

AIGC

The CAC Releases Draft Regulations and Standards for Labeling AIGC

With the rapid advancement of AI technology, generative synthesis tools have become more realistic, increasingly accessible and user-friendly. As a result, the ability to create and disseminate deepfake content at low cost has blurred the lines between reality and the virtual world, posing challenges to the traditional distribution of information on the internet. In September, China Central Television (CCTV) reported that criminals in various regions of China had used “face swapping” technology to forge indecent photos and manipulate victims’ selfies from social media for blackmailing purposes.

The foundations of trust required for the distribution of information on the Internet needs to be urgently restored, and labeling AI generated or synthetic content will play a critical role in rebuilding this trust. In 2022 and 2023, the Cyberspace Administration of China (CAC) issued the *Administrative Provisions on Deep Synthesis of Internet Information Services* (“**Deep Synthesis Provisions**”) and the *Interim Measures for the Administration of Generative Artificial Intelligence Services*, which broadly outline the identification obligations for generative AI service providers. To further clarify labeling requirements, the CAC released the drafts of the *Measures for the Identification of Artificial Intelligence Generated and*

Synthetic Content (“**Draft Measures**”) and the supporting mandatory national standard *Cybersecurity Technology – Method for Identifying Artificial Intelligence Generated and Synthetic Content* (“**Draft Standard**”) on September 14, 2024. The deadline for public comments on the Draft Measures is October 14, 2024, and on the Draft Standard, November 13, 2024.

I. Scope of Application

Under the Draft Measures and the Draft Standard, service providers that utilize algorithmic recommendation technology, deep synthesis technology or generative AI technology to provide Internet information services within China (“service providers”) are the primary parties responsible for the labeling requirements. Online information and content distribution platforms, Internet application distribution platforms, and their users are also imposed obligations to monitor and assist with the implementation of the labeling obligations.

Enterprises and institutions that are engaged in the research and development or application of AI generative synthesis technology but do not provide related services to the public in China are not subject to these requirements.

II. Obligations of Service Providers

Depending on the nature of the service, service providers are obliged to apply explicit and/or implicit labels to AI generated or synthetic content. The key obligations under the Draft Measures are outlined below.

(a) Obligations to Use Explicit Labels

Explicit labels must be added to AI generated or synthetic content. Explicit labels must be presented in a form that users can clearly perceive, such as through text, sound, or graphics and be applied according to specific context.

- **Text content:** Services that simulate human-generated or human-edited text such as intelligent dialogue and intelligent writing should incorporate text labels or universal symbol tags. These labels should be placed at the beginning, end, or other suitable locations within the text. Alternatively, explicit labels should be added to the interface or positioned around the text in interactive scenes.
- **Human voice and sound imitation content:** For services that involve generating synthetic human voices or sound imitations, or that significantly alter identifiable personal characteristics, it is necessary to include voice or rhythm markers at the beginning, end, or other suitable points within the audio. Alternatively, explicit labels should be added to the interface in interactive scenes.
- **Face content:** Services that offer editing features, such as face generation, face swapping, face manipulation, gesture manipulation, or any other features that substantially modify identifiable personal characteristics in character images or videos, must prominently display an explicit label on the image or video. For videos, an explicit label should be added at the beginning of the video and in a suitable location within the video display area. Additional explicit labels may be included at the end of the video or at

appropriate intervals within the video.

- **Virtual scenes:** For services involving the generation or editing of immersive virtual reality experiences, it is required to prominently display an explicit label on the start screen. Additional, during the ongoing service of the virtual scene, an explicit label can be added at an appropriate location to ensure clear disclosure to users.
- **Other scenarios:** Other scenarios of generative AI services that may cause confusion or misunderstanding among the public should be marked with conspicuous labels that have a significant notification effect based on the specific features of the scenarios.

According to Article 17 of *Deep Synthesis Provisions*, if a deep synthesis service provider provides services other than those specified above, it shall provide an explicit labeling function and inform users of the deep synthesis service that such explicit labeling is available. We understand that these scenarios likely include the generation and synthesis of non-human voices, non-face images, videos, and similar content.

The Draft Standard provides comprehensive guidelines, including specific methods (such as location, font and format) along with examples, for explicit or implicit labeling that companies can follow in various scenarios. For instance, when labeling text content explicitly, it should be presented as text or badges and should incorporate the following elements: 1) an identification of artificial intelligence, such as “artificial intelligence” or “AI”; and 2) a reference to generative or synthesis, such as “generative” and/or “synthetic”. These explicit labels should be positioned at the beginning, end, or an appropriate point within the text. The font and color of the explicit labels must be clear and legible to ensure easy comprehension. Below is an example for text content.



Figure C.1 The explicit label at the beginning of the text

(b) Obligations to Use Implicit Labels

According to Article 16 of the *Deep Synthesis Provisions*, deep synthesis service providers are required to take technical measures to add labels that do not affect the user experience to content generated or edited using their services, and shall maintain logs in accordance with the law.

Implicit labels are markers embedded within the data of files containing AI-generated or synthetic content, applied through technical measures that are not easily visible to users. Implicit labeling can be achieved through various methods.

According to the Draft Measures, it is required that implicit labels must be added to the metadata of files containing AI generated or synthetic content. These implicit labels may include, but are not limited to, content attributes, the service provider's name or code, and the content identification number. Service providers are also encouraged to utilize digital watermarks as a form of implicit labeling when generating or synthesizing content.

The Draft Standard specifies the elements to be included in implicit labeling, which are as follows:

- A generative or synthetic tag, indicating the AI generative or synthetic attributes of the content. This tag is categorized as certain, possible, and

suspicious;

- The name or code of the generative and synthetic AI service provider;
- A unique identification number assigned to the content by the generative and synthetic AI service provider;
- The name or code of the content distribution service provider; and
- A unique identification number assigned to the content by the content distribution service provider.

Appendices E and F of the Draft Standard provide specific formats and examples for metadata implicit labeling.

(c) Obligations regarding User Management

Service providers are required to include in the user agreement the method, style and other relevant specifications for labeling AI generated or synthetic content and remind users to read and understand the relevant label management requirements.

In the event that a user requires the service provider to provide the AI-generated or synthetic content without explicit labeling, the service provider must specify the user's obligations regarding labeling and their responsibilities for usage in the user agreement. Furthermore, the service provider must retain relevant logs for a minimum of six months.

III. Obligations of Content Distribution Platform Service Providers

The Draft Measures stipulate obligations for online content distribution platforms regarding the review and verification of content distributed on their platforms particularly concerning implicit labels. The requirements are as follows:

- **Verification of Implicit Labels:** Platforms must check whether the file metadata contains implicit labels. If implicit labels are present, an explicit label must be prominently applied around the post to clearly inform users that the content has

been generated or synthesized by AI.

- **User Claims and Explicit Labeling:** In case no implicit label is detected in the file metadata, but the user claims that the content is AI-generated or synthetic, an explicit label must be prominently applied around the post to indicate to users that the content is likely AI-generated or synthesized.
- **Suspicion of AI Generation or Synthesis:** If no implicit label is found in the file metadata and no user claims the content to be AI-generated or synthetic, but the platform detects signs of AI generation or synthesis, such content can be identified as suspected AI-generated or synthetic. In this case, an explicit label must be prominently displayed to inform users that the content is suspected to be generated or synthesized by AI;
- **Addition of Implicit Label:** For confirmed, possible or suspected AI generated or synthetic content, the file metadata should include content attributes, the name or code of the content distribution platform, content identification number, and other information about the AI generative or synthetic element.
- **User Notifications:** Service providers must offer necessary labeling functions and remind users to actively declare whether their posts contain AI-generated or synthetic content.

IV. Obligations of Internet Application Distribution Platforms and Users

The Draft Measures require that, in connection with the release or review of applications, Internet application distribution platforms must verify whether service providers offer the required labeling functions for AI generated and synthetic content.

The Draft Measures do not explicitly provide detailed requirements for application distribution platforms to ensure compliance with labeling obligations by the applications distributed in these platforms, nor do they

specify the legal consequences for failing to verify non-compliance. However, we believe that application distribution platforms have an obligation to assist regulatory authorities by removing non-compliant applications from their platforms. This could include taking action against applications that fail to comply with the labeling requirements.

When users upload AI-generated or synthetic content to an online information/content distribution platform, it is expected that they voluntarily declare this content and utilize the labeling function provided by the platform.

V. Our Observations

In the processes of algorithm filing and security assessment of Internet-based AI services, compliance with the obligation to label AI-generated or synthetic content has become a critical item of scrutiny for regulatory authorities. The introduction of specific regulations and national standards on AI content labeling provides comprehensive guidelines for service providers. Content creators, online information/content distribution platform service providers, and application distribution platforms each have specific obligations regarding content labeling in the production and distribution of information.

We recommend all these service providers conduct self-assessments to ensure compliance with the existing regulations, as well as the Draft Measures and Draft Standard. It is anticipated that the labeling of AI-generated and synthetic content will increasingly become a focal point for law enforcement in the AI sector in the future.

While there is a general agreement among lawmakers worldwide regarding the importance of increasing transparency for AI-generated content, there is an ongoing debate on the best technical methods to achieve this, such as metadata and watermarks. These technologies are still evolving, and critical factors like their effectiveness, interoperability, and resistance to tampering need careful assessment.

The availability of reliable tools for users, content

distribution platforms, and regulatory agencies to identify synthetic content will greatly influence the implementation of the legislative requirements discussed above and the allocation of responsibilities

among various stakeholders. These matter require further research and continued observation.

Marissa DONG	Partner	Tel: 86 10 8519 1718	Email: dongx@junhe.com
Jinghe GUO	Associate	Tel: 86 10 8553 7947	Email: guojh@junhe.com
Ye LI		Tel: 86 10 8519 1324	Email: liye@junhe.com

This document is provided for and only for information sharing. Nothing contained in this document constitutes any legal advice or opinion of JunHe LLP. For more information, please visit our official website at www.junhe.com or our WeChat public account “君合法律评论” /WeChat account “JUNHE_LegalUpdates” .

