Deciphering the AIGC Compliance Blueprint (Part III): Security Assessment for AIGC Products

Peng Cai, Pulingling Xiao

1. INTRODUCTION

Pursuant to Article 17 of the *Interim Measures for the Management of Generative Artificial Intelligence Services* ("AIGC Measures"), providers offering AI-related services with public opinion nature or capable of social mobilization shall conduct security assessment in line with applicable laws and regulations. With respect to the market entry of AIGC products, security assessment is a pivotal step in ensuring both product compliance and user security. As Part III in a series of articles intended to chart the regulatory course for AIGC and explore its potential trajectory under the current legal advancements, this article aims to dissect the complex nature of "security assessment" by tracing its historical development and exploring the current landscape. This article will also analyze how to formulate and execute effective security assessment strategies, offering insight into facilitating successful market entry of AIGC products.

2. THE SECURITY ASSESSMENT: VERSION 0.5

In 2017, the Cyberspace Administration of China ("CAC") already introduced the regulatory approach of "security assessment" in the *Regulations on the Security Assessment of New Technologies or New applications for Internet-based News Information Services* ("Double-new Measures"). The Double-new Measures stipulates that internet news service providers should conduct security assessment and compile a written security assessment report when (1) applying new technologies, adjusting or adding application functions related to news rendering or bearing the public opinion nature or social mobilization capabilities, or (2) causing changes in user scale, functions, technical implementation methods, basic resource allocation and other aspects which lead to significant changes in news rendering, public opinion nature or social mobilization capabilities. Such service providers must submit the required security

assessment report to the CAC or its provincial offices for review. This early form of security assessment is what the industry now refers to as the "Double-new Assessment" (new technology, new application).

However, under the Double-new Measures, the duty to conduct security assessment is confined to "providers of internet news information services", in light of which, the security assessment requirement under the Double-new Measures does not apply to AIGC products which does not concern news information services. We refer to this early form of security assessment as "Security Assessment 0.5".

3. THE SECURITY ASSESSMENT: VERSION 1.0

A year after the implementation of the Double-new Measures, the CAC issued the *Provisions on the Security Assessment of Internet Information Services with Public Opinion Nature or Social Mobilization Capacity* ("Public Opinion Provisions"). This time, the application scope of security assessment was no longer confined to news information services. According to Article 3 of the Public Opinion Provisions, providers of internet information services shall carry out security assessment under the following circumstances:

- Information services with public opinion nature or social mobilization capacity are implemented online, or corresponding functions are integrated into such online information services.
- There are new technologies or new applications that will cause significant changes in information services' functions, technical implementation methods, basic resource allocation, etc., thus leading to major changes in public opinion nature or social mobilization capacity.
- User scale is markedly increasing, resulting in major changes in the public opinion nature or social mobilization capacity of such information services.
- Unlawful or harmful information has been disseminated, which indicates that the existing security measures fall short to effectively prevent and control cybersecurity risks.
- Other circumstances occur where cyberspace administrations or public security bureaus at the prefecture level and above notify in writing that security assessment is required.

In such circumstances, providers of internet information services are obligated to conduct security assessment and submit the security assessment report to the cyberspace administrations or public security bureaus at the prefecture level and above via the National Internet Security Management Service Platform (https://www.beian.gov.cn/portal/index.do). The specific requirements for security assessment can be found in Article 5 of the Public Opinion Provisions.

Additionally, cyberspace administrations or public security bureaus have the authority to initiate on-site inspections based on their evaluations. Generally, providers of internet information services should furnish the designated public security bureau with relevant documents for on-site inspections. Looking from the submission window and on-site inspection bodies, the public security bureau is the primary body responsible for implementing "Security Assessment 1.0".

Even though Security Assessment 1.0, by law, is closely related to internet information services characterized by "public opinion nature or social mobilization capacity," in practice, interpretation of what qualifies as such services is somewhat expansive, virtually covering all products featuring internet information interaction functionalities or channels. The submission process for Security Assessment 1.0 is very transparent, and businesses simply need to accurately complete the forms and submit their assessment reports under the guidance of the Security Assessment User Manual available on the National Internet Security Management Service Platform. Generally, the review of the security assessment reports, from submission to approval, can be accomplished within a month.

4. THE SECURITY ASSESSMENT: VERSION 2.0

Fast forward to 2023, a year of exceptional growth and development in the field of artificial intelligence. AIGC product developers, primarily those working on large models, are now required to provide comprehensive Double-new Assessment reports to the CAC. Such reports must meet extensive requirements, normally including over a hundred pages of meticulous details, which are quite different from Security Assessment 1.0 reports submitted via the National Internet Security Management Service Platform. Hence, we refer to this as "Security Assessment 2.0".

The primary regulatory targets of Security Assessment 2.0 are AIGC developers. Based on our experience and observations, Security Assessment 2.0 differs from Security Assessment 1.0 in the following ways:

	Security Assessment 1.0	**Security Assessment 2.0**
Subject	Internet Information Services Provider	Developers of AIGC Products Which Are Primarily Based on Large Models
Conditions	• before the market launch of internet information services or related functions with public opinion nature or social mobilization capacity; • before the launch of new technology or new application of information services; • when the scale of information service users significantly increases; • when harmful information disseminates.	• before the launch of deep synthesis technology; • before the launch of facial recognition positioning, identity authentication, and attribute analysis services; • before the launch of internet information services based on algorithm recommendation; • before the launch of AI system; • before the launch of blockchain information services.
Regulatory Authority	Public Security Bureaus at the Prefecture level and above	CAC
Report Content	• setting of security department; • identity authentication; • retention of the logs of registration, published information, and weblogs; • measures for preventing and disposing of harmful information; • technical measures to prevent the spread of harmful information; • channels for complaints and rights protection; • mechanisms for cooperation and assistance in law enforcement.	• security guarantee of the subject (organization, system, technology and management of third parties); • security guarantee of information (information resource, content review, information release, monitoring and warning, information storage and destruction; • user security; • technology security (deep synthesis, facial recognition, algorithm recommendation, AI system and blockchain).
Assessment Key Issues	• business application security; • business platform security; • business operation security; • data security.	• ideological security; • legal and ethical security; • technology security.
Assessment Period	2-4 weeks	Unknown
Pass Rate	High	Low

Regulatory authorities have been closely monitoring the potential security risks of AIGC products for some time. Back in 2021, the CAC, along with the Ministry of Public Security ("MPS"), issued directives to strengthen the security assessment of emerging internet technologies and applications, especially those related to voice-centric social media and deepfake technologies. The cited legal basis then was the Public Opinion Provisions. Fast forward two years to today, and we now see significantly enhancements in the rigor of regulatory oversight, with major AIGC product developers currently grappling with the requirements of Security Assessment 2.0.

Interestingly, Security Assessment 2.0 appears to be implemented in practice despite a lack of clear statutory guidance. Pursuant to Article 6 of the draft AIGC Measures, service providers shall declare the security assessment to the CAC before launching any AIGC products to the public per the Public Opinion Provisions. Compared to Security Assessment 1.0, Article 6 of the draft AIGC Measures revise up the regulatory body from the "cyberspace administration offices or public security bureaus at the prefecture level and above" to the CAC. Hence, it was once perceived as the legal basis for Security Assessment 2.0.

But in the officially released AIGC Measures, the requirements for declaring security assessment to the CAC have been modified. It now states, "those providing AIGC services with public opinion nature or social mobilization capacity, security assessment should be conducted according to relevant provisions of the state." This updated, albeit somewhat nebulous, statutory requirement may be interpreted in several ways:

- **Possible Interpretation 1**. "AIGC services with public opinion nature or social mobilization capacity" fall under the category of "internet information services with public opinion nature or social mobilization capacity." Thus, AIGC service providers should satisfy the relevant requirements set forth in the Public Opinion Provisions, including the submission of Security Assessment 1.0.
- **Possible Interpretation 2**. The term "relevant provisions of the state" is not confined to the Public Opinion Provisions. Therefore, AIGC service providers should also comply with other applicable laws and regulations. If other applicable laws and regulations put forward similar security assessment requirements, AIGC

service providers should also fulfill their relevant assessment obligations in accordance with such laws and regulations.

In addition, the requirements for "security assessment" also recur in other AI-related legislations. For instance, providers of algorithmic recommendation services that bear public opinion nature or social mobilization capacity must conduct security assessment. Similarly, providers and technical supporters of deep synthesis services with specific functionalities must also perform security assessment according to relevant laws and regulations. The recurring requirements of "security assessment" across different legislative frameworks inevitably cause confusion among businesses. Consequently, we cautiously foresee that the CAC may introduce auxiliary regulations or specific assessment guidelines for Security Assessment 2.0 in the future, thereby filling the current legislative gap.

5. THE SECURITY ASSESSMENT: MIIT VERSION

In addition to Security Assessments 0.5, 1.0, and 2.0, all of which were supervised by the cyberspace administration offices or public security bureaus, the Ministry of Industry and Information Technology ("MIIT") introduced another variant of security assessment, referred to as the "MIIT Version".

The inception of the MIIT Version can be traced back to the *Internet New Business Security Assessment Management Measures (Draft for Comments)* (the "MIIT Measures"). It was promulgated by the MIIT in 2017, which requires telecommunications business operators to carry out security assessment - also known as "Double-new Assessments" - regarding potential cybersecurity threats that their new Internet services may pose. Despite the MIIT Measures not having officially taken effect, the MIIT has subsequently released a succession of industry standards that pertain to the security assessment of new internet technologies and services within the telecommunications sector. This progression has enabled the practical implementation of the MIIT Version of the security assessment:

- YD/T3169-2020 Guidelines on Security Assessment of New Internet Technology and New Internet Business
- YD/T3738-2020 Implementation Requirements on Security Assessment of New Internet Technology and New Internet Business
- YD/T3739-2020 Requirements on Security Assessment of New Internet Technology and New Internet Business: Instant Communication Business
- YD/T3740-2020 Requirements on Security Assessment of New Internet Technology and New Internet Business: Internet Resource Collaboration Services
- YD/T3741-2020 Requirements on Security Assessment of New Internet Technology and New Internet Business: Applications and Services Based on Big Data Technology
- YD/T3742-2020 Requirements on Security Assessment of New Internet Technology and New Internet Business: Content Distribution Business
- YD/T3743-2020 Requirements on Security Assessment of New Internet Technology and New Internet Business: Information Search and Query Services
- YD/T3503-2019 Agency's Capability Certification Criteria of Security Assessment of New Internet Technology and New Internet Business

The MIIT Version of security assessment currently operates under a well-established service procedure, and recognized assessment institutions are available to conduct the process. Notably, the content of the MIIT Version largely intersects with the requirements of Security Assessment 2.0. The extent to which the CAC, MIIT, and MPS will coordinate and harmonize their regulatory scopes and benchmarks in the future remains an open question.

6. CONCLUSIONS

The existing dialogue surrounding security assessment remains ambiguous, leaving businesses in dire need of further clarification and definition from the regulatory authorities. It's indisputable that Security Assessment 2.0 has emerged as the toughest hurdle to cross before AIGC products can enter the market. We would strongly advise businesses that haven't commenced preparation to deal with this matter with utmost seriousness. Initiating internal projects at an early stage, dedicating personnel, and enlisting external legal help if needed, could be instrumental in facilitating the smooth market entry of AIGC products.