世界互联网大会
World Internet Conference

中国人民大学
RENMIN UNIVERSITY OF CHINA

Outcomes of World Internet Conference
Think Tank Cooperation Program

Challenges and Responses of Generative Artificial Intelligence to the Governance of Online Information Content

Renmin University of China

APRIL 2026


世界互联网大会
World Internet
Conference


中國人民大學
RENMIN UNIVERSITY OF CHINA

Renmin University of China Law School Working Group

Chair

Yang Dong	Dean of Law School, Renmin University of China
Li Mingxuan	Lecturer of Renmin University of China

Members

Xu Xiaoxiong	Lecturer of Renmin University of China
Zhu Yutao	Lecturer of Renmin University of China
Wu Yaxi	Ph.D. Student of Renmin University of China
Lyu Haoran	Ph.D. Student of Renmin University of China

Foreword

Against the backdrop of the rapid development of generative artificial intelligence, the production model of online information content is undergoing profound transformation. As the risks to online information content brought about by generative AI become increasingly prominent, there is an urgent need to establish a governance framework for online information content that is suited to the era of generative AI.

This report systematically reviews how generative AI is reshaping the ecosystem of online information content, and provides an in-depth analysis of the challenges faced by existing governance mechanisms for online information content. In light of international trends in AI regulation, the report focuses on key issues such as the identification of AI-generated content, platform governance, and the allocation of responsibilities. It concludes by proposing policy recommendations for improving the governance of AI-generated content.

Contents

I. Generative Artificial Intelligence Reshapes the Online Information Content Ecosystem

(I) The Rise of Generative Artificial Intelligence

Generative artificial intelligence refers to "technologies that generate content such as text, images, voices, videos, and code based on algorithms, models, and rules."[1]One common classification method of artificial intelligence divides it into discriminative artificial intelligence and generative artificial intelligence. Discriminative artificial intelligence learns the conditional probability distribution within data, that is, the probability of a sample belonging to a specific category, and then makes judgments, analyses, and predictions for new scenarios. Typical examples include facial recognition, recommendation systems, autonomous driving, and other intelligent decision-making systems.[2]Generative AI, on the other hand, learns the joint probability distribution within data, which is the probability distribution of vectors composed of multiple variables in the data, summarizes and induces existing data, and creates entirely new content on this basis.[3]The underlying technology of discriminative artificial intelligence is relatively mature and has been widely applied across various fields for a longer period. In contrast, breakthroughs in generative artificial intelligence emerged relatively later, and it is only in recent years that it has begun to experience explosive growth at the application level.[4]

As early as the 1950s, models of generative artificial intelligence had already emerged. Researchers proposed some fundamental statistical models, such as Hidden Markov Models (HMM) and Gaussian Mixture Models (GMMs). Using these models, attempts were made to enable computer-aided creation. A notable example is the Illiac Suite of 1957, which is considered the first musical work created by a computer. At that time, the model structures were relatively simple, and their application relied on a large number of manually written rules with high development costs; therefore, the development of generative AI was relatively slow. After entering the 21st century, neural network models began to achieve significant results in various tasks, and generative artificial intelligence technology also advanced accordingly. Particularly with the progress in hardware technology and the development of deep learning algorithms, it became possible to train deeper neural network models. The increase in model complexity led to a qualitative leap in the quality of generated content.

In the field of computer vision, in 2014, Ian Goodfellow and others proposed Generative Adversarial Networks (GANs), bringing a milestone advancement to the domain of image generation. The quality of generated images reached a level virtually indistinguishable to humans.[5] Subsequently, models such as StyleGAN and Denoising Diffusion Probabilistic Model further enhanced the quality of image generation. In the field of natural language processing, in 2017, Ashish Vaswani and others at Google proposed the Transformer model, which became a benchmark for natural language generation.[6] Subsequently, various pretrained language models (such as BERT, GPT, and BART) further improved the quality of text generation. A recent technology of transformative significance is large language models. The emergence of this technology signifies a major breakthrough for generative artificial intelligence at the technical level and

[1] Measures for the Management of Generative Artificial Intelligence Services (Draft for Comments), Article 2.
[2] DING L. Generative Artificial Intelligence: The Logic and Applications of AIGC[M]. Beijing: CITIC Publishing Group, 2023: 6.
[3] DING L. Generative Artificial Intelligence: The Logic and Applications of AIGC[M]. Beijing: CITIC Publishing Group, 2023: 6.
[4] DING L. Generative Artificial Intelligence: The Logic and Applications of AIGC[M]. Beijing: CITIC Publishing Group, 2023: 8.
[5] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Networks[J/OL]. Advances in Neural Information Processing Systems, 2014[2023-10-06]. https://arxiv.org/abs/1406.2661.
[6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need[J/OL]. Advances in Neural Information Processing Systems, 2017[2023-10-06]. https://arxiv.org/abs/1706.03762.

has driven its large-scale adoption in people's daily lives and productivity.

Large Language Model (LLM) refers to a pre-trained language model with parameters on the scale of tens of billions or more.[7]As early as 2020, OpenAI released the GPT-3 model with 175 billion parameters, which already demonstrated extraordinary capabilities in many natural language processing tasks.[8]Building upon this model, OpenAI developed the GPT-3.5 model; the initial version of ChatGPT was based on this large language model, and its parameter scale was also at the 175 billion level. Subsequently, OpenAI released the larger-scale GPT-4 model, whose parameter scale is rumored to reach the 1.76 trillion level. The increase in scale has endowed the model with certain "emergent abilities," which are "abilities that do not exist in small models but appear in large models," such as in-context learning, instruction following, and step-by-step reasoning.[9]These abilities give ChatGPT better general capabilities to handle complex tasks that were previously difficult to solve. With the success of ChatGPT, people realized the powerful capabilities and huge potential of large language models and generative AI. Generative AI companies represented by OpenAI, DeepSeek, Anthropic, and Moonshot AI have risen rapidly, and traditional technology companies represented by Google, Alibaba, and ByteDance have also entered the development and application of generative AI, setting off a global boom in the development of generative AI.

(II) Generative Artificial Intelligence Transforms the Production of Online Information Content

Before the rise of generative AI, the production modes of online information content were mainly Professional Generated Content (PGC) and User Generated Content (UGC). PGC refers to content created, edited, and published by professional content creators or teams.[10]This content production mode had already existed as early as the traditional media era, such as newspapers, television, and film. After entering the digital era, the PGC mode also served as the primary production mode for online information content, whether it was news portal websites in the Web 1.0 era or long-form video and music websites in the Web 2.0 era, all of which are typical representatives of the PGC production mode. With the rapid growth of network users and the fast development of network technology (especially storage and transmission technologies), the UGC mode has also gradually become another important production mode for online information content. UGC refers to content where ordinary users or audiences participate in creation, editing, and publishing.[11] Typical representatives based on this mode include social networks, blogs, and knowledge-sharing platforms, most of which emerged in the Web 2.0 era.

[7] ZHAO W X, ZHOU K, LI J, et al. A Survey of Large Language Models[Z/OL]. (2023-06-29)[2023-08-15]. https://arxiv.org/abs/2303.18223.
[8] BROWN T, MANN B, RYDER N, et al. Language Models are Few-Shot Learners[Z/OL]. Advances in Neural Information Processing Systems, 2020[2023-08-15]. https://arxiv.org/abs/2005.14165.
[9] WEI J, TAY Y, BOMMASANI R, et al. Emergent Abilities of Large Language Models[J/OL]. Transactions on Machine Learning Research, 2022[2023-08-15]. https://arxiv.org/abs/2206.07682.
[10] Ryan. The Evolution of Content Creation: From PGC and UGC to AIGC[EB/OL]. (2023-05-05)[2026-03-09]. https://zhuanlan.zhihu.com/p/627012065.
[11] Ryan. The Evolution of Content Creation: From PGC and UGC to AIGC[EB/OL]. (2023-05-05)[2026-03-09]. https://zhuanlan.zhihu.com/p/627012065.

The rise of generative AI has had a profound impact on the production mode of online information content, primarily making Artificial Intelligence Generated Content (AIGC) a mainstream production mode for online information content. Some views hold that AIGC is "a new production method that uses artificial intelligence technology to automatically generate content, following Professional Generated Content (PGC) and User Generated Content (UGC)."[12]With the development and wide application of generative AI, the proportion of AI-generated content in online information content is becoming higher and higher. Some predictions claim that by 2026, as much as 90% of the content on the Internet may be generated or enhanced by AI.[13] This indicates that AIGC is gradually becoming an important source of online information content, and may even replace PGC and UGC in the future to become the most primary production mode for online information content.

(III) Generative Artificial Intelligence Triggers Online Information Content Risks

The development of technology often has two sides; the rise of generative AI has not only revolutionized the production mode of online information content but also triggered various risks. In 2021, a study by DeepMind summarized 21 risks in 6 categories that large language models might cause.[14] Among these risks, many are related to information content. For example, the study mentioned that large language models might generate toxic language, leak privacy or sensitive information, provide misinformation, and make the production of disinformation cheaper and more effective. These risks can all be classified as online information content risks triggered by generative AI. Regarding the types of risk, based on the types of generated online information content, the online information content risks triggered by generative AI mainly include personal information risks, harmful information risks, and erroneous information risks.

First, personal information risks. If personal information exists in the training data, large language models might "remember" and generate this information. For example, research has found that among 600,000 sampled texts generated by GPT-2, at least 604 (approximately 0.1%) contained text copied verbatim from training data, some of which included personally identifiable information from the training data.[15] There are also studies conducted experiments targeting ChatGPT and New Bing, finding that: (1) Compared to previous language models, ChatGPT can better prevent users from generating personal information through simple prompts; however, if users utilize elaborately designed Jailbreaking Prompts, ChatGPT will still generate personal information. (2) New Bing integrates ChatGPT and search engines, bringing greater privacy risks. Utilizing New Bing, people can not only generate personal information through simple prompts but even generate personal information beyond the training data.[16]

Second, harmful information risks. Generative AI may

[12] China Academy of Information and Communications Technology, JD Explore Academy. White Paper on Artificial Intelligence Generated Content (AIGC) [R]. Beijing: China Academy of Information and Communications Technology, JD Explore Academy, 2022.
[13] WANG D D. U.S. Media Predicts Eight Major AI Trends for 2026[EB/OL]. (2025-10-10)[2026-03-09].
[14] WEIDINGER L, MELLOR J, RAUH M, et al. Ethical and social risks of harm from Language Models[Z/OL]. (2021-12-08)[2023-08-15]. https://arxiv.org/abs/2112.04359.
[15] CARLINI N, TRAMER F, WALLACE E, et al. Extracting training data from large language models[C/OL]//The Proceedings of 30th USENIX Security Symposium: USENIX Association, 2021[2023-08-15]. https://www.usenix.org/system/files/sec21-carlini-extracting.pdf.
[16] LI H, GUO D, FAN W, et al. Multi-step Jailbreaking Privacy Attacks on ChatGPT[Z/OL]. (2023-04-11)[2023-08-15]. https://arxiv.org/abs/2304.05197.

also generate illegal and harmful information such as hate speech, fraudulent information, and infringing content. For example, according to OpenAI's report, ChatGPT may respond to harmful instructions or exhibit biased behavior.[17] Other research has found that if ChatGPT is instructed to play the role of a "bad person," it will have a higher probability of generating harmful information.[18] Furthermore, malicious users may utilize generative AI to engage in illegal activities such as fraud and infringement, generating various types of illegal and harmful information.

Third, erroneous information risks. Generative AI may generate erroneous and untrue information. For example, a test by NewsGuard showed that when false statements were entered into ChatGPT as prompts, the generated answers were full of various misinformation.[19] Possible reasons for AI-generated erroneous information include: (1) The training data contains erroneous information, which the AI model learns during the training process and generates during the output process; (2) Language models exhibit "Hallucination" phenomena, which may output meaningless information or information unfaithful to the source content, thereby generating erroneous information; (3) Users maliciously utilize generative AI to produce disinformation, such as using misleading instructions to guide language models into generating fake news, or using image generative AI to perform deepfakes.

[17] OPENAI. Introducing ChatGPT[EB/OL]. (2022-11-30)[2023-08-15]. https://openai.com/blog/chatgpt.
[18] DESHPANDE A, MURAHARI V, RAJPUROHIT T, et al. Toxicity in ChatGPT: Analyzing Persona-assigned Language Models[Z/OL]. (2023-04-11)[2023-08-15]. https://arxiv.org/abs/2304.05335.
[19] OREMUS W. The clever trick that turns ChatGPT into its evil twin[N/OL]. The Washington Post, 2023-02-14[2023-08-15]. https://www.washingtonpost.com/technology/2023/02/14/chatgpt-dan-jailbreak/.

II. Challenges of Generative Artificial Intelligence to the Governance of Online Information Content

(I) Information Disorder: Challenges in the Identification of Generated Content

As the proportion of AI-generated content in online information content continues to increase, how to accurately distinguish between AI-generated content and human-created content has become an important issue in online information content governance. AI-generated content differs from human-created content in at least two important aspects; therefore, it is necessary to distinguish it from human-created content and treat it differently at the governance level.

First, AI-generated content is more likely to trigger online information content risks. With the emergence of generative AI, online information content risks are further exacerbated, and the difficulty of governance will also further increase. On one hand, generative AI can automatically generate a large amount of illegal and inappropriate information, reducing production costs and increasing transmission efficiency, thereby leading to a flood of illegal and inappropriate information and increasing the difficulty of supervision. On the other hand, AI-generated content has high-quality effects, is more likely to influence the information audience, and is more difficult to identify. For example, an experiment based on GPT-3 found that under certain conditions, text created using GPT-3 was as persuasive as human-created propaganda.[20] With the development of generative AI technology, it has become difficult to reliably distinguish between AI-generated content and human-created content purely from the content itself. Research shows that the accuracy of humans in distinguishing between AI-generated text and human text is very low, with the identification accuracy for AI-generated text being only 10% and the identification accuracy for human text being only 17%, which is far below the level of random guessing.[21]Other studies show that in experiments distinguishing between AI-generated images and real images, human identification accuracy is about 62%, which is close to the level of random guessing.[22]

Second, AI-generated content receives weaker legal protection of rights. Human-created content generally obtains copyright protection in law as long as it meets a minimum degree of originality. However, it is difficult for AI-generated content to obtain legal rights protection under the existing copyright law system. In major jurisdictions represented by China and the United States, courts or copyright administrative authorities both regard human creation as one of the constituent elements of a work; therefore, content entirely generated by AI cannot be protected by copyright. In a series of cases decided by Chinese courts, the courts have all determined that content automatically generated entirely by AI does not constitute a "work" in the sense of copyright law. For example, in the case of "Beijing Feilin Law Firm v. Beijing Baidu Netcom Science and Technology Co., Ltd. regarding copyright infringement dispute," the court held that creation by a natural person is a necessary condition for a work under copyright law, and the "work-like" results generated intelligently by computer software are not works in the sense of copyright law.[23] The United States Copyright Office holds a similar view. For example, the US Copyright Office released the "Copyright Registration Guidance: Works

[20] GOLDSTEIN J A, CHAO J, GROSSMAN S, et al. Can AI Write Persuasive Propaganda?[Z/OL]. (2023-02-21)[2023-08-15]. https://osf.io/preprints/socarxiv/fp87b/.
[21] Cheng A, Lin Y, Reedy G, et al. Ability of AI detection tools and humans to accurately identify different forms of AI-generated written content[J]. Advances in Simulation, 2025, 10(1): 66.
[22] Roca T, Roman A C, Vega J T, et al. How good are humans at detecting AI-generated images? Learnings from an experiment[J]. arXiv preprint arXiv:2507.18640, 2025.
[23] Beijing Internet Court, Civil Judgment, (2018) Jing 0491 Min Chu No. 239.

Containing Material Generated by Artificial Intelligence," stating that if the applied material is generated by AI and lacks human creation, the Copyright Office will not grant registration.[24] In practice, the US Copyright Office has already rejected registration applications for a series of AI-generated contents including Zarya of Dawn, Théâtre D'opéra Spatial, and SURTAST.[25]

Therefore, effectively identifying AI-generated content is an important foundation for maintaining the current order of online information content. Currently, both academia and industry have invested significant effort in developing identification technologies for AI-generated content. Existing identification technologies mainly distinguish between AI-generated content and human-created works by analyzing semantic patterns and statistical features of "work-like" results such as text and images. Many studies indicate that AI-generated content often presents patterns in semantic structures and statistical characteristics that differ from human-created content; these differences can be identified through technical means to discover traces of AI generation.[26]However, the reliability of these technologies is still insufficient, and false positives or false negatives may occur. First, when the style and pattern of human-created content are relatively close to AI-generated content, there is a possibility it will be erroneously labeled as AI-generated content, thus creating a risk of false positives. Second, the identification accuracy of existing technology for AI-generated text is also relatively limited; especially after AI-generated content has been manually modified, many identification technologies find it difficult to detect that it was generated by AI, thus producing a high proportion of false negatives. Consequently, existing identification technologies have not yet reached the standard of certainty required by law, and their identification results are difficult to adopt as legally admissible conclusions.

(II) Platform Revolution: The Absence of Rules for New Platforms

The development and wide application of generative AI have spawned a large number of new platforms providing generative AI services, namely Generative AI Service Providers. Based on generative models, these platforms provide content production capabilities and services to users through conversational interaction, text generation, image generation, video generation, etc. Typical representatives include OpenAI's ChatGPT, Google's Gemini, Alibaba's Qwen, and DeepSeek. These platforms are both the main force in the development and application promotion of generative AI technology and important participants in the governance of AI-generated content. Through reasonable regulation of platforms, it is possible to effectively encourage platforms to actively participate in the governance of AI-generated content.

The platform regulation rules targeting generative AI service providers remain in the early exploratory stage. Currently, many countries and regions still rely on existing platform regulatory rules, particularly legal

[24] Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence[EB/OL]. (2023-03-16)[2026-02-01]. https://www.federalregister.gov/documents/2023/03/16/2023-05321/copyright-registration-guidance-works-containing-material-generated-by-artificial-intelligence.

[25] Zarya of the Dawn (VAu001480196) [EB/OL]. (2023-02-21)[2026-03-09]. https://copyright.gov/docs/zarya-of-the-dawn.pdf; Re: Second Request for Reconsideration for Refusal to Register Théâtre D'opéra Spatial (SR # 1-11743923581; Correspondence ID: 1-5T5320R) [EB/OL]. [2026-03-09]. https://www.copyright.gov/rulings-filings/review-board/docs/Theatre-Dopera-Spatial.pdf; Copyright Review Board of US Copyright Office, Re: Second Request for Reconsideration for Refusal to Register SURYAST (SR # 1-11016599571; Correspondence ID: 1-5PR2XKJ) [EB/OL]. (2023-12-11)[2026-03-09]. http://copyright.gov/rulings-filings/review-board/docs/SURYAST.pdf.

[26] Sardinha T B. AI-generated vs human-authored texts: A multidimensional comparison[J]. Applied Corpus Linguistics, 2024, 4(1): 100083; Georgiou G P. Differentiating between human-written and AI-generated texts using automatically extracted linguistic features[J]. Information, 2025, 16(11): 979.

frameworks designed for internet service providers, to address issues related to generative AI service providers. Some countries and regions have also implemented specialized legislation and regulations to establish specific provisions for the regulation of such providers. This paper argues that generative AI service providers differ from traditional internet service providers, particularly in aspects related to the governance of online information content, where they exhibit distinct characteristics, thereby posing challenges to the application of existing platform regulation rules.

First, from the perspective of platform attributes, generative AI service providers exhibit characteristics of complexity and diversity. Firstly, generative AI service providers are neither pure technical service providers nor traditional content service providers, but rather a special type with composite attributes. In the traditional sense, network service providers can be divided into technical service providers and content service providers, with significant differences in their obligations and responsibilities regarding online information content governance. For example, content service providers generally bear an obligation to review the content they provide; if content issues trigger damage, their fault is generally defaulted, and they need to bear corresponding responsibilities. Technical service providers, however, do not bear a general review obligation for the content users provide on their platforms; if content issues trigger damage, unless the technical service provider knew or should have known and failed to take necessary measures, they generally do not bear tort liability. However, generative AI service providers differ significantly from both traditional technical service providers and content service providers, playing a role more akin to being positioned between technical service providers and content service providers. Therefore, platform regulation rules designed based on the division of traditional network service providers may be difficult to apply directly to the special type of generative AI service providers. Secondly, the types of generative AI service providers are also very diverse. Currently, various operating models exist for generative AI services, and under different operating models, the service providers' ability to manage information content also varies. This means there should also be differentiated regulation rules for different types of generative AI service providers, and their obligations and responsibilities in online information content governance should also be distinguished.

Second, from the perspective of operating mechanisms, generative AI possesses characteristics of being a "black box" and operating at scale. Firstly, generative AI represented by large models relies on massive data for training, and its internal parameters and decision-making logic are often difficult to understand or verify. When a model outputs illegal and inappropriate information, it is difficult for people to accurately hold specific publishers and content creators accountable to allocate responsibility and prevent subsequent output as is done with traditional content platforms. Secondly, generative AI can generate large amounts of illegal and inappropriate information in a short period, and the risks it generates may be rapidly amplified. Most existing platform regulation rules are built upon the logic of ex-post disposal; facing automated and large-scale online information content production modes, relying solely on ex-post processing is clearly insufficient to address potential risks.

Therefore, due to the specificity of generative AI service providers as a new type of platform, existing platform regulation rules may be unable to address the challenges they bring to online information content governance, resulting in an absence of rules. How to build platform

regulation rules that adapt to the characteristics of this new type of platform while ensuring the vitality of technical development is also an issue that currently needs to be addressed urgently in online information content governance.

(III) Liability Dilemma: Allocation Problems of Content Damage

The online information content risks triggered by generative AI may lead to the occurrence of damage, thereby creating the problem of damage liability allocation. Such content liability issues involve multiple subjects and diverse scenarios, and are also a core difficulty in the governance of AI-generated content.

First, content liability issues may involve multiple subjects. AI-generated content may involve multiple subjects such as generative AI model developers, training data providers, service platform operators, third-party application developers, and users. For example, when using a certain generative AI application to generate content, the model might be developed and continuously trained by model developer A; training data provider B is responsible for providing A with cleaned training data; service platform operator C deploys the model developed by A and provides API services to the public; third-party application developer D calls C's API service to develop a generative AI application; and finally, user E uses D's application to guide the generation of corresponding content through prompts. If the generated content constitutes an infringement of others' rights, which of the aforementioned subjects should bear responsibility? Different subjects have different degrees of participation, control, and foreseeability in the content formation process; how to delineate reasonable responsibility boundaries among these subjects is the primary challenge that content liability issues may face.

Second, content liability issues may involve different scenarios, and there is controversy over whether it is necessary to differentiate treatment based on the characteristics of the scenario. For example, content damage may involve the infringement of different types of rights, including personality rights and intellectual property rights; when the type of rights infringed differs, is it necessary to apply different rules? Furthermore, when generative AI is applied in different scenarios, the risk of damage caused by its generated content also varies; should differentiated treatment be based on the level of risk? These disagreements reflect that the content liability issues triggered by generative AI may not be resolvable by a single liability rule.

III. International Development Trend of Artificial Intelligence Regulation

(I) United States: Development-oriented Soft Regulation

Within the global landscape of AI governance, the United States has generally adopted a development-oriented model characterized by soft regulation. This approach is not completely lacking in regulatory norms, but emphasizes risk control through principled guidelines and industry self-regulation within existing legal frameworks, striving to avoid premature or overly stringent specialized legislation that might impede technological innovation.

Firstly, regarding federal legislation, although AI regulation has garnered joint attention from legislators of both parties, Congress has hardly enacted any legislative proposals specifically concerning AI regulation in recent years. Currently, passing comprehensive AI regulation legislation in Congress remains a challenge, as neither party has reached a broad consensus on either the content or the process. Nevertheless, in the realm of online information content governance, reliance on the extended application of existing institutions remains possible. For instance, Section 230 of the Communications Decency Act (CDA) establishes the principle of "platform immunity," which has long provided online platforms with broad liability exemptions. In the field of copyright, the "safe harbor" principle established by the Digital Millennium Copyright Act (DMCA) also offers liability exemptions for specific types of online platforms. Although there is still controversy over whether generative AI platforms naturally apply these provisions, it can be seen that US legislation has always maintained a relatively modest position in regulating online information content.

Secondly, the relevant practices of the federal government in regulating artificial intelligence are gradually unfolding, but they are constantly changing.. In October 2023, President Biden signed the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, which stipulated measures concerning "ensuring the safety and security of AI technology," "promoting innovation and competition," "supporting workers," "advancing equity and civil rights," "protecting consumers, patients, passengers, and students," "protecting privacy," and "strengthening U.S. leadership abroad." This order was once regarded as a landmark event signaling the beginning of strengthened AI regulation in the United States.[27] However, with the ascent of Trump, most of the Biden administration's AI regulatory policies were rescinded, shifting towards a stance that emphasizes development over regulation. Given the lack of broad consensus between the two major U.S. parties on AI regulation, if there is another change of government, it is likely that the existing artificial intelligence policies of the Trump administration will not be able to be continued.

Thirdly, current AI governance in the United States remains heavily reliant on industry self-regulation and corporate commitments. Technology companies represented by OpenAI, Google, and Microsoft primarily promote the gradual formation of industry self-regulatory norms by issuing policies on the use of generated content, transparency reports, and safety frameworks. The government has also been encouraging enterprises to make relevant self-regulatory commitments. In July and September 2023, the Biden administration consecutively convened fifteen technology companies, including Amazon, Anthropic, Google, Inflection, Meta, Microsoft, OpenAI, Adobe, and IBM, to make voluntary commitments, including conducting internal and external red-teaming tests on models or systems; sharing information regarding risks between enterprises and the government; investing in cybersecurity and internal

[27] Biden J. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence[EB/OL]. (2023-10-30)[2026-03-09]. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

threat protection measures; encouraging third parties to discover and report issues and vulnerabilities; developing and deploying mechanisms that enable users to discern whether audio or video content is AI-generated, including reliable sourcing, watermarking, or both for AI-generated audio or video; publicly reporting the capabilities, limitations, and appropriate and inappropriate use cases of models or systems; prioritizing research into the social risks posed by AI systems; and developing and deploying frontier AI systems to help address society's greatest challenges.[28]

(II) European Union: Security-Oriented Strong Regulation

Distinct from the US emphasis on developmental innovation, the European Union highlights risk prevention and the protection of fundamental rights in AI governance, forming a security-oriented model of strong regulation. The EU proposed the draft Artificial Intelligence Act as early as 2021, and formally adopted the Act in March 2024, making it the world's first comprehensive AI legislation. The AI Act lays the foundational framework for the EU's AI regulatory regime. Regarding the object of regulation, the Act defines an AI system as "a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers from the input it receives how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments." In terms of regulatory methodology, it primarily adopts a risk-based tiered regulatory approach, distinguishing between prohibited AI practices, high-risk AI systems, limited-risk AI systems, and minimal-risk AI systems. It imposes varying degrees of regulatory measures on AI systems of different risk levels and includes specific provisions for general-purpose AI. As for the regulatory agency, the European AI office was established to supervise AI at the EU level and promote the implementation of the AI Act.

In terms of regulating generated content, the EU first places particular emphasis on transparency and traceability. For instance, it requires clear labeling of deepfake content to prevent public misleading, and mandates compliance explanations regarding the sources of training data for large models, especially strengthening requirements for the legality of data usage in the field of copyright protection. In addition, the EU also emphasizes ex-ante compliance. High-risk systems must complete conformity assessments before being placed on the market and are subject to continuous market surveillance. Although this approach has not fully implemented a licensing system in the field of generative AI, it strengthens the ex-ante responsibilities of platforms through methods such as risk assessment, technical documentation filing, and regulatory review.

This strong regulation adopted by the EU is aiding it in gaining greater influence and discourse power in AI regulation, forming a "Brussels Effect" in this domain. The "Brussels Effect" refers to the process by which the EU effectively globalizes its unilateral regulation through market mechanisms, extending its laws beyond its borders. The EU has consistently been at the forefront of constructing digital governance rules. Prior to the AI Act, the EU had already enacted a series of significant legislations including the General Data Protection

[28] The White House. Fact Sheet: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI[EB/OL]. (2023-07-21)[2026-03-09]. https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/; The White House. Fact Sheet: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI[EB/OL]. (2023-09-12)[2026-03-09]. https://www.whitehouse.gov/briefing-room/statements-releases/2023/09/12/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-eight-additional-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.

Regulation (GDPR), the Regulation on the Free Flow of Non-Personal Data, the Open Data Directive, the Digital Content Governance, the Digital Markets Act (DMA), the Digital Services Act (DSA), and the Data Act, exerting significant influence in the field of digital governance and forming a "EU Model" of digital governance. In the realm of AI governance, the EU is continuing many characteristics of this model, creating a stark contrast with governance models in other regions and countries, particularly the United States. This strict regulatory model primarily employs "hard law" methods that impose statutory obligations and liabilities on relevant subjects, backed by state coercive power, which can effectively urge relevant subjects to fulfill their obligations and assume responsibility. However, if the imposed obligations and liabilities are too burdensome, or if the rules are overly one-size-fits-all, it may adversely affect the development of AI.

(III) China: A Middle Path Balancing Development and Security

In the field of AI governance, China has explored and formed a middle path situated between the approaches of the United States and the European Union. China's regulation neither relies heavily on market self-discipline in the US style nor fully adopts the comprehensive legislation of the EU style. Instead, it establishes specialized norms at the level of administrative regulations first, strengthening the regulation of specific AI risks while encouraging technological innovation and industrial development.

Against the backdrop of the rise of generative AI, the Cyberspace Administration of China issued the "Provisions on the Administration of Deep Synthesis of Internet Information Services" in 2022 and the "Interim Measures for the Management of Generative Artificial Intelligence Services" in 2023, specifically regulating behaviors related to providing generative AI services to the public. Regarding the governance of online information content, these regulations mainly demonstrate developments in two aspects. First, in the generation phase of online information content, they refine the obligations of relevant subjects to prevent the generation of illegal and undesirable information. The "Provisions on Deep Synthesis" mainly regulate three types of subjects, namely, deep synthesis service providers, deep synthesis service technical supporters, and deep synthesis service users. The "Interim Measures" mainly distinguish between two types of subjects, which are generative AI service providers and generative AI service users. These regulations make detailed provisions for various sub-stages related to the generation of online information content. Taking the obligations of service providers as an example, the provisions in these regulations cover their duties in data management, content management, user management, and other sub-stages. In terms of data management, service providers shall take effective measures during the selection of training data to prevent discrimination; shall carry out training data processing activities such as pre-training and optimization training in accordance with the law; and if data labeling is conducted, shall formulate labeling rules, conduct quality assessments of data labeling, and train, supervise, and guide labeling personnel. In terms of content management, service providers shall adopt technical or manual methods to audit inputs and generated results, and shall take disposal measures upon discovering illegal and undesirable information. In terms of user management, service providers shall take disposal measures against relevant service users upon discovering illegal and undesirable information.

Second, in the dissemination phase of online information content, new obligations are imposed on relevant

subjects to label AI-generated content. Articles 16 to 18 of the "Provisions on Deep Synthesis" establish the basic rules for service providers to label AI-generated content, including: (1) General obligation for labeling. Service providers shall add labels to their generated content that do not affect user usage. (2) Special obligation for prominent labeling. If service providers provide specific deep synthesis services that may cause public confusion or misidentification, they shall conduct prominent labeling in reasonable positions or areas of their generated content. (3) Protection of labels. No organization or individual shall destroy the labels on AI-generated content. In March 2025, the issuance of the "Measures for the Labeling of AI-Generated Synthetic Content" further improved China's system of obligations for labeling AI-generated content. This measure makes centralized provisions on the labeling of AI-generated content, clearly distinguishing between explicit and implicit labeling forms, and specifies the labeling obligations of different subjects such as service providers and users in the links of content generation, dissemination, and platform management. To support the implementation of this measure, China also formulated the national standard "Cybersecurity Technology: Methods for Labeling AI-Generated Synthetic Content" (GB45438—2025), providing specific guidance on how to add labels.

In summary, ranging from the US's development-oriented soft regulation, to the EU's security-oriented strong regulation, and then to China's institutional exploration balancing development and security, the global regulation of generative AI content presents a pattern of parallel diverse paths. These different approaches reflect differences in value orientations, legal traditions, and industrial structures across various jurisdictions, also laying the foundation for future international coordination and rule convergence.

IV. Recommendations for Improving the Governance of AI-Generated Content

(I) Content Identification: Synergistic Development of Technology and Institutions

Facing the challenges of identifying AI-generated content, it is necessary to promote synergistic development at both the technological and institutional levels to further improve the governance mechanism for produced content.

Firstly, the construction of a technical system for identifying generated content should be strengthened. On one hand, research and development of identification technologies must be promoted, supporting relevant institutions in studying technologies for identifying generated content. Special emphasis should be placed on tackling difficulties in this area, including: strengthening research on text generation identification technology, focusing on improving its accuracy, exploring more robust labeling methods for generated text adaptable to short texts, and enhancing the generalization of post-event detection technologies for generated text identification; continuously tracking the development of adversarial identification technologies, and studying how to enhance the robustness of identification technologies to prevent the destruction of generated content labels and the evasion of detection technologies. On the other hand, supporting service mechanisms for identification technologies should be improved by establishing public platforms to facilitate user queries regarding known watermarked output models and their identification services, or by having platforms provide unified identification services for different models. Initiatives should be launched at the international level to formulate standards for identification technologies and labeling and to establish corresponding organizations, thereby expanding China's discourse power and influence in standard-setting and international organizations.

Secondly, the institutional framework for disclosure obligations regarding generated content should be perfected. On one hand, rules concerning labeling obligations for generated content need to be refined. Different scenarios should be distinguished, considering factors such as the difficulty of content labeling, the impact on user experience, and the magnitude of legal risks. Through departmental regulations or industry standards, specific requirements for the labeling obligations of service providers in different scenarios should be further detailed. Rules for handling misidentification must be clarified, granting stakeholders of the identified objects the right to object to identification results, and refining procedural rules for objection handling. For instance, if the parties fail to reach an agreement, stakeholders may file a lawsuit, and the court shall render a judgment; the identifying party shall handle the matter based on the court's determination. Responsibility for erroneous identification must be defined; if the identifying party is at fault, it shall bear liability for compensating damages caused to stakeholders due to the identification error. On the other hand, based on the existing labeling obligation system, a comprehensive disclosure obligation system for generated content needs to be constructed. The content of disclosure obligations should expand from "whether AI was used" to "how AI was used," and the methods should broaden from a single labeling obligation to a dual obligation of labeling and explanation. Furthermore, disclosure obligations for different subjects should be reasonably allocated based on information costs.

(II) Platform Governance: Rule Construction for New-Type Platforms

The emergence of new-type platforms, such as generative AI service providers, provides an opportunity to update platform regulation rules. It is necessary to

explore regulatory rules for these new-type platforms at the institutional level.

First, the legal positioning of generative AI service providers should be clarified. Based on existing types of network service providers, an independent category of "generative AI service provider" should be added to distinguish them from traditional types of network service providers. Specific regulations governing generative AI service providers should be stipulated through the enactment of specialized legislation or regulations.

Second, classified regulation should be implemented based on the information management capabilities of service providers. For example, generative AI service providers rely on models from different sources and serve different objects, which affects their information management capabilities. Based on the source of the models relied upon, service providers can be categorized into those based on self-developed models, those based on secondarily trained models, and those based on third-party models. These three types of service providers differ significantly in their management capabilities regarding training data and models. Service providers based on self-developed models possess strong management capabilities over training data and models, and thus may bear a higher duty of care regarding training data selection and model training. Service providers based on secondarily trained models generally only have strong management capabilities over the secondary training process, and thus primarily bear corresponding duties of care regarding the secondary training process. Service providers based on third-party models have the weakest management capabilities over training data and models and cannot take prevention and disposal measures related to training, because they typically do not participate in model training. Additionally, based on the objects served, service providers can be divided into those serving application developers and those serving end-users. These two types differ significantly in their capabilities to manage user behavior. Service providers offering Application Programming Interfaces (APIs) to application developers usually find it difficult to monitor the behavior of end-users within applications. In contrast, providers directly serving end-users have greater capacity to directly supervise and manage users. Therefore, they should bear a higher duty of care in this regard and ought to take more measures to prevent and dispose of infringing behaviors by users.

Third, rules regarding the burden of proof and information disclosure for platforms should be improved. Due to the black-box nature of generative AI, generative AI service providers should bear a certain degree of burden of proof and information disclosure obligations, including disclosing records of generation logs and preserving prompts and out-put content to facilitate liability determination in the event of disputes. If a platform refuses to provide necessary information, its fault may be legally presumed.

(III) Liability Allocation: Liability Rules for Content Damages

First, when allocating liability among multiple subjects, it is necessary to combine specific circumstances and consider the causal force of different subjects' behaviors in generating infringing content to determine the subject liable for bearing responsibility. For example, if the cause of generating infringing content is that the service provider included relevant infringing content in the training data and failed to take effective preventive measures, resulting in the AI generating infringing content even when the user inputs prompts with low relevance, then the subject of the infringing act should

be the service provider. Conversely, if the generation of infringing content is primarily due to the user's input containing relevant infringing content, causing the AI to memorize and reproduce the content from the input during output, then the subject of the infringing act should obviously be the user. Furthermore, if the service provider fails to take effective preventive measures and the user's prompt also contains infringing content, such that the generated content results from the combined actions of both the service provider and the user, then it can be considered that the service provider and the user constitute joint infringement.

Second, considering that liability allocation may involve different scenarios, the principle of scenario-based analysis should be followed. This involves evaluating the costs and benefits of specific countermeasures in preventing risks under different scenarios to select the liability allocation strategy that optimizes social efficiency. For instance, in China, courts often consider various factors to determine whether a service provider is at fault in a specific scenario. Judicial interpretations in China have previously stipulated rules for determining the fault of network service providers, listing various factors that should be considered when determining fault to guide courts in making judgments in specific cases. Although current Chinese laws, regulations, and judicial interpretations have not yet made similar provisions regarding the determination of fault for generative AI service providers, in judicial practice, some courts have referred to the methods for determining fault of network service providers for analysis. For example, in the Hangzhou Ultraman Case, the court held: "Regarding the rules for determining fault, one should comprehensively consider factors such as the nature of generative AI services, the current level of AI technology development, the feasibility and cost of alternative designs to avoid damage, the necessary measures that can be taken and their effects, and the impact of assuming tort liability on the industry. By dynamically adjusting the standards for determining fault, the duty of care of the platform should be controlled at a reasonable level."[29] It can be seen that the methods used to determine fault on the part of network service providers provide useful experience for determining fault on the part of generative AI service providers, enabling courts to make reasonable determinations in light of different scenarios.

[29] Hangzhou Intermediate People's Court of Zhejiang Province. Civil Judgment, (2024) Zhe 01 Min Zhong No. 10332.