

Risks and Legal Governance of Generative Artificial Intelligence

Hanpu Sun

College of LAW, Beijing Normal University, Beijing, China

ABSTRACT

Generative Artificial Intelligence (AI) represents a significant advancement in the field of artificial intelligence, characterized by its ability to autonomously generate original content by learning from existing data. Unlike traditional decision-based AI, which primarily aids in decision-making by analyzing data, generative AI can create new texts, images, music, and more, showcasing its immense potential across various domains. However, this technology also presents substantial risks, including data security threats, privacy violations, algorithmic biases, and the dissemination of false information. Addressing these challenges requires a multi-faceted approach involving technical measures, ethical considerations, and robust legal frameworks. This paper explores the evolution and capabilities of generative AI, outlines the associated risks, and discusses the regulatory and legal mechanisms needed to mitigate these risks. By emphasizing transparency, accountability, and ethical responsibility, we aim to ensure that generative AI contributes positively to society while safeguarding against its potential harms.

KEYWORDS

Generative Artificial Intelligence; Legal Liability; Risk Regulation.

1. INTRODUCTION

Generative AI is a class of AI systems capable of autonomously generating original content by learning the features and patterns of existing data. Before the birth of generative AI, AI models were mainly dominated by decision-based AI. Decision-based AI processes, analyzes, and outputs autonomous decision-making results based on input data. Typical application scenarios for this type of AI include algorithmic recommendation and autonomous driving. For example, algorithmic recommendation systems can recommend content that may be of interest to users based on their historical behavior and preferences; autonomous driving technology realizes autonomous driving of vehicles by sensing and analyzing the environment around the vehicle. Decision-based AI aims to achieve preliminary analytical functions and has existed for a long time as an auxiliary tool for human productive activities, helping humans make faster and more accurate decisions.

However, with the continuous growth of data volume and the improvement of computing power, decision-based AI gradually exposes its limitations. The main problem is that decision-making AI can only analyze and process existing data and lacks the ability to generate original content, while the emergence of generative AI makes up for this deficiency. Generative AI can not only analyze existing data, but also generate new content autonomously based on learned features and laws, and this ability makes generative AI show great potential in many fields.

At present, generative AI has become the technical key to the information age, empowering the development of different industries as well as different fields, and becoming the underlying support for the digital economy and industrial revolution. Generative AI can directly contribute to the process

of articulating arguments.[1] In news writing, literary creation, and advertising copywriting, generative AI can automatically generate well-structured and content-rich articles as long as keywords or topics are inputted. The GPT-4 model developed by OpenAI can generate smooth and logical texts based on input prompts, and it is widely used in writing assistants and content creation. This not only greatly improves the efficiency of content production, but also breaks through the limitations of human creators' thinking. In the field of visual arts, generative AI can create realistic images and artworks. By learning from a large amount of image data, these models can generate paintings that are similar in style to those of human artists, or even create entirely new art styles. Generative AI has also shown strong capabilities in areas such as music composition and game design. In music composition, generative AI can create musical works in a variety of styles, providing inspiration and material for music producers. In game design, generative AI can automatically generate complex game scenes and plots, enhancing the diversity and playability of games.

The emergence of generative AI marks a new leap in AI technology. Compared with traditional decision-making AI, generative AI can not only assist humans in decision-making, but also create brand new content on its own. This increase in access and information availability changed the dynamics of inventiveness and creativity.[2] With the continuous development of technology and the expansion of application scenarios, generative AI will surely fulfill its potential in more fields and create more value for human beings. This technological progress not only promotes the development of various industries, but also brings infinite possibilities for future AI applications.

2. RISKS OF GENERATIVE ARTIFICIAL INTELLIGENCE

Generative AI is facing many issues such as data security, privacy protection, algorithmic bias, and dissemination of false information along with its rapid development. Therefore, when applying generative AI, its risks and challenges need to be handled with care by taking appropriate measures to protect aspects such as user privacy, trade secrets, and technological security, as well as by complying with the relevant legal and ethical norms.

2.1. Data Security

Generative Artificial Intelligence is a technology that relies on large-scale data sets for training. In order to realize their powerful generative capabilities, these systems need to obtain large amounts of data from various sources, including user inputs, Internet resources, business data, and so on. These data contain a large amount of personal information and trade secrets, such as user's behavioral records, conversation content, and business data.

Against the backdrop of increasing competition in the generative AI market, many providers often choose to take a quick and easy approach to data collection and processing to gain a head start in technology development. While this can accumulate a large amount of data in a short period of time and drive rapid model iteration and performance improvement, it also sows serious data security risks. Many enterprises have imperfect data protection measures in the process of data collection, storage and use.

Specifically, data security issues are reflected in all stages of data collection, storage and use. In the data collection stage, some enterprises extensively collect user data without fully informing users, failing to protect users' right to informed consent. In the data storage stage, inadequate data protection measures can result in data that may be accessed or stolen by unauthorized personnel. In the data use stage, the generative AI model lacks transparency and standardization in the process of processing and analyzing data, and there is a risk of data misuse.

The existence of these problems leads to a significant increase in the risk of data leakage and misuse. Once data leakage occurs, users' personal privacy and business secrets may be disclosed or used for illegal purposes, bringing a series of negative impacts. For individual users, the leakage of private

data may lead to problems such as identity impersonation, property loss, and damage to social relationships. For enterprises, leakage of trade secrets may lead to competitive disadvantages, loss of market share, damage to goodwill and other serious consequences.

2.2. Rights and Interests of Users

Data collection and use without consent will violate users' personal information rights and interests. In terms of privacy protection, generative AI models require a large amount of data for training, and if this data is collected without the consent of the user, it may violate privacy protection and other relevant laws. For example, the EU's General Data Protection Regulation (GDPR) clearly states that data processing must be based on legitimate user consent and that users have the right to know how and for what purpose their data is being used.

Transparency means that the data processing process should be open and transparent so that users can understand and monitor it. For generative AI models, transparency means that information about the model's training data, algorithms, decision-making process, and generated content should be interpretable and understandable. Users have the right to know how their data is being used, and transparency in the operation of generative AI models is essential to protect user privacy. Data needs to be used in a manner consistent with privacy protection, and unauthorized data processing, including storage, analysis, sharing, and reuse of data, may involve privacy violations.

Based on the principle of data minimization, the amount of data collected should be minimized during data collection and processing, and only the minimum amount of data necessary to achieve a particular purpose should be collected and processed. By limiting data collection and processing, unnecessary data exposure can be reduced and user privacy protection can be enhanced. Generative AI models require large amounts of data for training, a requirement that may conflict with the data minimization principle. For generative AI, more data usually improves the accuracy and generation quality of the model, and in order for generative AI models to be broadly adaptable and generative, the training data needs to be diverse and representative, which means that a large amount of different types of data need to be collected from a variety of sources.

Data subject rights include rights of access and deletion, and rights of rectification, and generative AI systems need to provide mechanisms to support these rights. Users have the right to access their data and to request that it be deleted and errors in the data be corrected. The right of access refers to the right of users to know and access their personal data that has been collected and stored. The right to erasure refers to the right of the user to demand that their personal data be deleted, typically when the data are no longer necessary, the user withdraws consent, the data processing is unlawful or the user objects to the data processing. The right to rectification refers to the right of the user to have his/her inaccurate or incomplete personal data corrected to ensure the accuracy and completeness of the data. The realization of these rights in generative AI models may involve complex data processing and modification.

2.3. Algorithmic Bias

Generative AI systems typically learn and make inferences based on pre-existing data, but if that data is biased or discriminatory, then the generated intelligence may also be biased or discriminatory. Generative AI generates algorithmic bias. Algorithmic bias is a systematic bias in the output of an AI system towards certain groups or specific situations due to the data, algorithms, or the design itself during training, evaluation, or real-world applications. This bias may appear in various forms at various stages.

The first category, data bias. Data bias is primarily reflected in the underrepresentation of training data. When the dataset used to train generative AI does not adequately represent all relevant groups or situations, then the model may be biased against certain groups or situations in practice. For

example, if a language model is primarily trained using textual data from a particular geographic region or culture, it may lack an accurate understanding of the linguistic habits and expressions of other cultures. Also, data from different sources may contain different biases. Data from online media may be more colloquial and tendentious, while data from academic papers may be more formal and neutral. The biases in these training data are inherited and amplified by the model, and if the training data contains discriminatory content, the model may reflect or reproduce these biases when generating content.

The second category, algorithmic and modeling biases. In model selection, model architectures or algorithms may exhibit specific biases when dealing with specific types of data. Different neural network models may be more sensitive to certain features, thus unbalancing the amplification of these features when generating content. In addition, the choice of optimization objective and loss function of an algorithm may also introduce biases, and the excessive pursuit of accuracy may result in a lack of fairness. The selection and regulation of hyperparameters during model training may introduce bias. If the goal of hyperparameter regulation is based on some specific performance metrics that are inherently unfair to certain groups, the final performance of the model may also be biased.

The third category, application and feedback bias. Generative AI algorithms are capable of creating various forms of new content through its utilization of recognizable patterns over large data sets.[3] This type of bias mainly lies in the content aspect, where generative AI may show different biases in different application scenarios, and user feedback on generated content may also affect the further optimization and adjustment of the model.

2.4. False Information

The nature of artificial intelligence is to make it close to human behavior and way of thinking, which will bring about ethical and moral issues of generated content. Since generated content is related to arithmetic power, algorithms, and depth of learning, there is no way to ensure the accuracy of the generated content, creating the issue of generating content credibility. Generative AI systems may generate false information, harmful and misleading content, etc. A major feature of AI that distinguishes it from information technology is the strong human-computer interaction, and generative AI may have many implications. Generative AI can also be used for deep forgery, creating highly realistic fake video or audio, which could be used for fraud or to damage someone's reputation.

A major advantage of generative AI is its ability to generate high-quality text, images, audio, and video content. However, this powerful capability also raises a number of ethical and moral issues. Because generative AI relies on large amounts of training data and complex algorithms, the accuracy of the content it generates cannot be fully guaranteed. When generating textual content, models may generate biased or erroneous information based on their training data or model parameters, leading to a decrease in user trust in the information. More seriously, generative AI may be used maliciously to intentionally generate false information or misleading content, which not only affects the judgment of individuals, but may also have a negative impact on social stability.

Another distinctive feature of generative AI is its powerful human-computer interaction capability. Unlike traditional information technology, generative AI is able to simulate human language and behavior, making users feel more natural and real when interacting with the system. However, this strong interactivity also brings new challenges. In the process of interacting with AI systems, users' ideologies, value judgments, learning contents, and consumption habits may be subconsciously influenced, and false information may have a more negative impact. Generative AI systems inadvertently disseminate specific values or biases in their interactions with users, who may unconsciously accept these ideas and influence their personal decisions and behaviors.

In addition, generative AI can be used in deepfake techniques to create highly realistic fake videos or audio. The misuse of such techniques can have serious consequences. With deepfake, an attacker can create fake video or audio content to deceive the public or damage someone's reputation. For example,

forging a video of a politician's speech and spreading false information could have a significant impact on society. Deep forgery techniques can also be used in cyber fraud to create false authentication material, which can lead to illegal activities such as economic fraud. An AI trained on copyrighted works might progress to the point at which it becomes deceptive and power seeking, surpasses human intelligence, and poses a substantial risk to humanity.[4]

3. THE CHALLENGES OF GENERATIVE AI REGULATION

3.1. Algorithmic Regulatory System

Due to the complexity and novelty of generative AI technology, the existing regulatory system may face many uncertainties in its application. At present, the algorithmic standing regulatory system has not yet been fully established, such as the algorithmic assessment system to address risks and the algorithmic audit system for routine review, and this extensive standing regulation can also be applied to generative AI. The establishment of the algorithmic regulatory system needs to be continuously improved and adjusted with the development of technology and changes in application scenarios, and algorithmic regulation involves a number of fields and sectors, requiring cross-sectoral collaboration and joint efforts. With the continuous development of technology and the continuous improvement of the regulatory system, algorithm regulation will become more and more standardized and normalized.

The purpose of algorithm assessment is to establish a risk-adaptive algorithm governance system to ensure that algorithms can operate safely, fairly, and effectively in different application scenarios. Establishing a risk framework for analyzing and assessing algorithmic applications requires clear assessment indicators and assessment methods, which are customized based on industry standards and best practices and combined with specific application scenarios. In turn, algorithmic applications are comprehensively assessed, potential security vulnerabilities and risk points are identified, and possible negative impacts are analyzed and evaluated in depth. The regulator or third party summarizes and classifies the identified risks, prepares a risk assessment report, and provides explanations and interpretations to the management and relevant departments so that they can understand the risk profile of the algorithm and take appropriate management measures. The supervisee can conduct self-regulation based on the assessment results, determine corresponding risk management measures, such as strengthening monitoring, modifying algorithms, redesigning systems, etc., and follow up and evaluate the implementation of risk management measures to ensure the effectiveness of the measures. Regulators can strengthen the supervision and inspection of algorithm application based on the assessment to ensure that the algorithm operates in accordance with the established rules and standards, and that problems in the operation of the algorithm are detected and corrected in a timely manner to prevent the risk from expanding.

The algorithm audit system is a comprehensive review mechanism for the algorithm system and its operational activities, aiming at supervising the operation of algorithmic power, preventing digital risks and controlling algorithmic alienation through technical and non-technical measures. The system collects data on the performance of algorithms in specific application scenarios and evaluates their impact on people's interests, so as to judge whether the algorithms are legally compliant and whether they meet the requirements of fairness and transparency. The algorithm audit system is a powerful lever to synergize and articulate enterprise self-regulation and government regulation, which can not only make up for the governance conflict between algorithm transparency and proprietary technology protection, but also promote the dual-track synergy between enterprise self-regulation and government regulation, and promote the effective implementation of the triadic governance mechanism of law, technology and ethics. In the auditing process, the algorithm audit system focuses on the principle of dynamic auditing, taking the whole life cycle of the algorithm as a chain, and adopting different auditing paths in a classified and hierarchical manner, such as code

auditing, capture auditing, vest auditing, collaborative auditing and non-intrusive auditing, etc. At the same time, the algorithm audit system is also compatible with voluntary auditing. At the same time, the algorithm audit system is also compatible with voluntary and mandatory auditing, which on the one hand meets the diversified needs of enterprises for differentiated management and risk control of algorithmic systems, and on the other hand reserves a system interface for the government to intervene in the supervision of algorithms.

In the absence of a clear regulatory regime, companies may reduce their investment and innovation in generative AI for fear of compliance risks, which is not conducive to the continued development and application promotion of the technology. The imperfect regulatory system for algorithmic normalization is also not conducive to the regulator's basic risk prevention and control of generative AI, which may result in risk amplification.

3.2. New technology Regulatory

The emergence of generative AI as a new technology significantly increases the difficulty of regulation. For example, generative AI services show differences in subject diversity, and the emergence of generative AI generality makes regulatory coordination more complex.

Generative AI services weaken the boundary between technology and platform, the big model developer and the application provider may be the same subject, after the downstream application has been developed, the two will be gradually differentiated, accordingly, laws and regulations should pay attention to the distinction in terms of the responsibility of the subject. At the early stage of the birth of a new technology, in order to quickly verify its value and feasibility, developers tend to apply it directly in the environment they can control, thus accelerating the process from technology development to application landing. Developers are not only the creators of the technology, but also the pioneers of the application. They form a high degree of fusion between technology and application by continuously iterating and optimizing the model while exploring its application potential in different scenarios. This fusion helps provide rapid feedback on the effectiveness of the technology and promotes its rapid progress and maturity. As the technology matures and the market expands, generative AI services begin to attract more application developers and service providers to join. These newly joined players may not have the ability to develop big models directly, but they are good at applying existing big model technologies to a variety of specific business scenarios, such as content creation, customer service, edutainment, and so on. At this point, the line between big model developers and application providers begins to become clearer. Developers focus on continuous innovation of the technology and optimization and upgrading of the model, while application providers focus on transforming the technology into actual products and services to meet the market demand.

Generative AI has achieved generalization to a small extent in language, video, etc., away from the constraints of the strong scenario-based compartmentalization of AI, and it may be difficult to apply the scenario-based governance approach of AI in the traditional sense to generative AI governance. Generative AI, especially in areas such as language and video, has demonstrated a certain degree of generalization, which means that these technologies are no longer limited to specific, highly scenario-specific applications, but are able to span multiple domains and scenarios to provide diverse services and solutions. For example, a well-trained large-scale language model can be applied to a variety of scenarios such as chatbots, content creation, text summarization, and so on, without the need for specialized training or tuning for each scenario. Traditional AI technologies often need to be customized and optimized for specific application scenarios, the so-called "strong scenario compartmentalization". However, the versatility of generative AI makes such scenario-based compartmentalization less obvious or necessary, allowing it to exhibit similar performance in different scenarios and adapt to new and unknown application requirements. Traditional governance approaches tend to focus on regulating and managing technology applications in specific scenarios,

such as in face recognition and autonomous driving. However, in the context of generative AI, the corresponding governance measures may need to be more flexible and comprehensively considered to adapt to the trend of evolving technologies and expanding applications.

4. LEGAL GOVERNANCE OF GENERATIVE ARTIFICIAL INTELLIGENCE

With the rapid development of generative artificial intelligence technology, the original legal governance model shows certain deficiencies. The complexity of generative artificial intelligence technology and the specificity of the data processing method have increased the difficulty of regulation, and the regulatory authorities need to combine the technical characteristics of generative artificial intelligence to develop more effective regulatory measures for governance. Regardless of where one stands on the future use of generative AI, what is clear is that the legal profession must put some standards in place immediately to avoid injury and unfairness on both an individual and institutional level.[5]

4.1. Risk-preventive Legal Regulation

In order to guarantee the data security of generative AI and reduce the algorithmic bias in generative AI, the regulatory authorities should formulate and improve the laws and regulations for generative AI, clarify the regulation of data use, privacy protection, algorithmic bias, etc., and regulate the aspects of data, models and generated content in combination with the data life cycle.

In terms of data sources, the legitimacy of data sources, diversity and representativeness of datasets should be ensured. Generative AI service providers need to obtain explicit consent from users when collecting and using their data, and provide users with detailed information about the use of data, clearly informing them of the purpose and scope of data use.

Strict privacy protection and data security measures should be introduced during data processing. Cleaning and pre-processing of data before training, completing data anonymization, improving the quality of data annotation, and removing significant bias and discriminatory content. Technically, models can be additionally built to review, grade and filter data and information, remove sensitive and harmful information, and dynamically track all data throughout the process to avoid model parameter amplification effects. Strengthen security measures during data storage and transmission, and adopt advanced encryption technology to prevent unauthorized access and theft of data.

To address the possible risks in model training, it should be stipulated that data that has undergone standardized processing and qualified assessment should be used in the training and use of generative AI models to reduce the risk of privacy leakage. Model developers should focus on the diversity and representativeness of the training dataset, try to cover all relevant groups and situations, and also build development teams containing diverse members to look at the models and data from different perspectives to reduce the introduction of bias. In addition, a model evaluation mechanism needs to be established to introduce fairness metrics to ensure that the model not only excels in performance but also meets the requirements in terms of fairness.

In terms of protecting the rights of data subjects, generative AI service providers should improve data deletion and correction mechanisms. For example, provide easy-to-use channels that allow users to submit requests for data deletion or data correction. Implement efficient data erasure and update mechanisms to ensure that users' data are completely deleted from all storage and backup systems after requesting deletion, and process users' correction requests in a timely manner to ensure data accuracy. Send confirmation notifications to users after data deletion so that they are aware that their data have been successfully deleted and that the corrected data are accurate.

4.2. Enhancing Transparency, Credibility, and Interpretability

Enhancing the transparency, credibility and interpretability of generative AI systems is primarily aimed at enabling users to understand their workings and potential risks, while clarifying the responsibilities and obligations of service providers. The required way to achieve this lies in the generative AI service provider's obligation to explain, label, and accept external supervision, which requires both technical support and sound policies and systems.

First, generative AI service providers need to fulfill their disclosure obligations. On the one hand, generative AI service providers need to fulfill the necessary obligations to explain the use of their products, which can be done by adopting transparent algorithms, open-source technologies, and user interface prompts to introduce the methods of use and hint at the relevant operational risks, so as to dispel the public's concerns about generative AI, improve the users' AI application capabilities, and enhance their digital literacy. On the other hand, to disclose specific decision-making processes, especially parameters and considerations that may be biased, visualization tools for data and model decision-making processes can be developed to help users understand the model's internal working mechanism and decision-making logic, or transparent reports can be released to explain in detail the model's training and operation process. In addition, a user feedback mechanism is established to collect and respond to the problems and suggestions encountered by users in the process of using the model, so as to enhance user satisfaction by continuously optimizing and improving the model.

Second, generative AI service providers need to label the generated content. The credibility and transparency of generated content can be effectively improved by clearly labeling information such as the source of the content, the method of generation, the accuracy and authenticity, and the restrictions and responsibilities of use. The use of prominent marking suggests that the subject of generation of the relevant content, the source of information if publicly available online can reasonably disclose the basis for generation or the source of information, if there is no specific source should be the accuracy of the statement, marking the content is purely fictional or there is a risk of misrecognition, and is for reference only. Labels that are difficult to remove can also be added to avoid further dissemination and misinformation of the generated content.

Finally, the transparency, credibility, and interpretability of generative AI can be enhanced by building an auditing system for data, models, and algorithms. Regulatory authorities or third-party organizations can investigate the data sources and data processing methods used during model training, as well as the important operations during model training, testing, and reasoning, assess the stability and robustness of model performance, provide audit opinions on the safety and compliance of data and models, and publicly disclose the audit results to the public in accordance with laws, regulations, and technical standards, in order to enhance the interpretability of the system, and improve users' trust in generative AI.

4.3. Clarifying Accountability Mechanisms

Generative AI is a multi-party technology, and the responsibilities of different participants such as data providers, model developers, and applicators need to be clearly delineated.

Data providers should ensure that the data they provide are legitimate, accurate, and complete, and clearly communicate the purpose and scope of data use to avoid data misuse or leakage. If the data leakage is caused by the service provider's system security vulnerability, mismanagement or intentional leakage, then the service provider should bear the main responsibility. If the service provider stores the data with a third-party service provider and the third-party service provider fails to fulfill its security protection obligations resulting in data leakage, then the third-party service provider shall also be held liable.

Technology developers should ensure that the generative AI technologies they develop comply with the requirements of laws, regulations and industry norms, and should not develop technologies that

are illegal, harmful or potentially risky. Safeguard technical security, prevent data leakage, respect user privacy, minimize information collection and anonymize it, ensure that generated content is authentic and legal, and avoid misleading. Additionally, developers should consider ethical considerations, promote social justice, and actively participate in the popularization of technology education. Developers also need to accept regulation, be responsible for damage caused by the technology, and continuously improve and optimize the technology to promote its healthy development.

The responsibility of the applicant to use mainly lies in the legal and compliant use of generative AI services, and shall not utilize the technology to engage in illegal activities or infringe upon the rights and interests of others. When using generative AI services, the authenticity and accuracy of the generated content should be carefully assessed to avoid misleading others or causing adverse consequences. It has the right to supervise and provide feedback on the quality of the generative AI services, and to promote continuous improvement and optimization of the services by the service providers.

5. CONCLUSION

Generative AI offers vast potential and transformative capabilities, but also presents significant challenges and risks. Addressing these problems requires a combination of technical protocols, ethical considerations and robust legal frameworks. Balancing innovation with regulation is the key to harnessing the benefits of generative AI while mitigating its risks.

As generative AI continues to evolve, sustainable cooperation between service providers, regulators, and society at large will be essential. By fostering an environment of transparency, accountability, and ethical responsibility, we can ensure that generative AI contributes positively to the advancement of human knowledge and well-being.

REFERENCES

- [1] John Villasenor. "Generative Artificial Intelligence and the Practice of Law: Impact, Opportunities, and Risks", *Minnesota Journal of Law, Science and Technology*, vol. 25, no. Symposium Issue, pp. 25-48, 2024.
- [2] Raina Haque, Simone Rose and Nick DeSetto. "The Non-Obvious Razor & Generative AI", *North Carolina Journal of Law & Technology*, vol. 25, no. 3, pp. 399-446. 2024.
- [3] Sadie O'Connor. "Generative AI." *Georgetown Law Technology Review*, vol. 8, no. 2, pp. 394-404. 2024.
- [4] Matthew Sag. "Fairness and Fair Use in Generative AI", *Fordham Law Review*, vol. 92, no. 5, pp. 1887-1922, 2024.
- [5] S. I. Strong. "Rage against the Machine: Who Is Responsible for Regulating Generative Artificial Intelligence in Domestic and Cross-Border Litigation?", *University of Illinois Law Review Online*, 2023, pp. 165-178, 2023.