WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 2, Number 2, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v2n2.41
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Research on Generative Artificial Intelligence Management

Shuheng Wang

Hebei University, Baoding, 071000, China

ABSTRACT

As a new technology, generative artificial intelligence technology not only promotes technological change, but also brings many governance risks because of its unique technical logic. For example, the introduction of massive data sets in the training process of generative artificial intelligence has the risk of data infringement, and the reinforcement learning model based on human feedback may exacerbate the generation and dissemination of false information, thus bringing a series of risks. Although the current law of our country has formed a preliminary framework for the regulation of generative artificial intelligence, there are still difficulties. Based on the development status of generative AI, this paper briefly discusses the practical problems and challenges of generative AI governance, the current governance framework and policies, and the future governance measures.

KEYWORDS

Generative Artificial Intelligence; AI; AI management; Data security; Data infringement

1. INTRODUCTION

The Interim Measures for the Management of Generative Artificial Intelligence Services officially issued by the National Cyberspace Administration and other seven departments will come into effect on August 15, 2023, aiming to promote the healthy development and standardized application of generative artificial intelligence, and safeguard national security and social and public interests. As a key technology leading a new round of scientific and technological revolution and industrial transformation, generative artificial intelligence represented by ChatGPT has continuously spawned new scenarios, new formats, new models and new markets, changed the production mode of information and knowledge, reshaped the interaction mode between human and technology, and had a significant impact on industries such as education, finance, media and games. In this context, countries around the world have introduced artificial intelligence development strategies and specific policies to seize the strategic commanding heights. At the same time, security risks such as data leakage, false content generation, and improper use exposed by generative artificial intelligence have also attracted widespread attention from various countries. It is safe to say that the development, application and governance of generative AI is no longer a common challenge for a single country, but for the entire international community.

2. RESEARCH BACKGROUND

This section describes the potential impact of the development of prudent prevention technologies. Generative AI is different from traditional discriminative AI, whose primary task is to classify or label existing data. In contrast, generative AI is more focused on models learning the distribution of data and generating new data. Chat GPT is the crystallization of modern technology wisdom, relying on the Internet to apply to various industries, with fast speed and simple operation to bring

convenience to human life, but its own uncertainty also brings potential legal and ethical risks. Whether generative artificial intelligence creations are subject to copyright law should also be distinguished according to different forms of expression outside the creation, and the technology itself should also clarify ethical requirements in order to respond to the legislative intent of lawmakers and serve the original intention of industry development and cultural creation and dissemination.

3. PROBLEMS AND CHALLENGES

3.1. Social security and ethical risks from technology

By analyzing the input training samples and then imitating to generate the corresponding content, generative artificial intelligence achieves the effect of fake and real. Unlike simple PS, generative AI also learns biometrics such as microexpressions and speech of the target object. For example, after stars are banned, film and television crews save costs through AI face changing technology. And old photos come back to life, giving people a chance to talk to their dead loved ones. The use of generative artificial intelligence to produce highly realistic face replacement as the representative of deep forgery technology plays an important role in autonomous driving, medical diagnosis and entertainment education, but its flexible and difficult to distinguish effect once abused will pose a threat to social and even national security. If users with ulterior motives maliciously promote extreme content, it will not only bring wrong values, but also undoubtedly cause social panic and lead to social risks.

And the terrifying impact of generative AI is not limited to this, it is gradually evolving into a new public opinion war between countries. The United States' 2019 Global Threat Assessment has already mentioned that this technology poses a threat to the national security of the United States and is an influence infiltration campaign against the cooperation of Allies. In 2017, the official Twitter account of the state of Qatar was hacked, spreading false statements by the Qatari head of state about Iran and Islam, causing anger in the Gulf neighbors, and the incident was also related to the subsequent diplomatic crisis in the Middle East. In the report "Weaponization of deep forgery Technology" of the Australian Strategic Policy Institute, the public opinion attack and defense network information operations are detailed. Through automatic comments, a large number of users' unconscious actions are triggered, which not only reduces the cost of claiming, saves time, but also expands the scale of participating users. In particular, such fake news takes time to expose, adding unstable factors to social security.

3.2. Risk of judicial judgment function in protection of creative objects

In terms of law, generative artificial intelligence mimics the deep learning of the human brain, and the content produced has ambiguity on the boundary, who "pays" for the output content, and the copyright disputes caused by artificial intelligence bear the brunt. Two representative cases in judicial practice are analyzed.

The first case is Filin v. Baidu. The cause is that Beijing Filin Law Firm obtained a judicial data analysis report of film and television entertainment industry after setting search conditions through the data analysis software of Wolters Kluwer, and then uploaded the article to the wechat public platform combined with its own conclusions. Baijia, a website run by Baidu, deletes some of the content and bylines of articles and publishes them on the platform. Filin Law Firm is claiming economic losses on the grounds that its right of authorship, the right to protect the integrity of the work and the right of information network transmission are damaged. In the end, the court added that the work was not copyrighted, but could not be freely used.

The other is Tencent's lawsuit against Shanghai Yingxun Technology for copyright infringement, which is a financial article automatically written by Tencent through artificial intelligence Dreamwriter and marked at the end of the article. Yingxun Technology published it on its website

without permission, and Tencent claims to enjoy the copyright and property rights of the article involved. But in this case, the court found that generative AI creations fall within the category of works. The two cases were also generative AI creations, but had diametrically opposite verdicts. As for the academic and practical circles, there are different views on whether artificial intelligence products are works under copyright law. Whether artificial intelligence creations belong to the works stipulated by the copyright law, and whether artificial intelligence can obtain relevant rights as an independent subject, it is obvious that the unclear protection of legal creations makes it difficult for judicial judgments to maintain consistency and play the guiding function of judgments.

4. CURRENT GOVERNANCE FRAMEWORK AND POLICIES

4.1. Ai governance efforts beyond countries

Many international organizations have begun to pay attention to the ethical and governance issues of AI, and have taken a number of actions and initiatives to promote the development and implementation of relevant policies. For example, the United Nations Educational, Scientific and Cultural Organization (UNESCO) has developed the AI Principles to promote human rights, justice, and sustainable development. In addition, the European Commission has published the European Data Strategy and the White Paper on Artificial Intelligence, which set out a set of frameworks and principles for AI governance. These initiatives aim to harmonize the international community's common understanding and action on AI, and promote cross-border cooperation to address cross-border ethical and legal issues.

As the development and application of AI involves global companies, many multinational companies are also playing a key role in AI governance. Some companies have already taken a proactive stance, developing AI ethics codes and committing to compliance with relevant ethics and regulations. At the same time, some industry leaders are also actively participating in the development of international standards to ensure that the design and application of AI meet global standards. This proactive engagement not only helps strengthen ethical measures within companies, but also drives a shared focus on AI governance across the industry.

4.2. Policies and regulations within countries

The legal framework for generative AI varies from country to country, but many countries are working to develop regulations to regulate and govern the development of AI technologies.

4.2.1. Eu artificial intelligence legislation

Safety first and fairness first. From the perspective of the legislative process, in April 2021, the European Commission issued the legislative proposal "Regulation of the European Parliament and the Council on the formulation of uniform rules on artificial Intelligence (Artificial Intelligence Law) and the amendment of Certain EU legislation" (hereinafter referred to as the "Artificial Intelligence Act"), which opened the "hard law" road of artificial intelligence governance. The final version of the compromise draft of the Artificial Intelligence Act was formed in December 2022. In June 2023, the European Parliament adopted the draft negotiating mandate for the Artificial Intelligence Act and amended the original proposal. On December 8, 2023, the European Parliament, the European Council and the European Commission reached an agreement on the Artificial Intelligence Act, which provides for comprehensive regulation of the field of artificial intelligence. Overall, the AI Act establishes an ethical and legal framework for the development and use of AI in the EU, and is complemented by the AI Responsibility Directive to ensure its implementation.

4.2.2. Artificial intelligence legislation in the United States

It emphasizes self-regulation and supports technological innovation. In the global context of AI law and policy making, the United States is gradually developing a regulatory framework based on voluntary principles. One of the more comprehensive regulatory initiatives in the United States is the Bill of Rights Blueprint released by the White House Office of Science and Technology Policy (OSTP) in October 2022, which aims to support the protection of civil rights in the design, deployment, and governance of automated systems. Under the guidance of the Bill of Rights Blueprint, federal departments began to perform their respective roles and began to develop specific policies, such as the United States Department of Labor developed the "Fair Artificial Intelligence Action Manual" to avoid AI bias against job seekers and employees based on race, age, gender and other characteristics. At the core of the Bill of Rights Blueprint are five principles: Safe and effective systems: The public is protected from unsafe or ineffective systems; Algorithmic discrimination protection: The public should not face discrimination in algorithms and systems, and automated systems should be used and designed in a fair manner; Data privacy: Automated systems should have built-in safeguards to protect the public's data from misuse and ensure that the public enjoys control over the use of data; Knowledge and explanation: The public has the right to know that it is using an automated system and to understand what it is and how it produces results that affect the public; The principle of substitutability: where appropriate, the public should be able to opt out of using automated systems and use manual alternatives or other alternatives. Because these principles are unregulated and not binding, the Bill of Rights Blueprint is not an enforceable "bill of rights" with legislative protection, but rather a forward-looking governance blueprint based on a vision for the future.

4.2.3. Legislation in China

From the perspective of China's current legal basis, a preliminary framework has been formed for the regulation of generative artificial intelligence at the macro level, and there are also more clear normative guidelines at the micro level. From a macro perspective, the governance framework of generative AI can be divided into two main lines: soft law and hard law.

Hard law pays attention to the guarantee of national coercive force to make its effective implementation. The main dimensions of hard law are: (1) top-level law. The Civil Code of the People's Republic of China, the Cybersecurity Law of the People's Republic of China, the Data Security Law of the People's Republic of China, the Law of the People's Republic of China on Scientific and Technological Progress, the Protection of Personal Information and other laws all involve the regulation of generative artificial intelligence. ② Local laws and regulations. Local people's congresses in Shanghai and Shenzhen have promulgated local regulations such as the Regulations of Shanghai Municipality on Promoting the Development of Artificial Intelligence Industry and the Regulations of Shenzhen Special Economic Zone on Promoting the Artificial Intelligence Industry. ③ Departmental regulations. The CAC and other departments have issued departmental rules and regulations such as the "Regulations on the Management of Internet Information Service Algorithm Recommendation" and "Interim Measures on the Management of Internet Information Service In-depth Synthesis".

In the "soft law" level :① industrial policy. China has released a series of industrial policies, including the New Generation of artificial Intelligence Development Plan. ② Ethics of science and technology. The New Generation Ethics Code for Artificial Intelligence defines the basic ethics code for artificial intelligence. ③ Industry standards. Including "Artificial Intelligence deep synthetic image System Technical Specifications" Information security technology machine learning algorithm security assessment specifications (draft for comment) ". It can be seen that China is currently encouraging the innovation and development of artificial intelligence at the macro level, and paying extra attention to the transparency, interpretability, ethical norms and technical standards of generative artificial intelligence, and conducting multi-level comprehensive governance of generative artificial

intelligence through "hard law" and "soft law", and has initially formed a regulatory framework for generative artificial intelligence.

5. SUGGESTIONS AND MEASURES

5.1. Legal governance

First, set up a reasonable classification and classification of supervision strategy. The risks brought by artificial intelligence are complex and diverse, and risk classification is a prerequisite for effectively dealing with many risks of generative artificial intelligence technology. On the one hand, it should be based on the technical characteristics and product characteristics of generative artificial intelligence. Taking into account the possible algorithm defects and the risks generated by deliberate induction, the risk range of "low, medium and high" is divided, and the risk assessment classification standard with "high scalability" is established to avoid rigidity. On the other hand, we should recognize the dynamic changes and unpredictability of risks, establish a "malleable" risk assessment mechanism, detect and evaluate risks in a dynamic and changing way, and respond to newly generated risk types in a timely manner.

Second, establish an accountability system. Regular disclosure of regulatory data and reporting of regulatory governance results will ensure that regulatory agencies strictly comply with their statutory obligations. The Provisional Provisions did not set up a special department to supervise artificial intelligence, but seven departments to coordinate the supervision of generative artificial intelligence, which should clarify the division of responsibilities of each department, and avoid the buck-passing between regulatory departments by establishing an accountability system. In addition, in order to improve the ability of regulators to perform their supervisory duties, they should be ensured that they have appropriate powers. The legitimacy and universal recognition of power is the first prerequisite for regulators to perform their duties, especially in the face of frontier fields such as generative artificial intelligence with high entry barriers, regulators need to have the power to require technology giants that develop, deploy or use generative artificial intelligence to disclose information and comply with compliance requirements. At the same time, implementation rules and disciplinary measures should be introduced to resolve the difficulties encountered in law enforcement and promote the implementation of regulatory measures.

The third is to clarify the legal status of AI providers and establish the application of their liability models and defense grounds. First of all, in view of the fact that the type of network service provider in the current law cannot cover provider A, there are obstacles to the regulation of provider A by content producers. Therefore, the legal status of provider A should be clarified first in legislation. Secondly, in order to encourage innovation, AI providers cannot be directly required to bear strict liability for their content, and it should be made clear that M providers can apply special "safe haven rules" or the provisions of Article 1195 of the Civil Code by analogy, that is, AI providers can be exempted from liability by taking timely remedial measures to contain damage after being notified by the right holder. Third, although generative AI is not a product, if the damage is caused by vulnerabilities and defects that cannot be overcome by the technical conditions before it is put into use, it is more consistent with the concept of fairness to allow the AI provider to plead "insufficient prior art."

5.2. Technical governance

On the one hand, the service provider of generative AI should guarantee the interpretability of the algorithm and the transparency of the data in the internal algorithm model that should comply with the technical standards. The first is to improve data transparency, including the following requirements: publish training data sources and verify the authenticity of the data to ensure that the system uses only trusted data sources; Take steps that require designers to make as objective choices

as possible to reduce or compensate for deviations in input data: integrity checks on data, such as encrypted hashes, enable data recipients to ensure that data has not been tampered with in transit. The second is to ensure the interpretability of the algorithm has the following requirements: specify the purpose of the system and clarify the role of generative artificial intelligence in the process of realizing the purpose: determine the accuracy of the system's expectations, and measure the success probability of the model being able to identify relevant elements and make positive predictions each time; It should clarify who is responsible for a specific element within a generative AI system, disclose which entity is responsible for generating or maintaining that element, and establish a data record for each decision.

On the other hand, the external should establish a dynamic monitoring system of algorithms. First, in terms of regulatory agencies, an algorithm filing system should be established, requiring service providers of generative artificial intelligence to file algorithms that do not involve trade secrets, and finding algorithm risks in the process of supervising generative artificial intelligence will return to the algorithm itself to seek punishment basis and improvement measures. Second, in terms of industry norms, the industry should be encouraged to self-regulate and set industry standards, requiring enterprises to consciously form a compliance governance structure and improve the compliance level of enterprises.

Moreover, in terms of user governance, users should be guided to standardize the use of generative artificial intelligence services to enhance citizens' digital literacy. In addition, attention should be paid to the anonymization of personal information, the differential protection of personal privacy, to prevent the risk of privacy leakage and personal information abuse, and provide a mechanism for individuals to report and feedback, so as to reduce the infringement of technology on vulnerable groups. Under the requirements of transparency and explainability of the algorithm, to a certain extent, the possibility of Internet giants using their platform advantages to collect users' private information and privately use it for artificial intelligence data training is avoided. At the same time, the openness of algorithmic decision-making can allow regulators and users to more clearly understand the process of making generative artificial intelligence decisions, and the supervision of all parties will further ensure the rationality and scientific decision-making, which can protect the interests of digitally vulnerable groups and promote technological innovation to a more just direction.

6. SUMMARY

As a key technology leading a new round of scientific and technological revolution and industrial transformation, generative artificial intelligence represented by ChatGPT has continuously spawned new scenarios, new formats, new models and new markets, changed the production mode of information and knowledge, reshaped the interaction mode between human and technology, and had a significant impact on industries such as education, finance, media and games. In this context, countries around the world have introduced artificial intelligence development strategies and specific policies to seize the strategic commanding heights. At the same time, security risks such as data leakage, false content generation, and improper use exposed by generative artificial intelligence have also attracted widespread attention from various countries. The potential risks brought by the development of new technologies and new industries are inevitable, and generative artificial intelligence has various types and wide application fields, but there are also many fuzzy areas. The development, application, and governance of generative AI is no longer a common challenge for a single country, but for the entire international community.

To sum up, how to develop responsible artificial intelligence under risks requires multi-party efforts. In order to effectively deal with the new challenges of generative artificial intelligence to information content governance, it is necessary to balance the relationship between security and development, the relationship between technological innovation and technological governance, and other multi-set relationships. It is necessary to consider the social impact, but also to improve the legal identification

standards and rights allocation rules. Artificial intelligence systems should be developed and used in a sustainable and environmentally friendly way, and artificial intelligence technologies and products should be controlled for the benefit of all, so as to promote the healthy development of the artificial intelligence industry.

REFERENCES

[1] Yang Jun, Zhang Xiaoyun. Copyright analysis of intelligent Reports generated by Information base software -- Reflections on Fei Lin v. Baidu case [J] Journal of Changchun University, 2020.30 (11)

[2] CAI Shilin, Yang Lei. Research on Risk and collaborative governance of ChatGPT intelligent robot application. Information Theory and Practice, 2023. 46 (05).

[3] Zhan Ailan, Jiang Qi. Adjacency rights protection of artificial intelligence Products [J] Journal of Zhejiang University of Technology (Social Science Edition), 202.20 (04)

[4] Wang Zhaoxia. Analysis of copyright barriers of Artificial Intelligence Generation and exploration of protection approaches [J]. Internet Week, 2022 (11)

[5] Wu Han-dong, ZHANG Ping, ZHANG Xiao-Jin. The Challenge of Artificial Intelligence to legal Protection of Intellectual Property Rights [J]. China Law Review, 2018 (02).

[6] Huang Shanshan. Impact of Artificial Intelligence on Copyright System and Countermeasures [J]. Journal of Chongqing University (Social Science Edition), 2020. 26 (01).