligence, machine learning, copyright.

1. Introduction

Artificial intelligence technology is the key content of the current new round of scientific and technological revolution and industrial change. China attaches great importance to its application and development, and actively creates a legal environment conducive to the application of artificial intelligence technology and industrial development, responding to national strategic concerns while enhancing China's international competitiveness in the field of artificial intelligence. With the successive launch and rapid iteration of domestic and foreign generative AI products such as "ChatGPT" and "Wenxin Word", AI large models have become strategic "heights" for competition in various countries. However, when the industry is actively adding fuel to the development of generative AI's "overtaking", it is also of important theoretical value and practical significance to objectively understand and analyze the copyright risks faced by machine learning and strive to find ways to deal with and resolve them.

2. Research Background

The development of machine learning has penetrated into many fields, providing intelligent solutions to various problems. The continuous innovation and progress in this field indicates that more fields will benefit from machine learning technology in the future. For example, machine learning has made significant progress in processing natural language (NLP): Applications including speech recognition, text analysis, sentiment analysis, etc., have been widely used in virtual assistants, intelligent customer service, intelligent translation and other fields [1]. In addition, in the field of computer vision, machine learning applications for image and video processing have gradually matured (including image recognition, target detection, face recognition, etc.), and have been widely used in security, medical imaging, autonomous driving and other fields. For example, machine learning is also widely used in the financial field of financial risk management, fraud detection, stock market prediction and other aspects, among which algorithmic trading, credit scoring and other benefits from machine learning technology. But at the same time, when the "creation" began to become popular, the challenge brought by this intelligent "creation" to

the traditional copyright theory has become increasingly prominent.

3. Application of Copyright Law in Machine-generated Aspect

The specific application of copyright law in the field of machine generation is centered on the attribution and originality of creation. Copyright law usually deals with the protection of originality and the rights and interests of creators, but when it comes to machine-generated content, there may be some challenges and changes in the application of the law due to the creative behavior of the machine. As machine-generated content continues to evolve, copyright law may need to evolve to accommodate new technologies and creative methods. Lawmakers and legal bodies also need to keep a close eye on developments in this area to ensure that regulations and legal systems are able to effectively address the challenges posed by machine-generated content.

The first is the issue of attribution. In the traditional copyright law, the creator is the subject who is granted copyright. However, whether a work generated by a machine has an independent creator identity in machine-generated content is a complex issue. The legal community and legislatures may need to review regulations to clearly define the creators of machine-generated works and decide how their rights should be protected [2].

Besides, copyright law generally requires that works be original and innovative. For machine-generated content, it may be necessary to consider whether the design of the machine algorithm is innovative enough, and whether the specific work generated is original enough. The law may need to be adapted for these particular situations. Moreover, if the content generated by the machine is subject to human intervention or modification, there may be a human creative component involved. In such cases, copyright law may need to distinguish between machine-generated parts and human-created parts, and provide appropriate protection for each part. Finally, it is about the legal application of new creative forms. Machine-generated content may take forms that are not available in traditional creative methods, such as generated artwork or music. The law may need to follow up to ensure that these new forms of creation receive appropriate legal protection.

4. Copyright Disputes and Legal Challenges in Machine Learning

The main reason for the intelligent leap in generative artificial intelligence lies in the large-scale use of data and the performance leap of algorithms, but the accompanying risks are further amplified. China's "first case of AI large model data theft" and the United States "OpenAI sued" as typical cases of generative artificial intelligence machine learning infringement at home and abroad, also shows that the current risk of machine learning infringement needs to be solved. In the field of machine learning, copyright disputes and legal challenges relate to machine-generated works, as well as issues related to determining attribution and rights to works. These issues reflect the complexity of copyright in the field of machine learning, and addressing them requires the legal system to evolve to accommodate emerging technologies and forms of creative expression, taking into account factors from multiple fields of law, technology, and ethics.

The first is the recognition of creativity and originality generated by machines. Whether a work generated by a machine is creative and original enough to meet the requirements of copyright law is a point of contention. Sometimes machine-generated content can be based on large amounts of data, sparking controversy as to whether it is truly original. Secondly, it is necessary to consider the complexity of the work after manual intervention. If a machine-generated work is subject to human intervention or modification, rights can be more complicated. The law needs to consider the rights of artificial creators and determine the impact of human intervention on copyright. In the judgment of copyright ownership, generally speaking, the developer who designed the auxiliary artificial intelligence dominates the generation result of the product, and the copyright of the product should belong to the artificial intelligence developer. Furthermore, it involves determining the identity of the creator and the ownership of the machine, which is a challenge to determine the identity and ownership of the creator of the machine generated works. This may involve the designer of the machine, the maintainer, the provider of training data, etc., and the law needs to clarify the status of these parties in the ownership of copyright.

5. Suggestions and Solutions

It is an important prerequisite to protect the rights and interests of the right holders and the interests of the majority. The rapid development of machine learning technology has made existing legal frameworks lag in the face of emerging issues. The technological development and application scenarios of generative artificial intelligence are constantly updated, and limited by the lag of justice, it is difficult to effectively eliminate risks in the bud by relying only on the power of justice. In view of the risk governance difficulties and breakthroughs of generative artificial intelligence, it is more necessary to strengthen the optimization of the path of source governance, promote the synergistic effect of multi-party linkage, and build a pluralistic co-governance pattern of artificial intelligence.

5.1. Protecting legitimate rights and interests

Artificial intelligence creation itself does not have the ability to produce "new works", if we let artificial intelligence deprive the rights holders of works of due interests, then it will wipe out the passion of human creators, equivalent to

"killing the goose that lays the golden egg". First, to distinguish between the protection of artificial intelligence products and human creation. From the perspective of the construction of the copyright system and the protection of the copyright of the authors of literary, artistic and scientific works, there are essential differences between artificial intelligence products and human works in the subject of creation and originality, and the intensity of copyright protection should also be different. Second, accelerate the development of industry standards and regulations for copyright protection in the field of artificial intelligence. Guiding generative artificial intelligence to abide by the intellectual property system requires industry related personnel to pay attention to the protection and respect of copyright when training and using artificial intelligence to create [3]. Third, explore the income distribution and responsibility bearing system of artificial intelligence products. From the general trend of social development and scientific and technological progress, the rapid development of generative artificial intelligence is unstoppable, and "strong protection" does not mean "strong resistance", but actively explore a governance system that can promote industrial development and social progress on the basis of protecting traditional copyright. Starting from the principle that income is proportional to responsibility, we can clarify the income acquisition proportion and responsibility bearing capacity of each subject in the development of generative artificial intelligence industry, and find a Chinese-style governance way that can actively promote the development of generative artificial intelligence industry.

5.2. Achieving multi-party collaborations

First, it is necessary to establish a multi-party cooperation mechanism, including the government, industry organizations, academia, enterprises and social groups, to jointly participate in the discussion of machine learning copyright issues and the development of solutions. Second, there is also a need to develop widely applicable machine learning copyright guidelines and standards that can be used in different countries and regions and different industries to promote international collaboration and consensus, and provide targeted guidance. Third, strengthen international cross-border cooperation and knowledge sharing, establish an international cooperation network on machine learning copyright issues, and jointly deal with the challenges brought by transnational cooperation and data flow. Fourth, professional bodies and arbitration mechanisms need to be established to specifically handle machine learning copyright disputes and disputes, and provide independent and impartial solutions. Finally, continuously monitor and evaluate developments and changes in machine learning copyright issues, and timely adjust solutions and legal policies to address new challenges and issues [4].

6. Conclusion

In the current era of large model competition, when all circles call for a balance between the development of artificial intelligence industry and copyright protection [5]. It is urgent to explore the optimal path to resolve the copyright risk in machine learning. In the process of AI training and development, China has also been strengthening the protection of personal data, and has formulated relevant laws and regulations such as the Personal Information Protection Law to ensure that the use of sensitive information such as

personal data in the field of generative AI complies with legal provisions, and protects users' privacy rights and interests. In the future, China's artificial intelligence industry has laid a good foundation for competition, machine learning and generative artificial intelligence.

References

- [1] Liang Zhiwen. "On the Legal Protection of Artificial Intelligence Creations." *Legal Science (Journal of Northwest University of Political Science and Law)*, No.5, 2017, 156-165
- [2] Shi Guanbin, "On the Copyright Protection of Intelligent Robot Creations", *Oriental Law*, No. 3, 2018.
- [3] Zhang Yujie, "On Robot Rights and its Risk Regulation in the Era of Artificial Intelligence", *Oriental Law*, No.6, 2017.
- [4] Zhu Chengbin, Li Long: "Artificial Intelligence as a legal fiction cannot have the specificity of biological human beings", *Journal of Shanghai Jiao Tong University*, No.6, 2018.
- [5] Xiong Qi: "The Copyright Identification of Artificial Intelligence-Generated Content", *Intellectual Property Rights*, issue 3, 2017