SESEC V Report

Landscape of Artificial Intelligence Security Standards in China

December | 2024

Seconded European Standardisation Expert in China
(SESEC)

Landscape of Artificial Intelligence Security Standards in China

With the rise and application of artificial intelligence (AI) technologies, China is placing increasing emphasis on AI security governance. A series of policies, regulations, and standards have been issued or are under development to establish a comprehensive regulatory framework for AI, ensuring security across all stages of AI development and application.

1. AI Security-Related Policies and Regulations

In July 2017, the State Council issued the first systematic policy document for the AI industry, *The New Generation Artificial Intelligence Development Plan*. This document outlines a comprehensive roadmap for China's AI development through 2030, including strategic objectives, key tasks, and support measures. It elevates AI governance to a national strategic level. Regarding security, the document proposes establishing an AI security supervision and evaluation system, which includes, among others:

- Strengthening research and evaluation of AI's impact on national security and confidentiality.
- Building AI security monitoring and early warning mechanisms.
- Establishing a transparent AI regulatory system to oversee all stages, from algorithm design to product development and application.
- Advancing research on AI cybersecurity technologies and enhancing the network security of AI products and systems.

In July 2020, five ministries, including the Cyberspace Administration of China (CAC), Standardisation Administration of China (SAC), and Ministry of Industry and Information Technology (MIIT), jointly issued the *Guidelines for the Construction of the National New Generation AI Standard System*, outlining a framework for AI standards development in China. In June 2024, relevant authorities updated this framework by releasing the *Guidelines for the Construction of the National Comprehensive Standardisation System for the AI Industry (2024 Edition).*

According to this document, security/governance is one of the seven core components of China's AI standards system. This component aims to regulate the security requirements for the entire lifecycle of AI technologies, products, systems, applications, and services. Standards under this domain will cover:

- Foundational security requirements.
- Data, algorithm, and model security.
- Network, technical, and system security.
- Security management and services.
- Security testing and evaluation.
- Security labeling.
- Content identification.
- Product and application security.

In July 2023, CAC, MIIT, the National Development and Reform Commission (NDRC), and four other departments jointly issued the *Interim Administrative Measures for Generative Artificial Intelligence Services.* This is China's first regulation specifically targeting generative AI. Generative AI here refers to services that use generative AI technologies to provide the public in China with content such as text, images, audio, and video.

The document outlines regulatory requirements for all stages of generative AI services, including data labeling, model pre-training and optimization, algorithm and model evaluation, content identification and review, and service provision. This regulation has been in effect since August 2023.

In October 2023, the Chinese government released the *Global Initiative on AI Governance*, presenting China's philosophy on AI governance. Regarding AI security, the initiative emphasizes the principle of balancing development and security, ensuring the protection of personal privacy and data security during AI research and application.

2. AI Security Standardisation

To implement the *Global Initiative on AI Governance*, the National Cybersecurity Standardisation Technical Committee (TC260) released the *Artificial Intelligence Security Governance Framework* (hereinafter referred to as the Framework) in September 2024. The Framework outlines principles for AI security governance, including inclusive and prudent approaches to ensure security, risk-oriented and agile governance, integration of technology and management for coordinated responses, and open collaboration for co-governance and shared benefits. Following a risk management approach and considering the characteristics of AI technologies, the Framework analyzes the sources and manifestations of AI risks. It addresses inherent security risks such as model and algorithm security, data security, and system security, as well as application security risks across network, physical, cognitive, and ethical domains. Furthermore, the Framework proposes corresponding technical responses, comprehensive prevention and mitigation measures, and guidelines for the secure development and application of AI.

To implement the Framework, TC260 is developing the *AI Security Governance Standards System.* This system focuses on the various security risks identified in the Framework and the corresponding countermeasures proposed. Its objectives include:

- Providing standardized support to prevent and mitigate major AI security risks.
- Supporting the implementation of regulations such as the *Interim Administrative Measures for Generative Artificial Intelligence Services.*

The latest version of the *AI Security Governance Standards System* proposes 43 ongoing and planned standards projects. These are categorized into five key areas:

- Basic and General Standards: Covering terms, classification and grading, basic requirements, etc.
- Foundational Support Standards: Including data security, system security, and similar foundational aspects.
- Key Technology Standards: Addressing areas such as generative and synthetic security, agent security, etc.
- Security Management Standards: Encompassing development security, operational security, supply chain security, and more.

- Product and Application Standards.

Although the system is still under development, it provides a general outline of China's AI security standardisation work and its planning.

The system includes several projects specifically designed to support the implementation of various provisions in the *Interim Administrative Measures for Generative Artificial Intelligence Services.* These projects aim to translate the regulation's requirements into actionable standards.

Standard Projects and Their Correspondence with Regulatory Requirements

Standard Project	Relevant Standards	Corresponding Regulation Requirements
Data Annotation	GB/T *Procedures for Data Annotation in Machine Learning for Artificial Intelligence;* GB/T *Cybersecurity Technology - Security Specifications for Data Annotation in Generative AI*	Article 8: During the development of generative AI technologies, providers must establish clear, specific, and actionable annotation rules that comply with the requirements of these measures.
Model Training and Optimization	GB/T *Cybersecurity Technology - Security Specifications for Pre-training and Optimization Training Data in Generative AI*	Article 7: Providers of generative AI services conducting activities such as pre-training and optimization training must use data and foundational models from legitimate sources. They must not infringe upon intellectual property rights or violate personal information laws without consent or meeting other legal provisions. Measures must improve data quality, authenticity, accuracy, objectivity, and diversity in compliance with laws like the *Cybersecurity Law*, *Data Security Law*, and *Personal Information Protection Law*.
Algorithm and Model Evaluation	*GB/T Information Security Technology - Security Assessment Specifications for Machine Learning Algorithms;* *GB/T Security Assessment Provisions for Internet Information Services with Public Opinion or Social Mobilization Attributes*	Article 17: Providers offering generative AI services with public opinion or social mobilization attributes must conduct security assessments in accordance with national regulations.
Content Identification and Review	*GB Cybersecurity Technology - Identification Methods for AI-Generated Synthetic Content*	Article 12: Service providers must label generated content such as images and videos in accordance with the *Administrative*

		Provisions on Deep Synthesis for Internet Information Services[1].
Service Provider Filing	*GB/T Cybersecurity Technology – Basic Security Requirements for Generative AI Services*	Article 17: Providers offering generative AI services with public opinion or social mobilization attributes must comply with the *Administrative Provisions on Algorithm Recommendation for Internet Information Services*[2], including filing for algorithms, as well as handling changes or deregistration of filings.

In addition to the aforementioned standards, the system also includes the following key standards currently under development, which are expected to play a significant role in future AI security regulation:

- *GB/T Cybersecurity Technology - Security Framework for Artificial Intelligence Computing Platforms: Currently in the draft for public consultation stage.*
- *GB/T Cybersecurity Technology - Security Specifications for Deep Synthesis in Internet Information Services: Currently in the draft stage.*
- *GB/T Cybersecurity Technology - Security Requirements for AI Code Generation Services: Currently in the draft stage.*

During the second Standards Week event of 2024, held by TC260 in December 2024, a large number of new standard proposals were discussed. Among these, proposals related to large models supply chain security, generative AI security evaluations baseline data security, AI application security testing and evaluation, detection of AI-generated synthetic content, security of AI code generation services, and AI internet application security received significant support and are likely to be established as formal projects in the future.

3. Conclusion

China's current AI security regulation primarily revolves around the *Interim Administrative Measures for Generative Artificial Intelligence Services*. However, the implementation of this regulation still lacks sufficient technical standard support. TC260 is actively engaged in extensive research and development of AI-related standards, and those specifically designed to support this regulation are highly likely to become mandatory regulatory requirements in the future. These developments warrant close attention from overseas stakeholders.

Additionally, TC260 is building a comprehensive AI security standards system that will reflect China's future standardisation trends in this field. It is recommended that overseas stakeholders closely monitor its progress to stay informed of emerging regulatory and standardisation requirements.

[1] Effective in January 2023, it stipulates that providers of deep synthesis services with public opinion or social mobilization attributes (those providing internet information services using deep synthesis technologies within China) must file with regulatory authorities in accordance with the *Administrative Provisions on Algorithmic Recommendation for Internet Information Services.*

[2] Effective in March 2022, it stipulates that providers of algorithmic recommendation services with public opinion or social mobilization attributes must file with regulatory authorities within ten working days from the date of providing the service.

Introduction of SESEC Project

The Seconded European Standardisation Expert in China (SESEC) is a visibility project co-financed by the European Commission (EC), the European Free Trade Association (EFTA) secretariat and the three European Standardisation Organizations (CEN, CENELEC and ETSI). Since 2006, there has been four SESEC projects in China, SESEC I (2006-2009). SESEC II (2009- 2012), SESEC III (2014-2017), SESEC IV (2018- 2022) and SESEC V (2022-2025). Dr. Betty XU is nominated as the SESEC expert and will spend the next 36 months on promoting EU-China standardisation information exchange and EU-China standardisation cooperation.

The SESEC project supports the strategic objectives of the European Union, EFTA and the European Standardisation Organizations (ESOs). The purpose of SESEC project is to:

- Promote European and international standards in China;
- Improve contacts with different levels of the Chinese administration, industry and standardisation bodies;
- Improve the visibility and understanding of the European Standardisation System (ESS) in China;
- Gather regulatory and standardisation intelligence.

The following areas have been identified as sectorial project priorities by the SESEC project partners: Internet of Things (IoT) & Machine-to-Machine(M2M) communication, communication networks & services, cybersecurity & digital identity, Smart Cities (including transport, power grids & metering), electrical & electronic products, general product security, medical devices, cosmetics, energy management & environmental protection (including eco-design & labeling, as well as environmental performance of buildings).

SESEC V China Standardisation and Technical Regulation Bimonthly Newsletter

SESEC V China Standardisation and Technical Regulation Bimonthly Newsletter is the gathering of China regulatory and standardisation intelligence. Most information of the Monthly Newsletter was summarized from China news media or websites. Some of them were the first-hand information from TC meetings, forums/workshops, or meetings/dialogues with China government authorities in certain areas.

In this Bimonthly Newsletter

In this Bimonthly Newsletter, some news articles were abstracted from Chinese government organizations. All new published standards, implementation or management regulations and notice are summarized; original document and English version are available.