Scientific Research Publishing

Open Journal of Social Sciences, 2025, 13(8), 428-435
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

Legal Risks and Regulatory Pathways of Generative Artificial Intelligence

Siqi Ding

University of Manchester, Manchester, UK
Email: 1013436492@qq.com

How to cite this paper: Ding, S. Q. (2025). Legal Risks and Regulatory Pathways of Generative Artificial Intelligence. *Open Journal of Social Sciences, 13,* 428-435.
https://doi.org/10.4236/jss.2025.138028

Received: July 28, 2025
Accepted: August 16, 2025
Published: August 19 2025

Abstract

Generative artificial intelligence is advancing the iteration of artificial intelligence technology and expanding its application scenarios at an unprecedented speed, but the accompanying legal risks also highlight the tense game between technological innovation and legal protection. This article focuses on the core legal issues that generative artificial intelligence may cause, such as data privacy breaches, intellectual property infringement, and ethical dilemmas. It proposes countermeasures from the dimensions of building a data compliance regulatory system, improving intellectual property protection mechanisms, and perfecting diversified responsibility sharing models. The healthy development of artificial intelligence cannot be achieved without the joint efforts of the whole society. It is hoped that legislative bodies, regulatory departments, industry enterprises, and public opinion will form a joint force under the principles of "encouraging innovation, inclusiveness and prudence, and risk prevention", and use the light of the rule of law to lead technology towards goodness, and use generative AI to benefit humanity.

Keywords

Generative Artificial Intelligence, Legal Risks, Intellectual Property Right Burden-Sharing

1. Introduction

Generative artificial intelligence is one of the most prominent research directions in the field of artificial intelligence in recent years. This technology, through deep learning of massive amounts of data, can autonomously generate text, images, audio, and other content that is similar or even better than human creation, demonstrating broad application prospects in intelligent writing, assisted design, digital creativity, and other fields. However, with the advancement of technological ca-

pabilities, legal risks are becoming increasingly prominent. There are hidden issues such as privacy breaches, intellectual property infringement, and deep forgery in data collection, content generation, and ethical boundaries. It is urgent to clarify legal boundaries and balance innovation incentives with risk prevention and control. This not only tests the adaptability of the existing legal system, but also relates to the sustainable development of the artificial intelligence industry. The legal risks and regulatory challenges of generative artificial intelligence are not limited to specific countries or regions, but also manifest with distinct characteristics in different jurisdictions (such as China, the European Union, and the United States). This paper will focus on the development of AI legal frameworks in China, while also conducting a comparative analysis of relevant laws in the EU and the US to explore issues of universality.

2. The Core Legal Risks of Generative Artificial Intelligence

2.1. Data Privacy Breaches and Security Threats

There is a natural tension between the thirst for massive data and personal privacy protection in generative AI. On the one hand, model training requires large-scale and diverse data inputs, which is a key driving force for technological iteration. On the other hand, the aggregation of some data may infringe on personal information rights and even pose a risk of data abuse. Taking image generation models as an example, if facial photos are crawled from the network for training without sufficient desensitization, it is highly likely to infringe on citizens' portrait rights. The user behavior data required for text generation models, once collected beyond the necessary scope and improperly used, may also violate personal information protection regulations. Even more tricky is the challenge of massive training data to individuals' right to be forgotten. Even if users request the deletion of personal information in accordance with the law, it is technically difficult to completely remove it due to the highly abstract model parameters. The conflict between the rights of data subjects and the development of AI will become increasingly prominent.

Take the "Facebook Data Breach" as an example. Although this incident is not directly related to generative AI, the data privacy breach issue provides us with similar legal challenges. In 2018, Facebook revealed that the personal data of approximately 87 million users had been accessed by unauthorized third parties and used for political ad targeting. This incident triggered legal debates between user privacy rights and platform data usage. Although Facebook faced numerous lawsuits after the event, there is still no clear legal precedent on how to handle personal data involved in AI model training. Specifically, if an AI model collects facial images without authorization for training, it could face legal risks similar to those encountered by Facebook. Moreover, such data privacy issues often lack a unified cross-national legal framework, leading to varying interpretations of infringement across countries.

2.2. Intellectual Property Infringement and Ownership Disputes

The risks in the field of intellectual property mainly manifest in two aspects: the

copyright ownership of training data and the recognition of rights to generated works. As for the former, due to the huge demand for data in generative AI, some companies may neglect to respect the copyright of their works for the sake of convenient collection. Unauthorized use of copyrighted literary or artistic works by others for training undoubtedly constitutes infringement. Even if the data source is legal, it is not easy to determine the ownership of the generated work. Traditional intellectual property law regards human intellectual creation as a protected object, while intelligent generated content blurs the boundary between humans and machines on the creative subject. Taking AI painting as an example, the creativity of a painting comes from user input, and the creative process highly relies on algorithms. Both make substantial contributions to the formation of the work, making it difficult to simply define the ownership of rights (Merebashvili, 2025). There is still no consensus on whether new rights such as virtual character portrait rights and trademark rights can be applied to AI generated objects.

Take the 2018 "GAN Artworks Copyright Case" in the United States as an example. This case concerned whether AI-generated artworks should be attributed to the AI developer or the user of the tool. In this case, an artist used Generative Adversarial Network (GAN) technology to create an artwork and published it on a commercial platform. However, the focus of the case was whether the work could be granted copyright protection, as the creative agent involved was not only the artist but also the algorithmic system. The court ultimately ruled that while AI can assist in content generation, the final creative decisions were still made by the human artist, thus granting copyright to the artist. This case provides a reference point for determining the copyright ownership of AI-generated works, but it has not completely resolved the issue of the "originality" of AI-generated content, especially regarding the legal standards for AI-created works and ownership attribution.

2.3. Ambiguity of Responsibility Subject and Ethical Conflict

The determination of legal responsibility is a major challenge for generative AI governance. From the perspective of the composition of infringement, traditional attribution logic often relies on direct causal relationships between the parties involved. However, the illegal consequences of intelligent content generation may stem from joint negligence in multiple stages. For example, when AI outputs discriminatory statements based on biased data, data providers, algorithm developers, platform operators, end-users, and other entities may all be at fault. For example, there is no clear regulation in current law on whether providers of deepfake tools can be held accountable when users use AI "face swapping" technology to create obscene videos and spread them. In addition, the content generated by the algorithm exceeds the developer's expectations and the specific consequences cannot be foreseen, which also poses a dilemma for determining responsibility (Prinz, 2025). Therefore, it is urgent to clarify the liability boundaries of AI infringement within the existing legal framework and develop targeted attribution

rules.

In the 2018 "Deepfake Case," some users exploited AI technology to create and disseminate malicious fake videos involving false statements and defamation. Although technology providers (such as developers of deepfake tools) were not directly involved in content creation, the issue arose as to whether they should be held accountable. In this case, some video production companies failed to regulate the AI tools on their platforms appropriately, which allowed malicious users to use them to produce and distribute defamatory fake videos. While the development of such tools was not illegal, the law did not clearly define whether the tool providers should be held responsible for their users' illegal use. Ultimately, the court held these platforms accountable for failing to adequately regulate the AI-generated content and mandated them to strengthen their monitoring and review processes. This case highlights the uncertainty regarding the attribution of responsibility for AI technology and underscores the significant challenges faced by existing legal frameworks in keeping pace with the rapid development of AI.

3. Regulatory Strategies for Legal Risks of Generative Artificial Intelligence

3.1. Building a Data Compliance Regulatory System

Currently, there is a natural tension between the thirst for massive data and personal information protection in generative AI. On the one hand, model training requires large-scale and diverse data inputs, which becomes a key driving force for technical iteration. On the other hand, the aggregation of some data may violate personal privacy and even pose a risk of data abuse. It is necessary for regulatory authorities to develop differentiated data governance standards based on AI application scenarios. For the collection process, it is necessary to strictly implement the principle of user informed consent and eliminate the chaos of obtaining personal information, such as "riding on hot topics" and "killing acquaintances". In the usage process, data usage boundaries should be clearly defined for different business functions, and sensitive data must be desensitized before being used for AI training. The storage and transmission links should also be protected at different levels to reduce the risk of data leakage. To ensure the implementation of the system, it is necessary to establish a regular external audit mechanism and utilize third-party forces to conduct a full process evaluation of the compliance status of enterprise data (Corliss, 2025). At the same time, it is necessary to accelerate the construction of security assessment standards for cross-border data flow and maintain digital sovereignty security. Only by using the hands of effective systems to delineate the "red line" of data compliance and punishing data violations with the sword of law enforcement can we lay the foundation for the long-term development of generative AI. The EU Artificial Intelligence Act (2024) imposes strict requirements on data security and privacy, particularly for risk management in high-risk AI applications, mandating the use of transparent data governance standards.

3.2. Establishing a Sound Intellectual Property Protection Mechanism

The intellectual property challenges brought about by generative AI are mainly reflected in two aspects: the legitimacy of training data and the identification of ownership of generated content. For the former, although network data is inexhaustible, not all data can be freely included in the AI training set. Whether it is literary or artistic works, or code or patent information, once used for commercial training without authorization, it is highly likely to infringe on the legitimate rights and interests of the rights holder. Therefore, AI companies should attach great importance to the compliance review of data sources. For content involving intellectual property rights of others, they should obtain permission in advance or pay reasonable consideration in a timely manner after use. In the latter (Gong et al., 2025), which is the determination of ownership of generated content, the binary division logic of subject and object in traditional intellectual property law may be difficult to apply simply. From the essence of content generated by artificial intelligence, the degree of human creative input should be a key factor in determining the ownership of rights. If the generated works are mainly based on human creativity and reflect the substantial contributions of developers to core elements such as algorithms and parameters, then intellectual property rights should belong to individuals. On the contrary, if the work highly relies on intelligent algorithms to generate autonomously, demonstrating the "creativity" of the AI system itself, then giving the algorithms certain power may be more convincing. It is necessary to clarify the copyright standards for AI generated works as soon as possible in order to clarify the intellectual property boundaries under the human-machine game. Legislators should assess the situation and expand the scope of intellectual property objects, reserving institutional space for intelligent content generation.

The legitimacy controversy of training data as the "knowledge fuel" for AI systems stems from three dimensions: firstly, publicly available data on the internet does not necessarily constitute the category of "fair use". Although the Bern Convention stipulates that works in the public domain can be freely utilized, there are significant differences in the definition of the "public domain" in judicial practice among countries. The EU's Digital Single Market Directive requires commercial AI training to obtain copyright permission, while the flexibility of the US' fair use 'principle has sparked ongoing controversy. Secondly, there are technical blind spots in the ownership determination of structured data. The crawling behavior of professional databases such as code and patent literature may simultaneously touch multiple legal boundaries of copyright, trade secrets, and anti unfair competition. Finally, the annotation behavior during the data cleaning process may generate derivative rights. Whether the additional labor value of training data constitutes a new intellectual property object after manual screening, labeling, desensitization, and other processing urgently needs legal clarification. To solve the data compliance dilemma, a three-layer protection system needs to be established:

firstly, establish an AI enterprise data source list system, requiring enterprises to publicly disclose the channels for obtaining training data, processing procedures, and scope of use, and accept social supervision. Secondly, implement the "data usage license pool" mechanism, coordinate collective negotiations between copyright holders and AI companies through industry associations, and establish standardized authorization templates and fee systems. Thirdly, develop data compliance testing tools and use blockchain technology to achieve full chain traceability of data flow, ensuring that the legitimacy of each batch of training data can be verified. The U.S. Executive Order 14110 (2023) provides an initial exploration of the copyright ownership issues related to AI-generated content, proposing an intellectual property governance framework for the AI application domain.

3.3. Improving the Diversified Responsibility Sharing Model

The rapid development of generative AI technology has made the responsibility determination of data providers, algorithm developers, service deployment providers, end-users, and other entities in AI infringement more complex. To clarify the responsibilities of all parties, it is necessary to construct a diversified responsibility sharing model from multiple dimensions, such as legislation, judiciary, technology, and industry self-discipline.

One is at the legislative level, refining the determination of AI infringement liability: adding a special chapter on "Artificial Intelligence Generated Content" in the revision of the Copyright Law, clarifying the shift of originality standards from "human creativity" to "technology assisted creativity", and establishing a system for recording generated content, requiring a rights declaration before commercial use. At the same time (Hao & Zhang, 2025), referring to the Interim Measures for the Management of Generative Artificial Intelligence Services, specific provisions are made on the obligations of AI service providers such as data compliance and algorithm transparency.

The second is to promote the dynamic burden of proof at the judicial level: in AI infringement cases, the defendant is required to prove the legal source of its training data. For example, the Beijing Internet Court in the "AI Painting First Case" requires the defendant to submit a complete list of training data to ensure the operability of law application. By publishing typical cases through the Supreme People's Court, such as the AI generated content copyright case ruled by the Changshu Court in Jiangsu, the criteria for determining the copyrightability of AI generated content are clarified, providing rules and guidance for practice. At the same time, establish technical standards for "creating gene maps", analyze the creative path of generated content through algorithms, quantify the degree of human intervention, and provide technical support for judicial judgments.

Thirdly, at the technical level, a creative traceability mechanism should be established: drawing on the EU's "Generative AI Responsibility Directive", high-risk AI systems are required to have a built-in creative traceability module to record the creative path and degree of human intervention of the generated content, en-

suring the traceability of the generated content. Using blockchain technology to achieve full chain traceability of data flow, ensuring the legitimacy of training data sources. At the same time, establish a mechanism for recording annotation behavior during the data cleaning process, and record the additional labor value of data that has been manually screened, annotated, and desensitized, providing a basis for data ownership determination.

Fourthly, at the level of industry self-discipline, refine the reasons for exemption: Industry organizations such as the Data Rule of Law Research Institute of China University of Political Science and Law have released the "Artificial Intelligence Law (Scholar's Proposal Draft)", clarifying the regulatory system for key artificial intelligence and special application areas of artificial intelligence, and providing compliance guidance for AI enterprises(Gao & Qin, 2025). At the same time, establish a self regulatory mechanism for AI enterprises, such as OpenAI embedding a "copyright filter" in the GPT-4 model to actively block known copyrighted content. Enhance public legal awareness, strengthen publicity and education on the legal boundaries of AI applications through channels such as the Central Cyberspace Administration's "Clear and Rectify AI Technology Abuse" special action, and guide users to use AI technology legally; Guide users to use AI in compliance: As suggested by Wu Zifang, a lawyer at Beijing Rongtai Law Firm, users should choose legal and legitimate products when using AI, pay attention to the protection of others' copyright, portrait, and privacy, and avoid using AI to commit infringement. At the same time, establish a user credit record mechanism and impose credit penalties on users who have repeatedly committed infringement.

4. Limitations and Future Work

While this article provides an initial exploration of the legal risks and regulatory pathways of generative artificial intelligence, several limitations remain. First, the legal environment is rapidly evolving, and the existing legal frameworks may not fully address the emerging issues posed by generative AI, as the technology continues to evolve. Areas not covered in this paper, such as competition law and the potential antitrust concerns in AI development, as well as cross-border enforcement challenges, require further examination. The complexities of cross-border data flow and conflicts between legal systems of different countries add layers of difficulty to regulation. Additionally, the widespread adoption of generative AI could lead to more complex ethical dilemmas, for which traditional legal frameworks may not be well-suited. Therefore, future research should focus on legal adaptability and cross-sector coordination, dynamically adjusting legal policies to ensure a balance between technological advancement and legal protection.

5. Conclusion

While generative artificial intelligence creates new momentum for the digital economy, its legal risks cannot be ignored. These risks have strong technical attributes, complex ethical connotations, and far-reaching social impacts, and require a multi

pronged approach and collaborative governance. Guided by the principles of "encouraging innovation, inclusiveness and prudence, risk prevention, and legal supervision", we should actively improve institutional supply, strengthen key technology research and development, enhance industry self-discipline, and create a healthy ecosystem. Managers should follow the path of "rule of law, refinement, and specialization" in supervision, clarify red lines, set bottom lines, and provide basic guidance for the development of AI. Enterprises need to strengthen technical internal control, embed compliance requirements into algorithms, and build a strong security defense line. The public should also enhance their digital literacy and approach AI applications with an inclusive and cautious mindset.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Corliss, M. C. (2025). Redefining Rights: The Case for Limited AI Legal Personhood in Intellectual Property and Genetic Engineering. *The Undergraduate Law Review at UC San Diego, 3,* 252-270. https://doi.org/10.5070/lr3.47479

Gao, M., & Qin, P. (2025). The Five Major Criminal Legal Risks of Generative Artificial Intelligence. *Legal Person, 4,* 23-25.

Gong, X., Yang, Z., Li, Q. et al. (2025). Generative Artificial Intelligence Risk Identification and Countermeasures. *Cooperative Economy and Technology, 11,* 174-176.

Hao, W., & Zhang, Q. (2025). The Regulatory Dilemma and Legal Response to the Risks of Generative Artificial Intelligence. *Hebei Legal Vocational Education, 3,* 111-116.

Merebashvili, T. (2025). Generative Artificial Intelligence: New Dilemmas for Intellectual Property Law. *Socrates. Rīga Stradiņš University Faculty of Law Electronic Scientific Journal of Law, 1,* 80-84. https://doi.org/10.25143/socr.31.2025.1.80-84

Prinz, K. D. (2025). Managing the Legal Risks of Artificial Intelligence on Intellectual Property and Confidential Information. *Consulting Psychology Journal, 77,* 169-179. https://doi.org/10.1037/cpb0000287