

^{1,*}Taoye Wang²Li Li³Xiang Chen⁴Kunzhu Li

A Study on the Risks and Countermeasures of False Information Caused by AIGC



Abstract: - Generative artificial intelligence (AIGC) has changed the traditional information production mechanism and has a wide range of application scenarios. At the same time, the security risks it exposes such as data leakage, false content generation, and improper utilization have also attracted widespread attention from various countries. The development, application and governance of AIGC no longer seem to be a common challenge faced by one country but by the entire international community. In order to effectively respond to the challenges of AIGC to the false information governance system, this article uses multiple methods such as literature analysis and in-depth research to elaborate on the potential risks of AIGC, and conducts an in-depth analysis of the global challenges of false information risk governance. Finally, Proposing governance paths and countermeasures from various perspectives such as supervision and ecology provides intelligence reference for the healthy development of the AIGC industry.

Keywords: Artificial Intelligence (AI), Generative Artificial Intelligence (GAI), Risk, Governance, Countermeasures.

I. INTRODUCTION

2023 is known as the first year of AIGC. With the emergence of big models such as GPT and Sora, people without video production foundations can generate high-definition videos of up to 60 seconds in a short period of time. Workers without ideas can quickly generate a logically clear and complete document through dialogue with the big model, and those who are not good at socializing can compensate for some interpersonal interaction through chatbot [1-3]. Technologies such as large language models, multimodal models, and embodied intelligence are profoundly changing human production and lifestyle at an unprecedented speed and scale. However, while AIGC technology is advancing rapidly, problems such as telecommunications fraud and the proliferation of false information are becoming increasingly prominent, becoming obstacles to people's realization of a better life and high-quality economic and social development [4]. These are prominent problems that urgently need to be solved in the innovative development process of the new generation of information technology industry in the new era [5-6].

After reviewing existing literature, it was found that

After consulting the existing literature, it was found that Shi [7] put forward the data risks that AIGC may cause, and put forward the risk governance path from the perspective of legal regulation. Qi [8] summarized the realistic problems and future risks of AIGC's false information, and proposed the governance path of AIGC's false information risk from the perspectives of collaborative governance and multiple co-governance. Wang [9] conducted research on the legal countermeasures to infringement caused by AIGC, and believed that various infringement risks caused by generative artificial intelligence could be effectively prevented and dealt with through special legislation. Yu et al. [10] conducted an in-depth analysis based on the text content of the AIGC accident report, explored the action framework of generating AI governance, and believed that it was necessary for the government, enterprises and society to form a governance participation model of "diversity + coordination + checks and balances", and to carry out information governance under the action framework of "context-consciousness-action". The above scholars adopted the methods of theoretical research and data analysis respectively, and put forward suggestions on AIGC governance from the perspectives of legal regulation and collaborative governance. Based on the analysis of other researchers, this study combined literature research, case analysis and other methods to conduct an in-depth study on the multi-level reasons and potential risks caused by the proliferation of false information caused by AIGC, and carried out a detailed analysis of the current governance challenges. Finally, from the aspects of governance mechanism construction, regulatory system construction, technological innovation and development, and science popularization and training, this paper proposes a comprehensive governance path

¹ Guangdong Institute of Scientific and Technical Information, 171 Lianxin Road, Guangzhou, Guangdong, China

² Guangdong Institute of Scientific and Technical Information, 171 Lianxin Road, Guangzhou, Guangdong, China

³ Guangdong Institute of Scientific and Technical Information, 171 Lianxin Road, Guangzhou, Guangdong, China

⁴ Guangzhou Metro Materials Co., LTD, 204 Huanshixi Road, Guangzhou, Guangdong, China

*Corresponding author: Taoye Wang

Copyright © JES 2024 on-line : journal.esrgroups.org

for AIGC false information risk, which provides an important information reference for subsequent researchers to carry out relevant research in this field.

II. THE NEW RISKS OF AIGC GENERATING FALSE INFORMATION

A. *The Widespread Application of AIGC Has Led to the Proliferation of False Information*

1) *No guarantee of data quality, providing soil for the growth of false information:* The training of AIGC models relies on massive amounts of data. In terms of GPT, in 2019, OpenAI released GPT-2 with a parameter count of 1.5 billion and a data set containing 15 billion tokens; In March 2023, GPT-4 was launched with a total of 1.8 trillion parameters. In addition to 13 trillion tokens, the training data also includes quite a few epochs and millions of lines of instruction fine-tuning data [11]. However, the training data of the big model mainly comes from unverified public resources such as news, blogs, forums and books on the Internet, including some low-quality data. Although it has been cleaned and preprocessed, the authenticity, timeliness, and objectivity of the data cannot be fully guaranteed, as well as manual data input errors, resulting in uncontrolled generation of content. According to the model training process, the generated false information will be reused for training and deep learning. In this iterative process, the false information not only grows in large quantities, but also approaches the level of human thinking, making it more difficult to identify.

2) *The model training mechanism has flaws, providing operational space for the production of false information:* At present, the training of generative artificial intelligence models mainly adopts the “Reinforcement Learning from Human Feedback” mode (RLHF), which means that the output content of the machine is manually adjusted and optimized before being fed back to the machine as input data. After the final red blue confrontation, it forms an AI model with high accuracy and universality. But under normal circumstances, the ability of humans to evaluate tasks will not improve with the progress of AI models. Starting from a critical point, humans will not be able to provide good training signals for artificial intelligence systems. From this point on, generative artificial intelligence will confront problems such as model training mechanism failure, uncontrollable generated content, and unreliability. In addition, with the continuous iteration of technology, there is also a possibility of RLHF being breached [12]. In August 2023, researchers from Carnegie Mellon University (CMU) and the Center for Artificial Intelligence Security in the United States jointly published a paper stating that they bypassed security measures such as RLHF through a novel “Universal and Transferable Adversarial Attacks” approach, allowing mainstream large models such as ChatGPT, Bard, Claude 2, and LLama 2 to generate harmful content, such as methods for creating bombs, and the researchers involved in the study stated that there is currently no effective way to fix the problem [13].

3) *Low cost, low threshold information generation technology provides convenience for the rampant growth of false information:* AIGC technology has the characteristics of strong universality, efficient content generation, low usage threshold and cost. These characteristics have enabled generative artificial intelligence tools such as ChatGPT, Midjournal, Voicemod, etc., to be widely used in a short period of time. Among them, ChatGPT has just been launched for two months, and the monthly active users have exceeded 100 million. People use AIGC tools to program, write news, papers, create images, and videos, but their ease of use is also easily exploited by criminals to carry out illegal activities such as fraud [14]. On the one hand, due to the low threshold for use, users do not need to be proficient in programming or rely on other professional knowledge. They can directly call the already trained underlying model online and easily operate it using ordinary smart devices such as computers and mobile phones. On the other hand, large models can generate content at an extremely low cost in a very short amount of time, significantly reducing the cost of AI onomatopoeia and facial recognition. For example, a reporter once investigated and found that someone on the Internet sold tutorials for making synthetic face videos and “universal models” of star faces. A set of models only cost 10 yuan, and after purchase, “face changing” videos can be directly generated in the software; the complete set of AI real-time face changing models only costs 35000 yuan, and merchants claim that it can be applied to various live streaming platforms. For example, DeepMedia, a company dedicated to detecting deeply forged content, pointed out that by the end of 2021, the cost of cloning sound was \$10000, but now, with the continuous iteration of technology, the cost of cloning sound has dropped to a few dollars [15-16].

B. *New Risks Caused by False Information*

1) *Manipulating cognition, strengthening the negative impact of cognitive warfare, hindering national security and stability:* Cognitive warfare refers to an unconventional form of warfare that influences the behavior and decision-making of the target population by influencing their cognition, in order to achieve certain political goals.

The dissemination of false information is its basic method. In 2017, former Chief of Staff of the United States Air Force, David Goodfin, first proposed the concept of “cognitive warfare”, emphasizing that “the form of war is shifting from attrition warfare to cognitive warfare.” In 2022, the United States elevated cognitive warfare to a strategic position of equal importance as physical warfare in its National Security Strategy. Since the Russo Ukrainian War, cognitive warfare has, for the first time, demonstrated its war and political effectiveness through its integration with large-scale warfare, AIGC, and powerful computing power. Western media such as Ukraine and the United States have frequently released false messages using artificial intelligence technology, such as “Ukrainian military killed and injured senior Russian military leaders” and “Ukrainian father tearfully bid farewell to his daughter, and then prepared to fight against the Russian military”, to incite Ukraine’s fighting will, incite global hatred towards Russia, and strongly undermine the morale of the Russian military. China and the United States are currently in a period of intense competition, with the United States frequently launching cognitive warfare attacks against China in areas such as the South China Sea issue, Hong Kong, Taiwan, Xinjiang, Tibet, and military security, attempting to distort the world’s perception of China’s image [17]. For example, when searching for content on ChatGPT about the United States slandering China, ChatGPT will generally give false positive answers. From an external perspective, this will distort the correct global perception of China, leading to a deterioration of China’s international image; internally, it may mislead some Chinese citizens’ judgments and even threaten our national security and stability.

2) *Using deep fake technology to commit fraud, endangering the property safety of the public:* Although the use of artificial intelligence to carry out criminal activities such as defamation and fraud had already emerged a few years ago when deep fake technology was introduced, with the explosion of AIGC technology, ordinary people can also access various advanced production tools. The fraudulent methods of criminals have become more diverse, and the difficulty of identifying false information has further increased. Forged audio and video are spread more frequently and have a wider coverage, making it increasingly difficult to prevent AI fraud, bringing greater harm to the property safety of the people. DeepMedia stated that the number of deeply forged videos in 2022 is three times that of 2021, and the number of voice forgeries is eight times that of 2021. Scammers usually disguise virtual numbers as the phone numbers of relatives and friends of the parties involved, and then use voice synthesis software to simulate the voices of their relatives and friends in order to obtaining property. Sometimes, in order to further gain the trust of the parties involved, criminals may also play pre-prepared AI face changing videos through social media platforms such as WeChat to confuse the parties involved. According to the Hubei Internet Police Inspection and Law Enforcement on May 6, 2023, the success rate of new online fraud with the support of AI facial recognition, voice synthesis and other technologies is close to 100%. With the release of the OpenAI cultural and biological video model Sora in early 2024, the increasingly fake AI video generation and deep forgery technology is capable of turning into a sharp blade in the hands of fraudsters and deniers, targeting unsuspecting individuals and becoming a severe social “cancer” that threatens the property security of the people [18].

3) *Spreading false and erroneous information, triggering a cognitive crisis and infringing on user rights and interests:* OpenAI pointed out in the GPT-4 technology report that the content generated by GPT-4 and early GPT models is not entirely reliable, and there may be hallucinations, that is, the production of absurd or unreal content unrelated to certain sources. For example, ChatGPT may provide seemingly logically consistent incorrect answers, which are severe nonsense. In daily life, it is common to generate false news, papers that violate scientific laws, images, videos, etc. If users lack discernment ability and professional knowledge in relevant fields, blindly believing in the generated information can lead to a deviation in their understanding of facts, especially for teenagers who have not yet formed a stable and mature view, they are prone to be eroded by false content and biased towards value orientation; for professionals in legal, medical, and other professional fields, ChatGPT may fabricate non-existent legal provisions, cases, or incorrect medical advice to answer questions, hindering judicial fairness and causing medical risks. In addition, improper use of AIGC tools can also lead to serious infringement issues [19]. Researchers have found that since August 2022, there has been an increase in the amount of highly realistic AI generated child molestation materials circulating on the dark web. These new materials are mostly based on the appearance of real victims, and are generated through certain content auditing software with flawed images and videos, seriously infringing on the legitimate rights and dignity of minors; meanwhile, the illegal use of AIGC tools to generate text, images, and videos may also pose potential risks of intellectual property infringement.

III. GLOBAL CHALLENGES IN MANAGING THE RISK OF FALSE INFORMATION GENERATED BY AIGC

A. *It is Difficult to Grasp the Degree to Which National Public Power Intervenes in AIGC Regulation*

Generative artificial intelligence brings tremendous momentum to high-quality economic development. McKinsey's research shows that AIGC is expected to contribute approximately \$7.9 trillion in positive economic impact to the global economy [20-21]. However, while AIGC is promoting social changes, it has also spawned a series of new risks such as the proliferation of false information, frequent fraud, and content infringement. In addition, the iteration speed of artificial intelligence is constantly accelerating, and the old "development first, governance later" approach is no longer applicable to the rapid development of artificial intelligence. How to balance the innovation and development of artificial intelligence with regulatory governance has become an urgent issue to be solved. In terms of regulatory thinking, the United States and Europe have formed two distinct camps. The regulatory model in the United States is "encouraging" and focuses on using the development of artificial intelligence systems to consolidate its leading position in the global economy and meet its national security needs. Encouragement based regulation is beneficial to the advancement of artificial intelligence technology, but there may be a problem of ignoring development risks. The regulatory model of the European Union belongs to a "conservative" type, with strict regulation in data protection and privacy. On March 31, 2023, Italian data protection agencies investigated ChatGPT and stated that Italy will not only block OpenAI's chatbot, but also initiate an investigation into whether they comply with the General Data Protection Regulations. Although the conservative regulatory model effectively solves the problem of information leakage, it also greatly limits the development of AIGC technology. Our country adheres to an inclusive and prudent regulatory concept for the development of artificial intelligence, and places encouragement of innovation as the top priority at all times, focusing on enhancing originality, and adopting classified and graded regulatory measures. However, we still face many practical difficulties, such as the inability of the legislative process to match the speed of artificial intelligence technology iteration, the unresolved issue of system connection between legislative documents, and there is a debate between the legislative focus on safety risk governance and guarantee of industrial development. To regulate or to develop is not a binary choice, but in the field of information technology, achieving a balance between development and regulation is quite difficult [22].

B. *The Ownership of Generated Content Rights and the Determination of Infringement Liability are Still being Explored*

In terms of the intellectual property ownership of AIGC generated content, there is currently no clear provision in the laws of various countries. Therefore, there is a certain degree of uncertainty in judicial practice regarding responsibility determination and rights confirmation. The main controversy lies in whether artificial intelligence has the ability to create, whether the creation process of generated content involves human factors and randomness, and whether the generated results are unpredictable [23]. On November 29, 2023, the first domestic AI generated image copyright case was concluded, causing industry discussions. In the case, the plaintiff Li used Stable Diffusion software to generate the involved images and published them on the Xiaohongshu platform. The defendant Liu, without obtaining the plaintiff's permission, intercepted the signature watermark of the involved images and used them in their blog posts. After discovery, the plaintiff filed a lawsuit to the Beijing Internet Court, which ruled in the first instance that the pictures created by Li using AI technology were original and could be identified as works and should be protected by the copyright law. This case attempts to explore the issue of copyright determination regarding the content generated by AIGC, but the judgment can only be based on the resolution of disputes in the case, and standardized judgment standards still need to be further deepened. In terms of determining the infringement liability of AIGC for generating false content, due to the technical characteristics of AIGC itself, it may lead to "hallucinations", which may infringe on the reputation rights of others. Someone once asked ChatGPT for information about someone, and ChatGPT provided feedback on the content of their alleged crime, citing specific legal documents and seemingly authoritative news reports as sources of information. However, upon investigation, the above information is purely fictitious, and the legal documents and news reports quoted by it do not exist at all. In this situation, the false information generated by AIGC seriously infringes on the reputation rights of others. Although China promulgated the "Interim Measures for the Management of Generative Artificial Intelligence Services" on July 10, 2023, which requires service providers to guide users in using content generated by generative artificial intelligence to avoid generating content that harms the legitimate rights and interests of others, the text only proposes the requirement that "if providers discover illegal content, they should take timely measures such as stopping generation, stopping transmission, eliminating, and taking measures such as

model optimization training for rectification, and reporting to relevant regulatory authorities”, without proposing supporting measures or implementation rules, which may lead to difficulties in the implementation of the regulation.

C. *The R&D of False Information Recognition Technology is Proceeding Slowly*

At present, the research and development progress of false information recognition technology is not satisfactory. In terms of technology itself, due to limitations in training data volume and data quality, existing false information detection algorithms have problems such as high misjudgment rate and poor generalization ability, and the accuracy of classifying unknown data is also not high, which cannot fully meet practical application needs. In addition, with the continuous development of technologies such as deep forgery, the forms and methods of generating false information are also constantly updating and changing, and existing detection algorithms are currently unable to fully cope. In terms of external factors, on the one hand, false information identification business does not directly generate economic benefits, and social resources have a low willingness and intensity to invest in false information identification technology; In addition, false information recognition technology requires a large number of training samples, computing resources, and the participation of professionals, which poses certain technical and financial barriers for small and medium-sized enterprises as well as individual users. On the other hand, due to the lack of clear legal regulations on the specific responsibility of AIGC product developers for identifying false information, the existing fact verification mechanism has not clearly defined the rule system for verifying technology development and its application from a governance perspective. In addition, the industry has not yet formed a good mechanism for identifying false information and sharing data, which further slows down the progress of related technology research and development.

IV. COUNTERMEASURES AND SUGGESTIONS

A. *Introducing New Agile Governance Ideas, Shaping AIGC Content Generation Governance Mechanisms*

Agile governance is a governance concept proposed for emerging technologies. Compared to traditional governance models, it has characteristics such as agility, adaptability, sustainability, self-organization, and inclusiveness, which coincides with China’s regulatory philosophy of inclusiveness and prudence towards AIGC. Based on this, when carrying out AIGC content generation governance work, the advantages and characteristics of agile governance should be combined, and AIGC technology and industry development should be continuously observed and evaluated. Risk identification should be done well, and risk types should be divided into known and unknown risks based on the predictability of risks. Pre-prevention mechanisms should be adopted for the former, and post response mechanisms should be adopted for the latter to adapt to maximize the unpredictability of adapting to AIGC technology risks; continuously innovating governance tools and methods, exploring the establishment of a regulatory sandbox system, encouraging enterprises to actively carry out technological innovation within specific sandboxes, promoting the sharing of enterprise datasets within the sandbox, collaborating on system operation testing, exploring solutions for combating false content, and continuously improving governance flexibility; establishing a “notification deletion” rule to address violations intentionally induced by users. If the service provider takes necessary measures such as deletion, blocking, and disconnection in a timely manner upon receiving notification from the rights holder, and can prove that it has taken necessary measures such as output interception and filtering during the model development and operation stage, it is not required to bear infringement liability. This leaves enough room for trial and error in the development of artificial intelligence technology and promotes the advancement of emerging technologies Develop towards goodness.

B. *Accelerating the Pace of Regulatory Technology R&D, Providing Multiple Protections for Content Security Governance*

Strengthening the supervision and governance technical support of AIGC system design and application throughout the entire chain, firmly grasp the initiative of AIGC development. Conducting research on robust optimization, adversarial training, defense distillation, and other technologies to enhance the ability of AIGC systems to resist adversarial attacks. We should research digital watermarking technology, improve intellectual property and Internet information digital fingerprint, enhance information authenticity, traceability and accountability, minimize the risk of information misinformation, and forge new tools for intellectual property protection and information fidelity. By utilizing blockchain technology, we can enhance the credibility, transparency and regulatory compliance of AIGC services via recording data transmission, access control and user permissions on immutable ledgers. We can also build a network information supervision platform, combining AI image tampering detection technology, false information intelligent perception technology, rule-based content

filtering algorithms, and deep learning based content auditing models, to conduct uninterrupted network inspections 24/7, monitoring, warning, and filtering generated content in real time, and promptly handling false and harmful content. To establish a public algorithm library for AI generated content identification, enriching deep forgery identification tools, and continuously strengthening the support of technology for AIGC regulation.

C. *Improving the Regulatory System, Creating a Green Ecosystem for Content Generation*

Further improving the construction of the AIGC regulatory system, formulating the “Interim Measures for the Management of Generative Artificial Intelligence Services” and other current laws and regulations, guiding providers of generative artificial intelligence services to take effective measures to prevent users from engaging in illegal activities such as false and harmful information generation; actively exploring the formulation of supporting local policies and regulations, and effectively supplement and connect national laws and administrative regulations. Standardizing the application scope of AIGC technology, services, and products in specific industries, and improving the multidimensional standard specification system and evaluation and verification mechanism for the development, review, application, and security of AI content generation tools. Relying on leading enterprises and universities to explore the establishment of artificial intelligence anti fraud platforms and laboratories, and collaborating with various social forces to provide comprehensive network security protection for the public. Guiding enterprises and institutions that provide generative artificial intelligence services to establish professional review mechanism, establish an internal review system for model algorithms and safety testing standards, encouraging them to actively participate in AIGC global governance, and contribute Chinese wisdom and solutions to the robust development of artificial intelligence, especially the inclusive use of generative artificial intelligence technology in human society.

D. *Strengthening Science Popularization and Training, Enhancing the Public’s AI Safety Accomplishment*

Vigorously promoting the basic knowledge of artificial intelligence to the general public, so that the public can rationally approach the development of artificial intelligence technology and eliminate their misunderstandings and panic about artificial intelligence. Strengthening the publicity and popularization of new artificial intelligence fraud, effectively enhancing the public’s cognitive ability and false information identification literacy towards artificial intelligence technology. Strengthening science and technology ethics education and professional ethics training for artificial intelligence technology developers, offering engineering ethics courses in the academic education stage of majors such as artificial intelligence and computer science, and adding assessment content on ethical, professional knowledge and personal moral qualities in relevant professional qualification exams, comprehensively shaping the technological ethics and network security awareness system of AI practitioners.

V. CONCLUSION

As a new type of information generation mode, the rapid development of AIGC has put forward new challenges and requirements to the existing technical governance. This paper starts from the underlying causes of false information in AIGC, summarizes the situations that cause real problems and new risks, analyzes the global challenges of false information risk governance, and then puts forward governance paths and countermeasures. These studies contribute to a better understanding of the characteristics of AIGCs and their impact on society, law and technology. These findings have important reference value for policy makers, industry practitioners and academic researchers to help formulate more effective AIGC governance strategies to ensure the safe and controllable development of AIGC.

REFERENCES

- [1] Noy S, Zhang W. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 2023, 381(6654): 187-192.
- [2] Berg J M, Raj M, Seamans R. Capturing value from artificial intelligence. *Academy of Management Discoveries*, 2023, 9(4): 424-428.
- [3] Jovanovic M, Campbell M. Generative artificial intelligence: Trends and prospects. *Computer*, 2022, 55(10): 107-112.
- [4] Mandapuram M, Gutlapalli S S, Bodepudi A. Investigating the Prospects of Generative Artificial Intelligence. *Asian Journal of Humanity, Art and Literature*, 2018, 5(2): 167-174.
- [5] Wach K, Duong C D, Ejdy J. The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, 2023, 11(2): 7-30.
- [6] Longoni C, Fradkin A, Cian L. News from generative artificial intelligence is believed less//*Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 2022: 97-106.

- [7] Shi Y. Study on security risks and legal regulations of generative artificial intelligence. *Science of Law Journal*, 2023, 2(11): 17-23.
- [8] Chenhang Qi. Research on the risks of disinformation from Generative artificial intelligence and its governance paths. *Information Theory and Practice*, 2024, 47(03):112-120.
- [9] Liming Wang. Legal Response to Infringement of Generative Artificial Intelligence. *Chinese Journal of Applied Law*, 2023 (05): 27-38.
- [10] Yu Zhu, Chen Guanze, Lu Yongrong. A Generative Artificial Intelligence Governance Action Framework: Content Analysis Based on AIGC Accident Reporting Text. *Library and Information Knowledge*, 2023, 40 (04): 41-51.
- [11] Uthamaputhran S, Yusuff Y Z, binti Bahari N. Effectiveness of Adaptation of Artificial Intelligence on Management Domains in Industries: Bibliometric Analysis. *Migration Letters*, 2024, 21(S4): 1572-1588.
- [12] Duffourc M, Gerke S. Generative AI in health care and liability risks for physicians and safety concerns for patients. *Jama*, 2023.
- [13] Zou A, Wang Z, Kolter J Z, Wei Z, Chen J. Enhancing the self-universality for transferable targeted attacks//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 12281-12290.
- [14] Elgesem D. The AI Act and the risks posed by generative AI models//CEUR Workshop Proceedings. 2023, 3431.
- [15] Barrett C, Boyd B, Bursztein E. Identifying and mitigating the security risks of generative ai. *Foundations and Trends® in Privacy and Security*, 2023, 6(1): 1-52.
- [16] Walkowiak E, MacDonald T. Generative AI and the Workforce: What Are the Risks? Available at SSRN, 2023.
- [17] Allen D, Weyl E G. The Real Dangers of Generative AI. *Journal of Democracy*, 2024, 35(1): 147-162.
- [18] Elgesem D. The AI Act and the risks posed by generative AI models. *CEUR Workshop Proceedings*. 2023, 3431.
- [19] Joshi M A. The Security Risks of Generative Artificial Intelligence. Available at SSRN 4735175, 2024.
- [20] Fui-Hoon Nah F, Zheng R, Cai J. Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, 2023, 25(3): 277-304.
- [21] Ooi K B, Tan G W H, Al-Emran M. The potential of generative artificial intelligence across disciplines: Perspectives and future directions. *Journal of Computer Information Systems*, 2023: 1-32.
- [22] Harreis H, Koullias T, Roberts R. *Generative AI: Unlocking the future of fashion*. McKinsey & Company, 2023.
- [23] Knutsen L Z, David Patón-Romero J, Hannay J E. A Survey on the Perception of Opportunities and Limitations of Generative AI in the Public Sector//World Conference on Information Systems for Business Management. Singapore: Springer Nature Singapore, 2023: 503-520.