

# Legal Commentary

May 11, 2023

## Highlights of the Draft Measures for the Administration of Generative Artificial Intelligence Services

Author: Kevin DUAN | Kemeng CAI | Yi ZOU<sup>1</sup>

Generative AI has become a worldwide sensation recently with the launch of ChatGPT, Stable Diffusion, Midjourney, and other eye-catching products. Large language models such as ChatGPT have shown their phenomenal capacity for human language comprehension, human-machine interaction, text writing, programming, reasoning, etc. by generating output that is often on a par with the level of human intelligence, if not better. Despite this, however, the use of generative AI has raised concerns about privacy violations, trade secrets leakage, misinformation, information cocoons, cybercrimes, and other potential risks, which has aroused global attention and regulatory responses in different countries and regions. For instance, Garante, Italy's data privacy watchdog, has imposed a nationwide ban on using ChatGPT due to privacy violation concerns. Following this trend, on April 11, 2023, the Cyberspace Administration of China ("CAC") issued an exposure draft of the *Measures for the Administration of Generative Artificial Intelligence Services* (the "**Draft Measures**"), which is open for public comments until May 10, 2023. The Draft Measures, consisting of 21 articles, begin by clarifying its administrative objectives of facilitating healthy development and regulated application of generative artificial intelligence services, leaving space for further policies to regulate the development and use of generative AI. In this commentary, we summarize and comment on key aspects covered by the Draft Measures, the challenges they may pose in practice, and our suggestions to address those challenges.

### Scope of application: services to "the public in the territory of the [PRC]"

The scope of application of the Draft Measures is presented in Article 2, which is: "the research, development, and utilization of generative AI products to provide services to the public in the territory of the [PRC]". For the purpose of the Draft Measures, "generative AI" refers to "technologies that build on algorithms, models, and rules and are used for producing texts, images, audio, videos, codes, and other forms of content". Therefore, the Draft Measures, once adopted, will apply to popular generative AI products such as ChatGPT, Google Bard, Stable Diffusion, and Midjourney, as well as large language models rolled out by Chinese tech giants. A question invited by this provision is how to understand the

<sup>1</sup> Han Kun intern Yuxin XIANG also contributed to this legal commentary.

part of “providing services to the public in the territory of the [PRC]”. In our opinion, considering the meaning of the provision *per se* and the overall legislative purpose, the Draft Measures should apply to all providers that offer generative AI services to customers in China, regardless of whether the providers themselves are located in or outside China, and regardless of whether they provide services directly to end users or indirectly by linking to services from other carriers.

## AIGC regulation

The Draft Measures also emphasize the regulation of AI-generated content (“**AIGC**”) and ideological security, with a focus on the following aspects.

- **Service providers are responsible for AIGC security.** The Draft Measures stress that organizations and individuals (i.e., service providers) bear responsibility as the AIGC producer where they offer chat services and text, **image**, or audio generation or similar services by using generative AI products. In reality, however, a user may nonetheless find a means to create illegal or harmful content by using generative AI services. As such, it is debatable whether it is fair to hold a service provider responsible for AIGC outputs in all instances.
- **Generated content should be truthful and accurate.** Controversy has arisen regarding the requirement in the Draft Measures that “the content created by generative AI must be truthful and accurate, and measures shall be taken to prevent the generation of false information”. At present, it seems unavoidable that, sometimes, large language models such as ChatGPT deliver “confident nonsense”, a phenomenon known as an AI hallucination, which may derive from technological limitations such as divergences in the source content and errors in decoding by the transformer. Therefore, an overemphasis on the truthfulness and accuracy of generated content may impose onerous duties on service providers.
- **Measures to curb violative content.** Article 15 would require service providers to counteract generated content that is found to violate the Measures by means such as content screening, as well as retraining of the AI generator model for optimization within three months to prevent reproduction of such content. However, in practice, there might be hurdles to implement this provision given the existing technological bottlenecks which make it difficult to identify the origin of the violative content and to retrain the model in question to prevent such violative content.

In addition to model optimization, the Draft Measures impose more conventional, *ex post* obligations on service providers to curb violative content, which include: (1) taking measures to stop the generation of any text, image, audio, video or other content that, to their awareness or knowledge, has infringed upon others’ portrait rights, reputation rights, personal privacy, trade secrets, or that has violated any requirement of the Measures, as a way to discontinue the harm caused thereby; (2) suspending or terminating services to users who they find have violated relevant laws and regulations, business ethics or social morals in the course of using their generative AI products, namely users who have committed acts such as social media hyping, malicious posting and commenting, spam creation, malware programming, and improper business marketing.

- **Labelling requirements.** Article 16 of the Draft Measures would require service providers to label generated images, videos, and other content in accordance with the *Provisions on Administration of Deep Synthesis of Internet-based Information Services* (“**Deep Synthesis Provisions**”), though the Draft Measures would not expressly require the labelling of AI generated texts, as is prescribed in the Deep Synthesis Provisions.

## Training data compliance

The quality of training data is essential to ensure the accuracy and integrity of AIGC and to avoid AI discrimination and bias. Given that, the Draft Measures would hold service providers responsible to ensure that the data used to pre-train and retrain their generative AI models are obtained from legitimate sources, and impose detailed requirements for training data compliance in the following aspects.

- **Personal information protection.** According to the Draft Measures, where personal information is used for pre-training or retraining a generative AI model, service providers must obtain the consent of the personal information owner, or, under other circumstances, comply with requirements of applicable laws and administrative regulations. Specifically, pursuant to the Draft Measures, service providers must obtain consent from users for using their personal information to pre-train or retrain the relevant generative AI models, or, in any other circumstance, they must comply with requirements as prescribed in applicable laws and administrative regulations. In addition, service providers are prohibited from illegally retaining input data which can be used to infer users’ identities. Service providers are also banned from profiling based on users’ inputs and their use of the services, nor may they provide users’ inputs to any other party.
- **No infringement of intellectual property.** The Draft Measures would require that the data used for AI training must not contain any content that infringes upon intellectual property rights. This requirement may cause disagreements in practice. Training data used for developing and improving generative AI models are usually scraped from open sources on the internet, which inevitably involve many copyrighted works. It is currently a highly controversial issue worldwide as to whether the use of copyrighted works for algorithm training infringes the right of the copyright owner or whether it falls within the ‘fair use’ exception. Some argue that restricting the use of copyrighted works for AI training may significantly compromise the quality and diversity of training data. Therefore, striking a balance between the interests of the copyright owner and the service provider remains a question to be discussed at both theoretical and policy levels.
- **Training data must be truthful, accurate, objective, and diverse.** This requirement in the Draft Measures would also pose great challenges for service providers when selecting AI training data.

## Improvement of existing rules on recommendation algorithm-based services and deep synthesis services

The Deep Synthesis Provisions define “deep synthesis technology” as that which “employs deep learning, virtual reality, and other synthetic algorithms to produce text, images, audio, videos, virtual scenes, and other online information”, including but not limited to technologies used for text generation, text-to-speech

conversion, music creation, face generation, image generation, as well as 3D reconstruction, digital simulation and other technologies that create or edit 3D characters and virtual scenes. The *Provisions on the Administration of Algorithm-generated Recommendations for Internet Information Services* (“**Recommendation Algorithm Provisions**”) also expressly include “generative and synthetic” algorithms into its scope of application. This means that generative AI, which by definition constitutes both a “deep synthesis technology” and a “recommendation algorithm-based service”, also falls under the umbrella of the aforesaid AI regulations concerning deep synthesis technologies and recommendation algorithm-based services. Given that, the Draft Measures would incorporate and improve upon these existing rules in the following aspects.

- **Ethics and fairness of algorithm-based services.** The Draft Measures reiterate and would refine the Recommendation Algorithm Provisions and other rules that are in place to promote algorithm ethics and fairness and avoid algorithm-related discrimination. The Draft Measures stress that providers of generative AI products and services should “take measures to prevent discrimination on the basis of race, ethnicity, belief, nationality, region, gender, age, occupation, etc. in the process of algorithm design, training data selection, model generation and optimization, and service provision”, should “respect intellectual property rights and business ethics and not engage in unfair competition by using advantages such as algorithms, data, and platforms”, and should not “generate discriminatory content based on the race, nationality, gender, etc. of their users”.
- **Security assessment and registration for algorithm-based services.** Article 6 of the Draft Measures stipulates that, prior to the provision of services to the public by using generative AI products, service providers must conduct and report on security assessment to the competent cyberspace administration in accordance with the *Provisions on the Security Assessment for Internet-based Information Services with Public Opinion Attributes and Social Mobilization Capability*, and shall complete procedures for registration, change of registered particulars, and deregistration (as applicable) of services by following the Recommendation Algorithm Provisions. Based on the above provision, all generative AI products would be deemed as information services “with public opinion attributes and social mobilization capacity”, and thus be subject to security assessment and registration requirements under the applicable laws.
- **Transparency of algorithms.** According to the Draft Measures, service providers must, as required by the CAC and relevant competent authorities, provide necessary information that may affect users’ trust in and choice of the relevant services, including description of the source, scale, type and quality, etc. of pre-training and retraining data, rules for manual labelling, the scale and type of manually labelled data, basic algorithms, and technical systems, among others.
- **Disclosure requirements and measures to prevent addiction.** Article 10 of the Draft Measures would require service providers to specify and disclose the intended users, occasions, and purpose of their services and to take proper measures to prevent users from over-relying on and becoming addicted to generated content. This provision, placed in tandem with Article 8 of the Recommendation Algorithm Provisions which prohibits service providers from “setting up algorithms to induce users toward addiction or excessive consumption”, requires service providers

to ensure proper use of relevant products on various fronts from public disclosure to algorithm management.

### **Conclusion: impact and outlook**

According to Article 20 of the Draft Measures, violations of the Measures may be punished pursuant to the *Cybersecurity Law of the PRC*, *Data Security Law of the PRC*, *Personal Information Protection Law of the PRC*, and other applicable laws and regulations. Where a violation is not covered by the abovementioned laws and regulations, the service provider concerned may be given a warning, subject to public criticism, or be ordered to make rectification within a time limit; the service provider may even be ordered to suspend or terminate its use of generative AI for service provision and be subject to a fine of up to RMB 100,000. Behaviors in violation of administrative rules for public security will be subject to punishment in accordance with law, and behaviors that constitute criminal offences will be subject to criminal liability.

On the whole, by issuing the Draft Measures, Chinese regulators have directly responded to new issues posed by the recent generative AI breakthroughs under the current regulatory framework, which also conveys China's overarching AI regulatory principle of providing guidance and rules for the purpose of promoting growth of the industry. Nevertheless, the Draft Measures would impose some compliance requirements that seem to be difficult to implement in practice given current technological bottlenecks. Therefore, companies should consider responding with creative solutions by wisely integrating technology and law, so as to help assuage regulators' security concerns and create more policy space for further development of the generative AI industry.

## ***Important Announcement***

This Legal Commentary has been prepared for clients and professional associates of Han Kun Law Offices. Whilst every effort has been made to ensure accuracy, no responsibility can be accepted for errors and omissions, however caused. The information contained in this publication should not be relied on as legal advice and should not be regarded as a substitute for detailed advice in individual cases.

If you have any questions regarding this publication, please contact:

### **Kevin DUAN**

Tel: +86 10 8516 4123

Email: [kevin.duan@hankunlaw.com](mailto:kevin.duan@hankunlaw.com)