

Kluwer Copyright Blog

Generative AI Services in China: Public Consultation on the Regulation for Generative Artificial Intelligence Services

Baiyang Xiao (University of Szeged) · Monday, May 22nd, 2023

The booming industry of generative artificial intelligence (AI) is facing its first regulatory attempt in China. On April 11th, the Cyberspace Administration of China released a [draft of the Regulation for Generative Artificial Intelligence Services](#) (the ‘draft Regulation’) for public consultation, which includes 21 articles detailing the proposed regulatory framework for the generative AI industry. The draft Regulation outlines several core concerns surrounding generative AI, including content control, prevention of discrimination, intellectual property (IP) protection, curbing misinformation and data protection.



Image by Gordon Johnson via Pixabay

In general, the draft Regulation endeavors to ensure that the content produced using generative AI services adheres to legal and ethical principles and aligns with core socialist values. Among many other things, such as security, social morality, unfair competition, data accuracy and privacy, the draft Regulation turns to two specific copyright aspects that are addressed in Article 7, namely providers’ liability for IP violations, and access to copyright-protected materials for data training.

1. An overview of the draft Regulation

The draft Regulation supports and encourages the growth of the generative AI industry, while imposing strict burdens and liabilities on providers of generative AI services. Article 3 emphasizes the state’s support for independent innovation, promotion and application of AI algorithms, frameworks, and other basic technologies, as well as international cooperation. Additionally, the state encourages the prioritization of secure and trustworthy software, tools, computing, and data resources. Moreover, the draft Regulation includes a brief definition of generative AI as the technology of generating text, images, sound, video, code, and other content based on algorithms, models, and rules (Article 2 (2)). Furthermore, the draft Regulation places a significant emphasis on safeguarding personal data and information by imposing on providers the obligation to protect

the input information of users and usage records during the service provision process. In particular, when the data used in generative AI products or services contain personal information, the provider must obtain the consent of the personal information holder or comply with other [relevant laws or regulations](#) (Article 7(3)). In addition, illegal retention of input information that can be used to infer user identities, creation of user profiles based on input information and usage, and disclosure of user input information to third parties are strictly prohibited (Article 11). Finally, providers who violate terms of this regulation will be subject to penalties, such as warnings, criticisms, orders to rectify the situation, and fines by the Cyberspace Administration and relevant competent departments (Article 20).

2. Stringent compliance requirements for providers of generative AI services

The draft Regulation imposes a strict liability rule on providers of generative AI products or services, including those who support others to generate content through programmable interfaces.

The draft rules hold that providers are considered content producers and are thus responsible for all content generated (Article 5), and IP rights must not be violated in the provision of generative AI services (Article 4). To avoid potential IP-infringing output, providers have to moderate the generated content by adjusting the algorithm or employing a reliable ex-post content filtering mechanism to conduct an effective review of output (Article 15).

The draft Regulation also adopts a controversial article that requires providers of generative AI products or services to ensure the legality and accuracy of both the content generated and the sources of pre-training and optimization training data. To avoid potential liability, providers should ensure that the pre-training and optimization training data used for generative AI products meets the following requirements: (1) it complies with the requirements of laws and regulations such as the *Cybersecurity Law*; (2) it does not contain content that infringes on IP rights; (3) if the data contains personal information, the consent of the holder of the personal information shall be obtained or comply with other statutory circumstances; (4) the authenticity, accuracy, objectivity, and diversity of the data can be guaranteed; and (5) it complies with other regulatory requirements (Article 7). Obviously, the above stipulations greatly restrict the amount of data available for training generative AI. In compliance with these regulations, Chinese providers have to invest significant resources on rights clearance and ex-ante filtering when exploiting web-scraped datasets that include large amounts copyrighted materials.

For non-complaint generative content discovered by providers or reported by users (users may report unlawful content to relevant departments (Article 18)), in addition to content filtering and other measures, the provider should prevent the content from being generated again through optimization training of the algorithm or other methods within three months (Article 15). Furthermore, generated content should be marked as having been generated by AI in accordance with the [other regulations](#) (Article 16).

In addition to the aforementioned obligations, providers of generative AI services have to take measures to monitor and moderate undesirable content identified by themselves or reported by users. On the one hand, providers should take measures to stop generating content that violates others' legitimate rights (second sentence of Article 13). On the other hand, if providers identify that certain users violate laws, regulations, commercial ethics and social morality while using

generative AI services, they should suspend or terminate the disputed services (Article 19). However, the draft Regulation requires providers to implement a user complaint mechanism to handle requests regarding personal information in a timely manner, while leaving other claims unmentioned (first sentence of Article 13). The excessively broad scope of monitoring and narrow complaint mechanism may lead to the over-removal of lawful content in the name of copyright protection, thus adversely impacting users' freedom of expression.

3. Access to copyright-protected materials for data training

Generative AI routinely gathers public data from the internet for model training, including substantial volumes of IP-protected content, notably copyrighted materials. There is considerable controversy over whether using such protected works for model training infringes on IP rights and whether it is subject to limitations and exceptions, both domestically and internationally. The EU has adopted two [TDM exceptions](#) that allow inter alia researcher organizations and commercial entities to perform text and data mining (TDM) activities on materials they have lawful access to. Notably, the exception that allows commercial organization to carry out TDM in Article 4 of the CDSM Directive is [subject to express reservation](#) by right holders in an appropriate manner. Meanwhile, [the UK also recognizes](#) that restricting data access for training sets could disadvantage and impede domestic development of generative AI, and suggests collaboration between the government and the AI and creative industries to facilitate TDM for any purpose, and to include the use of publicly available content including that covered by IP as an input to TDM (including databases).

In contrast to those relatively TDM-friendly copyright rules or approaches, the limited statutory categories of limitations and exemptions in Article 24 of [Chinese Copyright Law](#) do not cover TDM exceptions. Article 24 (13) further indicates that whether TDM should be subject to limitations and exceptions should be clarified through the legislative process rather than left to be determined by the courts on a case-by-case basis. Meanwhile, due to excessively high administrative costs, incorporating TDM within the scope of statutory licensing appears problematic and suboptimal. In addition, statutory licensing TDM may fall short as copyright holders may preclude usage in TDM by express reservation. The draft Regulation specifies that providers are responsible for the legality of the data used for training and shall ensure the data is free of IP-infringing content. It seems impractical to demand that providers only use data from legal sources for data training because they are unable to evaluate the legality of massive input data collected from the Internet. Thus, an ex ante license should be obtained from the right holder for the use of training data in order to avoid IP-infringing content in the data.

However, the process to obtain such a license for input data still remains unclear. . One possible solution could be collective licensing, as proposed recently proposed by Authors Guild in the U.S.. Concretely, the Authors Guild [proposed possible legislative changes](#) to regulate generative AI, such as enabling various collective licensing models for AI training that would entitle authors to earn licensing income if their works are used to train generative AI. It is doubtful whether such an approach would work in China, as the massive number of licenses for input data might pose a huge challenge for the poorly designed state-controlled system of copyright collective management.. Further guidance is needed to clarify the legality of pretraining and optimization training data and to support AI companies in accessing copyright-protected materials for use in training data.

The Draft regulation underscores the intricate and challenging task of constructing a prompt regulatory framework for Generative AI services to mitigate potential legal risks without stifling innovation within this emerging field. The onerous requirements could potentially discourage domestic providers and international counterparts from entering the Chinese market, thus, necessitating close observation by all interested stakeholders.

To make sure you do not miss out on regular updates from the Kluwer Copyright Blog, please [subscribe here](#).

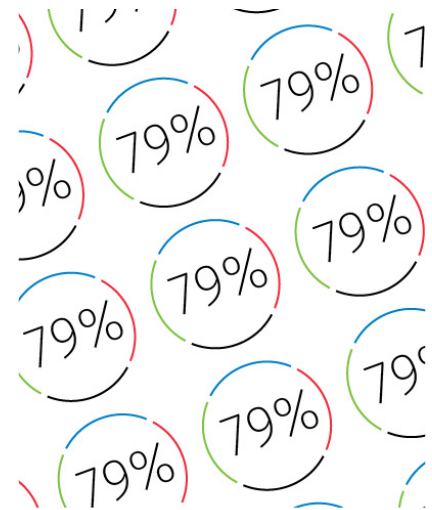
Kluwer IP Law

The **2022 Future Ready Lawyer survey** showed that 79% of lawyers think that the importance of legal technology will increase for next year. With Kluwer IP Law you can navigate the increasingly global practice of IP law with specialized, local and cross-border information and tools from every preferred location. Are you, as an IP professional, ready for the future?

Learn how **Kluwer IP Law** can support you.

79% of the lawyers think that the importance of legal technology will increase for next year.

Drive change with Kluwer IP Law.
The master resource for Intellectual Property rights and registration.



2022 SURVEY REPORT
The Wolters Kluwer Future Ready Lawyer
Leading change

This entry was posted on Monday, May 22nd, 2023 at 2:18 pm and is filed under [Artificial Intelligence \(AI\)](#), [Legislative process](#), [Liability](#), [Text and Data Mining \(TDM\)](#)

You can follow any responses to this entry through the [Comments \(RSS\)](#) feed. You can leave a response, or [trackback](#) from your own site.

