China Releases New Draft Regulations on Generative AI

May 30, 2024Posted by China BriefingWritten by Giulia InteresseReading Time: 6 minutes

China is striving to regulate generative AI while promoting innovation and technological advancement. The NISSTC has issued draft regulations outlining security measures for generative AI service providers, underscoring China's commitment to responsible AI development.

On May 23, 2024, the National Information Security Standardization Technical Committee (NISSTC) released new draft regulations titled *Cybersecurity Technology – Basic Security Requirements for Generative Artificial Intelligence (AI) Service* (hereinafter referred to as the "draft").

The draft, open for public comments until July 22, 2024, outlines several security measures for generative AI services. It covers important areas such as securing training data, protecting AI models, and implementing overall security protocols. It also provides guidelines for conducting security assessments.

In this article, we provide an overview of the comprehensive security requirements for

- **Security measures:** Specifies essential security measures to be implemented to effectively mitigate risks.

Additionally, in the context of the draft, the following key terms are clarified to ensure clarity and consistency:

- **Generative AI service:** Refers to services that utilize generative AI technology to produce various types of content like text, images, audio, and video for public consumption.
- **Service provider:** Refers to organizations or individuals who offer generative AI services through interfaces like interactive or programmable interfaces.
- **Training data:** Includes all data directly used to train AI models, covering both pre-training data and optimized training data.

The draft serves as a reference for both service providers and regulatory authorities. It offers guidance for conducting security assessments and establishing pertinent regulations.

Security requirements for training data

Before gathering data from specific sources, service providers must conduct a comprehensive security assessment. If a source contains more than 5 percent illegal or "harmful" content (explained in the next section), data collection from that source should

If training data from overseas sources is required, it should be reasonably combined with training data from domestic sources.

Requirements vary based on data collection method and type, as illustrated in the table below.

Security Requirements Based on Data Source and Collection Method

Training data source and collection method	Requirements
When using open-source training data	The open-source license agreement or relevant authorization documents for the data source should be obtained.
When using self-collected training data	› Collection records should be maintained. › Data explicitly prohibited from collection by others should not be collected.
When using commercial training data	When gathering user input for training purposes, providers must ensure users have control over their data: › Users should be provided with

contracts, cooperation agreements, etc., should be in place.

- If the transaction party or cooperating party cannot provide commitments and related proof of data sources, quality, security, etc., the training data should not be used.
- The training data, commitments, and materials provided by the transaction party or cooperating party should be audited.

When treating user input information as training data	Records of user authorization should be kept.

What kind of data will be deemed 'harmful'?

The draft categorizes "harmful data" into the following risk areas:

- **Violations of core socialist values:** Content that incites secession or undermines national unity and social stability is considered harmful. Extremism, in any form, including the promotion of terrorism, extremism, or ethnic hatred, is particularly dangerous and must be excluded from training data.

- **Commercial violations:** This refers to content that encompass infringements on intellectual property rights, breaches of business ethics, and the disclosure of commercial secrets.
- **Infringement of legal rights**: Harmful data also includes actions that infringe upon individuals' rights and well-being. This encompasses harming the physical or mental health of others and violating portrait rights. Moreover, defamation, infringement on personal honor, and breaches of privacy rights are significant concerns. Additionally, violating personal information rights and other legal rights of individuals must be rigorously avoided to uphold ethical standards in AI development.

How to ensure the non-harmfulness of the training data?

According to the draft, data content for all types of training data (e.g., text, images, audio, video) must be filtered before use. This can be done through keyword filtering, classification models, or manual inspection to remove illegal or harmful information.

With regard to intellectual property rights, data providers must implement a

For sensitive personal information, explicit consent must be obtained, or the use must comply with relevant legal or regulatory standards.

Security requirements for AI models

The draft highlights the importance of robust security measures throughout the entire lifecycle of generative AI model development and deployment. The following guidelines are proposed for each phase:

- **Model training:** Security should be prioritized as a key metric during training, and regular security audits should be conducted to identify and fix vulnerabilities, particularly in open-source frameworks.
- **Model output:** Technical measures should be implemented to align generated content with user intentions and mainstream understanding.
- **Model monitoring:** Inputs should be continuously monitored to prevent malicious attacks, and ongoing evaluation methods and emergency management measures should be established to promptly address security issues.
- **Model updates and upgrades:** A comprehensive security management strategy should be developed for updates or upgrades and security evaluations should be conducted post updates.

Security measures proposed in the draft

Understanding the applicability and scope of generative AI services	Service providers must justify the necessity and safety of employing generative AI within their service scope. This includes critical areas like: › Medical information services; › Financial transactions; and › Psychological counseling. **When catering to minors**, specific measures are imperative, including: › Allowing guardians to set anti-addiction measures; › Restricting access to paid services incompatible with minors' legal capacity; and › Actively promoting content beneficial to minors' physical and mental well-being. **For services not intended for minors**, proactive steps should be taken to prevent their access, either through technical solutions or effective management practices.

	about: › Limitations of the service; › Overview of models and algorithms used; and › Purpose of personal data collection. For services accessible via programmable interfaces, documentation should comprehensively cover the above information.
Handling user input responsibility	When gathering user input for training purposes, providers must ensure users have control over their data: › Users should be provided with convenient opt-out options; and › The process for opting out should be straightforward and clearly communicated.
Addressing complaints and reports	Establishing channels for receiving and addressing complaints and reports is vital:

	keyword filtering and classification models to detect harmful content. Adequate monitoring personnel should be in place to ensure compliance and address issues promptly.
Ensuring service continuity	Backup mechanisms and recovery strategies for critical data, models, and tools should be established. This ensures seamless service delivery and minimizes disruptions.

Implications for generative AI service providers

The introduction of the draft regulations marks a key moment for generative AI service providers in China, signaling a shift towards more stringent security standards and regulatory oversight.

increased user trust and loyalty, as consumers prioritize security and privacy in their interactions with AI-powered platforms.

Overall, while the new security requirements may pose initial challenges for generative AI service providers, they present greater opportunities for differentiation, innovation, and enhanced user trust in the long run. By embracing these regulations as a framework for responsible AI development, providers can position themselves for long-term success in an increasingly regulated market.

Events in China

ALL EVENTS

Our free webinars are packed full of useful information for doing business in China.

15
MAY

Navigating China's Compliance Framework in the Digital Age – Solutions for Forward-Thinking Enterprises

In-Person Event | Thursday, May 15, 2025 | 4:00-5:30 PM Shanghai, China

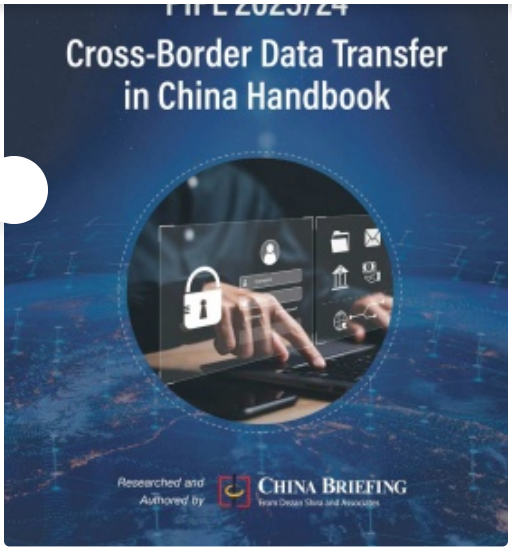
Adam Livermore
Partner

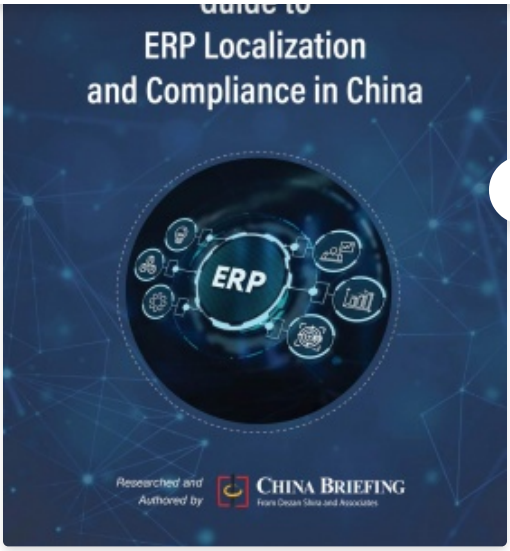
Lee Ding
Manager

Nathaniel Rushforth
Senior Data Security & Compliance Consultant

JOIN EVENT

Related reading





DEZAN SHIRA & ASSOCIATES

Meet the firm behind our content. Visit their website to see how their services can help your business succeed.

About Us | Find an Advisor

Get free access to our subscriptions and publications

Subscribe to receive weekly China Briefing news updates,
our latest doing business publications, and access to our Asia archives.

Sign Up Now

About Us

HR & Payroll

Bookstore

Visit Publication

My Account

My Order History

Products

Media Partners

Partner