

Authoritarian Innovation

How China Is Building Control into Generative AI Systems



Authoritarian Innovation

How China Is Building Control into Generative AI Systems

By Owen Au and Aiden Gu

© 2026 Resilience Innovation Lab

This publication is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0).

You may copy and redistribute this publication in any medium or format, provided that appropriate attribution is given, the material is not used for commercial purposes, and no modified versions are distributed. For more details, visit the official license page: creativecommons.org/licenses/by-nc-nd/4.0/

Contact: info.ril@proton.me

ACKNOWLEDGEMENTS

ACKNOWLEDGEMENTS

This report is jointly published by the Resilience Innovation Lab (RIL) and the Strategic China Policy Group (SCPG). This report is co-authored by Owen Au and Aiden Gu. The RIL, SCPG, and the authors would like to thank Dr Eric Yan-ho Lai, Daniel Li, and Morgan Lee for their valuable insights.

ABOUT RESILIENCE INNOVATION LAB

Resilience Innovation Lab (RIL), founded by Asian veteran researchers and human rights defenders in 2023, serves to promote innovation, experimentation, and the exchange of research and educational initiatives that aim to develop and strengthen the democratic and digital resilience of organisations and individuals working in repressive environments.

More about RIL: resilienceinnovationlab.org

ABOUT STRATEGIC CHINA POLICY GROUP

The Strategic China Policy Group (SCPG) is an independent, non-profit research initiative providing primary-source analysis of the development of China's strategic capabilities and its implications for Europe's resilience and competitiveness.

More about SCPG: strategicchinapolicy.org

TABLE OF CONTENTS

Foreword	5
Executive Summary	7
Introduction	8
Part I: Foundations of China's Information Control	10
Quest for Ideological Security	10
Content Security Risks in the Age of Generative AI	10
Part II: Extending Control in the Age of Generative AI	13
Strengthening Cyberspace Governance	13
Governing the Generative AI Pipeline	15
Industrial Standardisation	15
Part III: The State-Corporate Nexus	20
Reconfiguring the Dynamics	20
Incorporation into the Governance System	21
Standardisation as Political Incorporation	22
Part IV: Exporting China Model	24
Scale of Adoption	24
Persistent Control	24
Dependency Risk	25
Discussion	26
Annex A	27
Annex B	29
Annex C	30
Endnotes	31

Foreword

The governance of artificial intelligence (AI) has moved, within a remarkably short span, from a specialist preoccupation to one of the defining questions of global order. Yet the debate it has provoked is not converging towards a shared framework; it is diverging. As states, corporations, and multilateral bodies scramble to set rules for a technology whose trajectory none of them fully controls, the underlying disagreement is less about technical standards than about the values, interests, and conceptions of authority that such standards are meant to encode.

At least two broad currents run through this debate. The first is avowedly value-based. In his first encyclical *Magnifica Humanitas* (15 May 2026), Pope Leo XIV frames the age of artificial intelligence as a test of whether humanity will build “yet another Tower of Babel” or instead rebuild a shared civic home.¹ He places human dignity, the common good, subsidiarity, and solidarity at the centre of any legitimate technological order, and warns against a technocratic paradigm in which efficiency, control, and profit are permitted to dictate what matters and what may be discarded. The second current is frankly pragmatic, organised around national interest and control. The Trump administration’s recent decision to vet access to OpenAI’s most advanced model customer by customer — restricting its release to a small circle of approved partners on national-security grounds — illustrates how even the most market-oriented systems are now folding frontier AI into the machinery of strategic competition.² Between these two impulses, moral universalism and security pragmatism, much of the rest of the world is left to find its footing.

Against this backdrop, China occupies a distinctive and frequently misread position. Amid mounting geopolitical uncertainty and the visible erosion of multilateral capacity, Beijing has signalled an ambition that extends well beyond the management of its own information space: it seeks to help write the rules for everyone else. In its new White Paper of 17 June 2026, titled *More Just and Equitable Global Governance: China’s Principles, Proposals and Actions*, the government presents itself as a champion of “AI for good and for all”, pledging to support a central role for the United Nations, to assist developing countries in bridging the AI divide, and to convene new institutions — among them a proposed World Artificial Intelligence Cooperation Organisation and an array of capacity-building initiatives — through which China would contribute its own approaches to the emerging global order.³ The register is inclusive, developmental, and reassuringly multilateral.

That outward-facing confidence, however, sits atop a domestic debate that is conspicuously muted. As observed by a China human rights expert, the quiet surrounding AI in China reflects control rather than consensus: governance is shaped overwhelmingly by state mandates and regulatory agencies, while the courts, independent researchers, journalists, civil-society organisations, and labour and digital-rights groups that contest AI policy elsewhere have been systematically constrained or dismantled.⁴ Within politically permitted boundaries, experts may advise and citizens may complain, but these inputs do not form independent centres of pressure capable of shaping policy from outside the state-led framework. The consequence is that what is presented internationally as “China’s” position on AI governance is, in practice, an official position, largely insulated from the pluralistic contestation that might otherwise temper it.

The present report, *Authoritarian Innovation*, makes its contribution, examining China’s approach to AI

governance through the lenses of national security, information control, and political economy. Within China's broader framework of AI safety and security, content that departs from Party narratives or unsettles its ideological worldview is treated as a content-security risk to be managed. As Chinese generative models gain traction abroad, the Party-aligned content controls built into them travel with the technology, carried into foreign information spaces without the need for any overt information operation. With Beijing seeking a larger role in global standard-setting, international regulators and developers would do well to recognise that the language of "AI safety and security" can carry its own political and ideological freight. To overlook that political context is to risk allowing authoritarian approaches to content control to shape the rules of global AI governance from within.

As China advances an approach to generative AI organised around the safeguarding of "content security", it becomes all the more essential for policy communities and civil society to continue preserving, archiving, and disseminating genuine information and authentic government records — the better to resist the censorship, erasure, and fabrication of manipulated content that generative tools, instrumentalised by authoritarian regimes such as China's, increasingly make possible.

We hope this report reaches a wide audience, and that, by engaging the current debate strategically from a Chinese-speaking, stakeholder perspective, it contributes a voice too often absent from these discussions. For in an age of pervasive censorship and information manipulation, the production and dissemination of knowledge are not peripheral activities but essential tools of resistance — and it is in that spirit that we offer this report, in the hope that it helps, in some measure, to forestall the autocratisation of AI governance.

Resilience Innovation Lab

Executive Summary

- China's use of generative AI for information manipulation has drawn growing attention worldwide, but from Beijing's perspective, the technology is also a risk vector that could weaken its own control over the information environment. Based on qualitative review of publicly available Chinese sources published between 2013 and 2026, this report shows that **Beijing is building political control into generative AI systems through an increasingly comprehensive governance mechanism, which enables information control to operate at the technical layer even without deliberate information operations.**
- **Within China's broader AI safety and security framework, content that deviates from Party narratives or challenges the Party's ideological worldview is treated as a content security risk.** This has driven efforts to embed control into generative AI systems before content is generated.
- **China's policy response centres on an expanding system of industrial standards** across training data, data annotation, model behaviour, application testing, and content traceability. These standards translate high-level political requirements into implementable technical procedures, ensuring that information control is built into the process before content is generated.
- **Private technology firms are incorporated into China's AI governance system, working with state actors to embed the Party's political requirements into technical standards and model behaviour.** This role follows the early-2020s reconfiguration of state-private sector dynamics, when China's private sector was disciplined and incentivised to align its business and technical capacity with the Party's political objectives.
- **As Chinese generative AI models gain traction abroad, Party-aligned content controls embedded in these systems are travelling beyond China's domestic information space.** Without relying on overt information operations, their diffusion creates a less visible form of influence than conventional information manipulation, with risks deepening as foreign developers, platforms, and applications become dependent on Chinese models.
- With China seeking a greater role in global AI governance and standard-setting, **international regulators and developers should recognise that the country's language of AI safety and security often carries its own political and ideological requirements.** Overlooking this political context risks allowing authoritarian approaches to content control to shape the rules of global AI governance.

Introduction

In January 2025, a Chinese start-up released DeepSeek-R1, a generative artificial intelligence (AI) model that reportedly achieved performance comparable to leading U.S. models such as OpenAI's o1, while being significantly more cost-efficient.⁵ While the “DeepSeek moment” signalled a narrowing gap in technological competition with the U.S., it also raised concerns that China's growing technological capabilities could extend its authoritarian influence. Soon after, in the article *Why DeepSeek is so Dangerous*, Valentin Weber argued that generative AI could have “massive potential to enhance China's already pervasive surveillance state”.⁶ Later that year, a report by the Australian Strategic Policy Institute, *The Party's AI: How China's new AI systems are reshaping human rights*, highlighted how the technology is already being deployed at scale in China to automate censorship, enhance surveillance, and pre-emptively suppress dissent.⁷

Beyond domestic control, a growing body of evidence also suggests that China is actively using generative AI to strengthen information operations abroad. A leaked 2025 document from GoLaxy (中科天璣) illustrates how the technology is being used in China to shape public discourse in Hong Kong, Taiwan, and the U.S.⁸ Some intelligence analysts have described this as “a new weapon in information warfare”, while other observers have expressed concern about the country's expanding systems of Foreign Information Manipulation and Interference (FIMI), powered by its growing technological capabilities.⁹

While extensive research has focused on China's use of generative AI to manipulate information environments at home and abroad, less attention has been given to how Beijing seeks to control the technology itself. This report examines how China embeds information control into generative AI systems. For Beijing, the technology is not only an instrument for consolidating state control over the information environment, but also a risk vector that could weaken that control by enabling users to generate and disseminate content across the digital space at scale. Driven by this anxiety, China is increasingly seeking to intervene at the stage of technological development itself, through expansive policy efforts, industrial standard-setting, and collaboration with the private sector. The aim is to ensure that information control is built into the process before content is generated.

This matters because generative AI is increasingly central to today's information environment. Successive surveys over the past year point to an international trend in which generative AI is being integrated into people's daily lives and increasingly used as a source of information.¹⁰ Meanwhile, as Beijing pursues its official ambition to build China into a global technology powerhouse (科技强国), Chinese AI models and applications are gaining traction worldwide.¹¹ According to OpenRouter, by May 2026, seven of the world's top 10 most popular large language models (LLMs) are from China, accounting for approximately two-thirds of token use.¹² These trends, in combination, are expanding the channels through which the control logic embedded in Chinese technologies could influence the global information landscape.

This report is based on a qualitative review of publicly available Chinese policy documents, laws, administrative measures, national and sectoral standards, technical specifications, corporate announcements, certification materials, and third-party assessments published between 2013 and 2026. These materials were selected for their relevance to generative AI governance, content security,

standard-setting, model evaluation, and state–corporate coordination in China’s AI sector. Drawing on these sources, the report first examines the political context of China's information control regime and explains why Beijing treats generative AI as both an opportunity and a risk. It then analyses how this control logic is extended across the generative AI pipeline through platform governance, training data controls, model evaluation, application testing, and content traceability. The report next examines the state–corporate relationship that makes this system operational before assessing the international implications of the growing global diffusion of Chinese AI models. As the analysis relies primarily on publicly available sources, it may not capture internal regulatory instructions, closed-door standards deliberations, informal guidance to firms, or unpublished enforcement practices that further shape implementation.

Foundations of China's Information Control

Quest for Ideological Security

China's information control is driven by Beijing's quest for ideological security (意识形态安全). Over the last decade, President Xi Jinping has persistently expressed his paranoia over how democratic values promoted by "Western hostile powers" (西方敌对势力) could subvert the rule of the Chinese Communist Party (CCP), asserting that the "complete difference" in systems and ideology determines that the "struggle and competition between [China] and the West are irreconcilable".¹³ In the 13th Five-Year Plan, released in 2016 as the first Five-Year Plan under Xi Jinping, safeguarding ideological security was highlighted for the first time as a priority of national security, while the concept was incorporated into the overarching "holistic approach to national security" (总体国家安全观).¹⁴ As John Garnaut, a former China policy adviser to the Australian government, commented in 2017, "Xi Jinping has reinvigorated ideology to an extent we have not seen since the Cultural Revolution".¹⁵

Beijing's prioritisation of ideological security under Xi Jinping is well reflected in a political instruction issued by the General Office of the Central Committee (中共中央办公厅) in April 2013 — just one month after Xi became China's president. This Communiqué on the Current State of the Ideological Sphere (关于当前意识形态领域情况的通报), or more commonly known as the Document No. 9, explicitly highlights concepts such as constitutional democracy, universal values, civil society, liberalism, and freedom of the press and information as "mistaken thinking trends" (错误思潮) that could destabilise the social and political order in China.¹⁶ In this context, China's ideological security agenda has twofold objectives: first, to strengthen the dominance of Party-aligned narratives and ensure that these narratives remain mainstream in the information space; second, to reject — if not undermine — those "mistaken thinking trends" considered to represent the interests of "Western hostile powers".¹⁷

With cyberspace long treated in China as the 'main front of ideological struggle' (意识形态斗争的最前沿), securing control over the online information environment has long been a priority within Beijing's ideological security agenda.¹⁸ As concluding instructions in the Document No. 9, Party members were required to take measures to strengthen "online public opinion guidance" (网上舆论引导), in order to "purify online narrative environment" (净化网络舆论环境).¹⁹ This priority was further elevated in the 2015 amendment of the National Security Law, when Beijing introduced an article on "cyber sovereignty" (网络主权) that seeks to strengthen state control over online information flows and eliminate online content deemed "illegal and harmful" by the authorities.²⁰ It has been accompanied by a series of state efforts to expand the mechanism through which the authorities can systematically monitor, censor, and shape information flows in the digital space.²¹

Content Security Risks in the Age of Generative AI

Emerging generative AI technology threatens to disrupt the Party's control mechanism. Generative AI offers a powerful technological means of upgrading China's information operations, enabling more automated and adaptive operations across the information environment.²² Yet even as the country enthusiastically adopts the technology as an advantage for state information control, Chinese policy discussions also acknowledge its double-edged nature.

A paper published under the Chinese Academy of Social Sciences (中国社会科学院, CASS) in 2025 highlights the risk of generative AI being used to construct “distorted and chaotic narrative discourses” and, in worse cases, a “false social consensus” that deviates from the “mainstream”.²³ These concerns are conceptualised as “information and content security risks” (信息内容安全风险) and “cognitive domain security risks” (认知域安全风险) in the AI Security Governance Framework (人工智能安全治理框架) established by China’s national cybersecurity standards body. The framework expresses concern over the potential for the technology to be used to generate content that harms national and ideological security, or even to be exploited by foreign forces to conduct “cognitive warfare” (认知战) against the country’s sovereignty.²⁴

The risk is inherent to the technology and does not depend only on deliberate hostile use against the Chinese authorities. Undesired content may still be generated in ordinary use cases. Two technical realities make this risk more than hypothetical. First is the inherent “black box problem” in generative AI technology. The “black box” problem refers to the lack of transparency and interpretability in AI systems, making it difficult to determine how specific outputs are generated.²⁵ While this problem is not unique to China, it raises particular concerns in the Chinese context, where opacity in model reasoning may create potential loopholes in information control and ideological security. The inability to fully trace or explain generated content complicates efforts to ensure alignment with official narratives, thereby posing challenges to comprehensive control over the information environment.

The second technical reality is the reliance on training data that is heavily influenced by Western information environment. Policy discussions in China also express concern that the dominance of English-language content in cyberspace may introduce ideological risks through the training data of generative AI systems. With roughly half of global online content produced in English, large-scale datasets used to train models are heavily shaped by Western-origin information.²⁶ Even domestically developed models, such as DeepSeek, rely extensively on English-language materials or datasets derived from them.²⁷ In this context, generative AI is seen as a potential pathway through which Western values and ideological frameworks may be embedded in model outputs and subsequently diffuse into China’s information environment.²⁸

The Chinese authorities take these content risks seriously. In the Security Incident Response Guidelines for Generative AI Services (生成式人工智能服务安全应急响应指南), a national-level guidance document issued by a cross-agency body, security incidents derived from generative AI are classified into four levels according to their impact (Table 1).²⁹ The generation and spread of a “large volume of content that violates core socialist values” is notably placed in the highest severity category, which is treated as more disruptive and unacceptable than most other incidents, including the “leakage of a large amount of users’ sensitive personal data”.

SEVERITY	ILLUSTRATIVE CASES
Level I Severe Incident	CASE 1 Due to a hacker attack, the core model parameters of a highly critical generative AI service are maliciously tampered with, resulting in the platform being unable to generate any content normally, causing a complete service shutdown. CASE 2 A text-based generative service is induced to produce a large volume of content that violates core socialist values, such as disseminating speech that incites national division or undermines ethnic unity. This content spreads widely, causing exceptionally severe impacts on social stability and national security.
Level II Major Incident	CASE 1 An important generative AI service is subjected to a DDoS attack, leading to prolonged service disruption. CASE 2 An important text-based generative service produces content that results in the leakage of a large amount of users' sensitive personal data.
Level III Relatively Major Incident	CASE 1 An important generative AI service experiences a temporary interruption due to software errors or similar causes. CASE 2 A general-purpose generative AI service produces misleading financial information, which spreads and affects market stability.
Level IV General Incident	CASE 1 In a code generation task, a general-purpose generative AI service produces SQL query code that could be maliciously exploited for SQL injection attacks. Although such issues may be identified and corrected during code review, failure to detect them in time could allow vulnerabilities to enter production environments, posing certain security risks. CASE 2 In a natural disaster prediction system, part of the underlying data of a generative AI model is tampered with, resulting in minor errors in generated meteorological warning reports—for example, underestimating the intensity of an impending heavy rainfall. While this may not directly cause disasters, it could affect local residents' preparedness and pose certain risks to public safety.

Table 1: Classification of Security Incident Levels for Generative AI Services (Source: SAC / TC 260, 2025)

These concerns are increasingly reflected in China's policy discussions through the language of "content security" (内容安全).³⁰ For Chinese authorities, the rise of generative AI creates an urgent need to strengthen information control in a digital environment where content can be generated and circulated at scale.

Extending Control in the Age of Generative AI

The risk that generative AI may produce politically unacceptable content has pushed Chinese authorities to build a layered governance system around the technology. The immediate concern is not only that generative AI could be used in organised information operations against Chinese authorities, but also that ordinary use of the technology may produce content that challenges official narratives or weakens the Party's control over the information environment.

China's policy response operates through two layers: first, further tightening control over information flows at the platform level by advancing the country's established cyberspace governance framework; and second, pushing control upstream into training data and model development before content is generated.

Strengthening Cyberspace Governance

The first layer of policy efforts is a continuation of the country's established cyberspace governance framework. As assigned under the Cybersecurity Law in 2016, this part of work is largely overseen by the Cyberspace Administration of China (CAC), which is responsible for supervising and coordinating the governance of cyberspace.³¹ The logic of this layer largely focuses on regulating information at the level of content distribution, by placing online platforms and service providers under state scrutiny.

The 2019 "Provisions on the Governance of the Online Information Content Ecosystem" (网络信息内容生态治理规定) demonstrates the state logic of content control. It classifies online content into three political categories: "encouraged" content (鼓励性内容), "illegal" content (违法信息), and "harmful" content (不良信息).³² While comprehensively covering political, economic, social, cultural and religious issues, it lays an official framework to place the information environment into a hierarchy of political acceptability (*Table 2*). Party-aligned content, such as material promoting "Xi Jinping Thought" (习近平新时代中国特色社会主义思想) and "core socialist values" (社会主义核心价值观), is not merely permitted, but encouraged to be spread in the digital space. Content labelled "illegal", including material that challenges the Party's "fundamental principles" (基本原则) and historical narratives, is strictly banned. Content deemed "harmful" but not yet "illegal", such as "improper commentary" (不当评述) on disasters, is also subject to suppression.

Encouraged content	Illegal content	Harmful content
› Promotes Xi Jinping Thought on Socialism with Chinese Characteristics for a New Era, and explains the path, theory, system, and culture of socialism with Chinese characteristics › Communicates the Party's theories, policies, and major decisions of the Central Committee › Highlights the progress in economic and social development, as well as people's everyday lives › Promotes core socialist values, moral culture, and a positive national spirit › Responds to public concerns and helps building social consensus › Enhances the global influence of Chinese culture and presents the genuine image of China › Reflects good taste and responsibility, promotes truth and positive values, and supports social unity and stability	› Opposes the fundamental principles of the Constitution › Endangers national security, leaks state secrets, subverts state power, or undermines national unity › Harms national honour or interests › Defames or disrespects heroes and martyrs › Promotes terrorism or extremism › Incites ethnic hatred or discrimination › Undermines the State's religious policy or promotes cults or superstition › Spreads rumours that disrupt economic or social order › Involves pornography, gambling, violence, or incitement to crime › Infringes on others' reputation, privacy, or legal rights › Violates other laws and regulations	› Uses misleading or exaggerated headlines › Sensationalises gossip, scandals, or misconduct › Provides improper commentary on natural disasters or major accidents › Contains sexual innuendo or suggestive elements › Depicts bloody, horrifying, or cruel scenes › Incites discrimination › Promotes vulgar or low-quality content › Induces underage individuals to imitate unsafe behavior or harmful habits › Harms the online content environment

Table 2: *Online Content Classifications (Source: CAC, 2019)*

On this basis, the CAC released in 2021 the “Provisions on the Administration of Algorithmic Recommendation for Internet Information Services” (互联网信息服务算法推荐管理规定) together with the Ministry of Industry and Information Technology (MIIT), Ministry of Public Security (MPS) and State Administration for Market Regulation (SAMR), to regulate the algorithm of online platforms.³³ Adopting the categorisation framework outlined in the 2019 Provisions, the 2021 Provisions require platforms to ensure that recommendation algorithms promote “encouraged” content and “core socialist values”, while suppressing “illegal” and “harmful” content.

During the same period, China further institutionalised the state's leverage over online platforms. At the period when the CAC released the 2021 Provisions, it meanwhile issued the “Measures for Cybersecurity Review” (网络安全审查办法) along with twelve other departments.³⁴ Under the Measures, a cross-department cybersecurity review mechanism would be established to monitor the operations of online platforms in China and to assess their potential national security risks, with its office set in the CAC. The Measures also requires online platform operators with a large user base and business engagement abroad to regularly report to the Cybersecurity Review Office, ensuring close scrutiny by the authorities.

As generative AI grew more popular in the early 2020s, the Chinese authorities began applying the existing cyber governance framework to AI services. For instance, the 2022 “Provisions on the Administration of Deep Synthesis for Internet Information Services” (互联网信息服务深度合成管理规定) by CAC, MIIT and MPS extends the requirements for online platforms operators under the 2021 Provisions

to deep synthesis service providers.³⁵ As required by the 2022 Provisions, deep synthesis service providers are obliged to guide public opinions towards “politically correct direction” (正确政治方向), while constraining “illegal” and “harmful” content. The same pattern appears in the 2023 “Interim Measures for the Administration of Generative Artificial Intelligence Services” (生成式人工智能服务管理暂行办法), which require AI-generated texts, images, audio and video clips to comply with the Party’s ideology and oblige providers to suspend services, rectify problems, and report to the authorities when users engage in “illegal activities”.³⁶

Governing the Generative AI Pipeline

The second layer moves control beyond generated content and into the technical process that produces it. In its broader understanding, generative AI is not only a technology that produces texts, images, audio or video, but an end-to-end process through which content is produced, commonly referred to as the generative AI pipeline (*Figure 1*).³⁷ This generative AI pipeline ranges from data ingestion and preprocessing to model training and content generation, with each stage giving regulators a point of intervention before content reaches users.

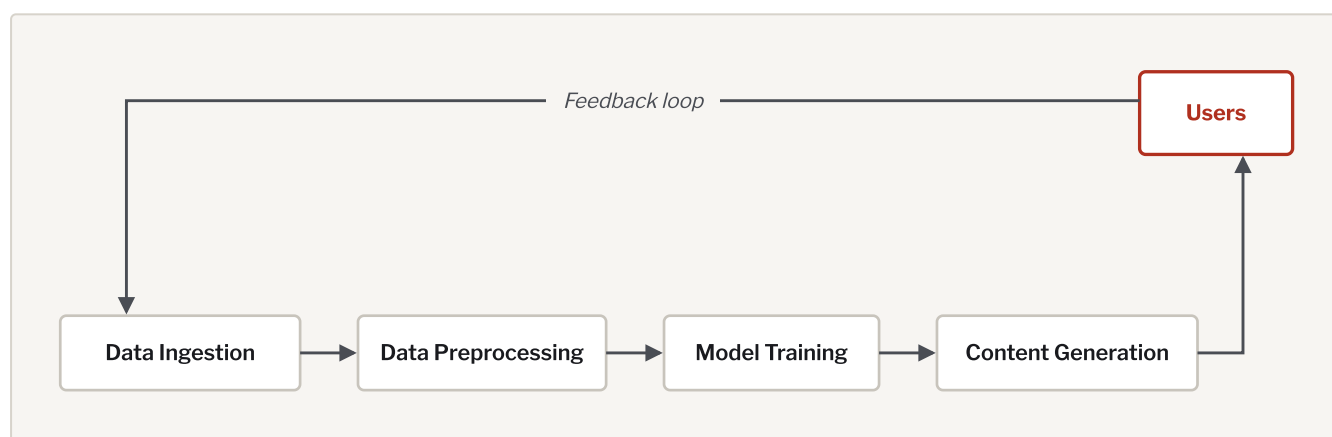


Figure 1: The Generative AI Pipeline (Source: IBM, 2024)

The latest “Interim Measures for the Administration of AI Anthropomorphic Interactive Services” (人工智能拟人化互动服务管理暂行办法) captures this logic clearly. Issued by the CAC in April 2026, it requires providers to manage training data used for pre-training and optimisation before services reach users. Such data must come from “lawful sources” and comply with “core socialist values”, while being cleaned, labelled and regularly checked to prevent manipulated or politically unsafe material from shaping model behaviour.

Industrial Standardisation

Standardisation is the main tool to enforce upstream control. While policies and regulations set out the political requirements for generative AI governance, industrial standards specify the procedures through which these requirements are implemented. They turn broad demands for content security into concrete practices for data processing, model assessment, service testing and content traceability. Since 2024, China has been building an expansive system of AI industrial standards to make these practices consistent across the industry (*Annex B*). This process is mainly driven by the MIIT and the National Standardisation Administration (SAC)³⁸, with MIIT providing sector-specific policy direction and technical leadership, such as defining standardisation priorities and coordinating industry stakeholders, and SAC overseeing the standardisation framework and formal approval of national

standards. In practice, standards are developed through technical committees administered by SAC.

In June 2024, four departments led by MIIT issued the “Guidelines for the Construction of a National Comprehensive Standardisation System for the Artificial Intelligence Industry” (国家人工智能产业综合标准化体系建设指南), setting out a standard system across algorithms, training data, models, and applications.³⁹ The following year, the “Guidelines for the Development of a Standard System for AI Security Governance in the Industrial and Information Technology Sector” (工业和信息化领域人工智能安全治理标准体系建设指南) gave this system a more explicit security function. It maps AI security governance across seven areas: “governance capacity” (治理能力), “infrastructure security” (基础安全), “cybersecurity” (网络安全), “data security” (数据安全), “algorithm and model security” (算法模型安全), “application security” (应用安全) and “AI-enabled security” (赋能安全) (Figure 2).⁴⁰

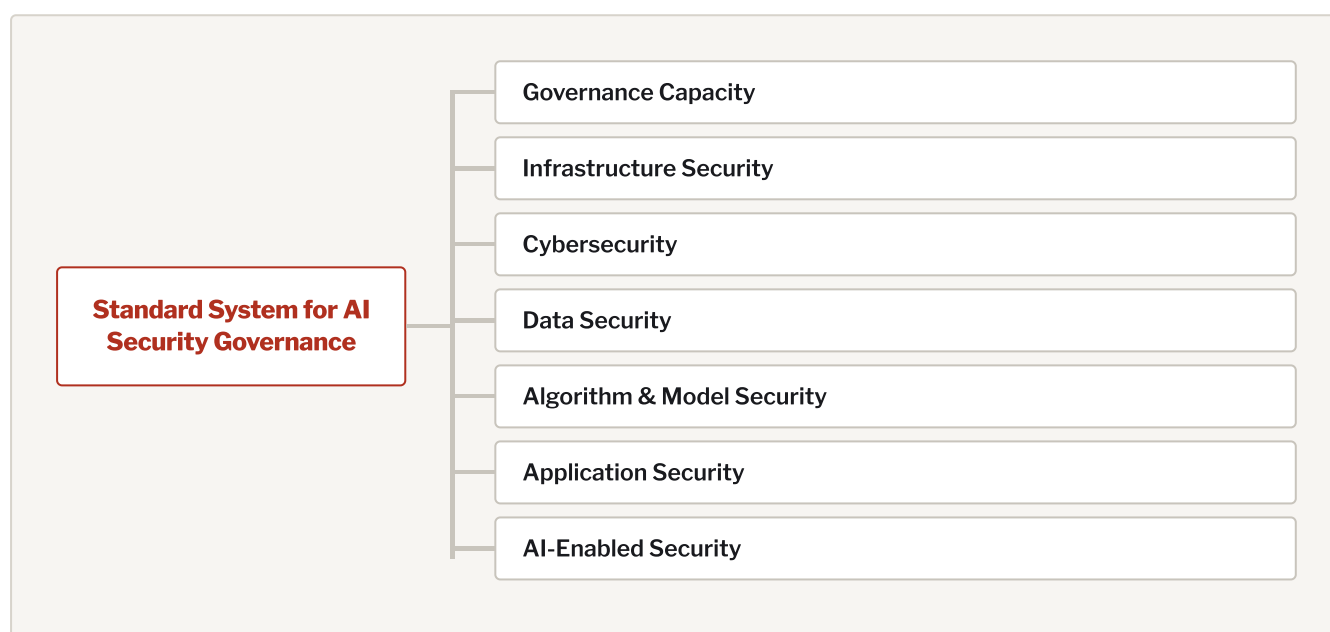


Figure 2: Standard System for AI Security Governance (Source: MIIT, 2024)

Within this standard system, content security is no longer confined to the point at which AI-generated content is displayed to users. It is distributed across the technical process of generation. Training data are assessed before they enter models, algorithms and models are evaluated for their tendency to produce or suppress certain types of content, AI services are tested against politically sensitive prompts, and generated content is made easier to identify and trace after release. Standardisation therefore gives China’s information control a more procedural form, systematically embedding it into the routine technical requirements that firms are expected to follow.

Data Security

Training data is the earliest point at which content control enters the pipeline. Alongside issues such as data quality, privacy, and personal information protection, these data security standards require developers to identify and remove “illegal and harmful” content from the source of data. And in China’s regulatory vocabulary, as discussed, this category is political as much as legal, targeting content that challenges the Party’s narratives and ideological authority. For example, the “Cybersecurity Technology – Generative Artificial Intelligence Data Annotation Security Specification” (网络安全技术 生成式人工智能数据标注安全规范) differentiates different categories of data annotation, with one of them is “security data annotation” (安全性数据标注), which explicitly sets to train AI models to generate content that aligns with the country’s “positive narratives”.⁴¹ Of the whole training data set, at least 3 per cent of

annotated data is required to be such security-related labels under the standard. Beyond security-related annotation, the document also standardises training of data annotation personnel on “legal compliance awareness” in order to strengthen their ability to spot and remove “illegal and harmful” content from training data.

The “Cybersecurity Technology – Security Specification for Generative Artificial Intelligence Pre-Training and Fine-Tuning Data” (网络安全技术 生成式人工智能预训练和优化训练数据规范) requires AI developers to pre-screen all data before using them to train AI models.⁴² With each data type, including texts, images, audio and video clips, AI developers are obliged to establish respective mechanisms to identify, contain, and filter out “sensitive content” that may be “illegal and harmful”. If those “illegal and harmful” content consists more than 5 per cent of the dataset, the dataset should be removed and not to be used as training data.

Algorithm & Model Security

Algorithm and model security standards move control from the data used to train models to the behaviour expected from the models themselves.⁴³ This is where the authorities’ preferred contents are required to be promoted, and those undesired contents being suppressed. The “Information Security Technology – Assessment Specification for Security of Machine Learning Algorithms” (信息安全技术 机器学习算法安全评估规范) directly draws reference from the 2019 “Provisions on the Governance of the Online Information Content Ecosystem”, that algorithmic technologies in China are mandatory to incorporate the content classification laid in the Provisions.⁴⁴ Under this framework, AI developers and service providers have to establish a “positive content reserve” (正能量内容储备), ensuring those “encouraged content” could be more frequently generated and pushed; and on the other hand, to set up a “database of illegal and harmful content features” (违法和不良信息特征库) in order to more effectively filter out “illegal and harmful” content, including even those indirect and implicit.

Application Security

Application security standards test whether these upstream controls work when users interact with AI services.⁴⁵ They bring the older content-control taxonomy into model evaluation, turning political boundaries into test prompts, refusal rates and review procedures.

The 2025 “Cybersecurity Technology – Basic Security Requirements for Generative Artificial Intelligence Service” (网络安全技术 生成式人工智能服务安全基本要求) outlines a detailed framework to assess content security of generative AI.⁴⁶ It first defines major security risks across five categories and 31 subcategories (*Table 3*). This risk classification framework is not only applied at the application level, but is also reused upstream as a reference for assessing training data, demonstrating how content control logic is standardised across the pipeline.⁴⁷

1. Content violating core socialist values	› inciting the subversion of state power or the overthrow of the socialist system › endangering national security and interests, or damaging the national image › inciting separatism, undermining national unity and social stability › promoting terrorism or extremism	› promoting ethnic hatred › promoting violence, obscenity, or pornography › disseminating false or harmful information › other content prohibited by laws and administrative regulations
2. Discriminatory content	› ethnic discrimination › discrimination based on belief or religion › nationality-based discrimination › regional discrimination	› gender discrimination › age discrimination › occupational discrimination › health-related discrimination › other forms of discrimination
3. Commercially unlawful content	› infringement of others' intellectual property rights › violation of business ethics › disclosure of others' trade secrets	› abuse of advantages in algorithms, data, or platforms to engage in monopolistic or unfair competition practices › other unlawful or non-compliant commercial activities
4. Content infringing others' rights and interests	› harm to others' physical or mental health › infringement of others' portrait rights › infringement of others' reputation rights › infringement of others' rights to honour	› infringement of others' privacy rights › infringement of others' personal information rights and interests › infringement of other lawful rights and interests of others
5. Inaccurate or unreliable content	› inaccurate content that seriously contradicts scientific knowledge or mainstream understanding	› unreliable content which, although not containing serious errors, fails to provide meaningful assistance to users

Table 3: Content Categories to Be Identified by Generative AI Service Providers Related to Content Security (Source: SAC, 2025)

The standard turns these risk categories into evaluation tools. Providers are required to build keyword libraries, test-prompt datasets, and classification models covering the identified risks. Politically sensitive content receives particular emphasis, with denser keyword coverage and more targeted test prompts for content violating “core socialist values”. The result is a testing system in which political reliability becomes part of service security.

To support implementation, the China Electronics Standardisation Institute (中国电子技术标准化研究院) and Fudan University (复旦大学) jointly developed the “Benchmark Test Set for Generative Artificial Intelligence Service Security” (生成式人工智能服务安全基准测试集) in 2024.⁴⁸ Of the 2,000 questions included, approximately one-third test whether generated content violates Party ideological requirements (*Figure 3*). Political alignment is therefore built directly into the testing infrastructure.

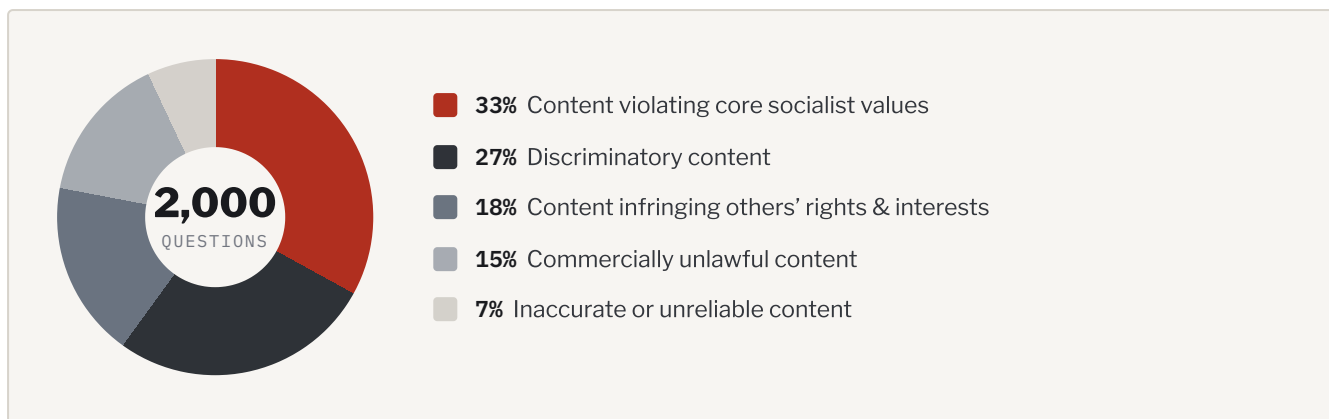


Figure 3: *Types of Questions Testing Generative AI Services in the Benchmark (Source: CESI and Fudan University, 2024)*

The standard also sets concrete evaluation procedures, combining automated screening, model-based detection and manual sampling. For high-risk categories, including politically sensitive content, systems must meet stringent refusal standards, such as a refusal rate above 95 per cent for prohibited prompts. Ideological compliance is therefore made measurable, allowing regulators to assess not only whether a provider has controls in place, but whether those controls produce the required behaviour.

The framework also extends control to users. Providers must monitor repeated attempts to generate prohibited content and impose restrictions on users who repeatedly trigger high-risk outputs. Content security is therefore enforced not only through model behaviour, but also through control over user interaction.

Governance Capacity

Governance capacity standards address the problem of content that still escapes earlier controls. These standards focus on making such content easier to identify, monitor, and trace, thereby ensuring that, in cases where undesired content is still generated and disseminated, responsibility can be effectively traced and enforced.⁴⁹ The “Cybersecurity Technology – Labelling Method for Content Generated by Artificial Intelligence” (网络安全技术 人工智能生成合成内容标识方法) requires both explicit and implicit labels for AI-generated content.⁵⁰ Explicit labels allow users to identify generated content directly, while implicit labels embed identifiers into metadata. These labels make generated images, audio and video easier to trace, allowing responsibility to be assigned when prohibited content circulates.

AI-Enabled Security

AI-enabled security marks the point at which AI becomes not only the object of control, but also a tool for control. In this domain, AI systems are used to monitor online narratives, identify harmful or undesirable content, and support real-time analysis of information flows. Although many standards in this area remain under development, the direction is already clear. The same technology that creates new content security risks is being incorporated into the state’s capacity to detect and manage those risks.⁵¹

The State-Corporate Nexus

This increasingly expansive yet granular control across the generative AI pipeline in China is not unilaterally imposed by the state, but is joined by the private sector. Despite the central role played by MIIT and SAC in standardisation, many of the discussed industrial standards see active private sector involvement, which sits in the drafting committees and provides technical expertise to the policy efforts (*Annex C*). This collaborative relationship between the state and private sector is often praised in China as the effective combination of “proactive government” (有为政府) and “efficient market” (有效市场) that together serve the “socialist market economy” (社会主义市场经济) – essentially a state-led economy with participation and endorsement from the private sector.⁵²

Unsurprisingly, many of China’s most prominent tech firms have consistently engaged in the drafting committees of these standards. Their engagement is significant as these firms occupy key positions across the country’s AI stack. Alibaba and Huawei, through their respective cloud platforms Alibaba Cloud (阿里云, Aliyun) and Huawei Cloud (华为云), together account for half of China’s cloud market, placing them at the centre of the data and computing environments on which AI systems are built and run.⁵³ DeepSeek ranks among the most significant foundation models in China, serving as a base for many generative AI applications in both public and private settings.⁵⁴ More recently, Tencent’s Hunyuan (混元, HY) model has also risen rapidly into one of China’s leading foundation model systems, strengthening Tencent’s position in the country’s AI stack on top of its established advantages in platforms and application services.⁵⁵ The same group of tech firms that is dominant to the AI stack in China is meanwhile joining with the state to set the industrial standards across the generative AI pipeline, determining how the Party’s political control over information could be technically implementable.

This dynamic should not be reduced to a simple story of state control and private-sector compliance. Through standardisation, the Party-state brings major technology firms into the process of governing generative AI. The Party remains central in setting the political direction and defining the boundaries within which firms operate. But by drawing on private firms’ technical expertise and position in the AI stack, the state can make its control system more effective and operable across the generative AI pipeline. Tech firms should therefore be understood not only as targets of regulation, but as part of the governance mechanism itself, with their business agendas expected to remain aligned with state interests.

Reconfiguring the Dynamics

The role of the private sector in China’s expansive technology governance system can be better understood through the reconfiguration of state–private sector dynamics in recent years. The growing influence of the private sector has long concerned the Party, which has worried that powerful business leaders could challenge its political authority and the ideology of “socialist market economy”.⁵⁶ Such concerns were not without basis from the Party’s perspective. Jack Ma, co-founder of Alibaba, declared in 2015 that a weak government would be an opportunity for the market and entrepreneurs to drive the economy, which clearly challenged China’s official narrative on the superiority of state-led economy.⁵⁷ Tycoons such as Sun Dawu, Geng Xiaonan, and Ren Zhiqiang, were also known for their public

criticisms over the country's policies.⁵⁸

In the early 2020s, Beijing took a series of measures to crack down on its private sector, during which the Party consolidated political control over major tech firms. The authorities filed investigations against companies including Alibaba, Didi, and Tencent for a range of regulatory and compliance issues.⁵⁹ The state logic behind was explained during a high-level official meeting chaired by Xi Jinping in 2021, which stressed that the country's economy must adhere to the “politically correct direction” (正确政治方向), while “curbing monopolies and preventing the disorderly expansion of capital” (反对垄断，防止资本无序扩张).⁶⁰

The “Opinions on Strengthening United Front Work in the Private Economy in the New Era” (关于加强新时代民营经济统战工作的意见), issued at the time by the General Office of the Central Committee in 2020, set a clear direction for this reconfiguration. While acknowledging the private sector's economic contribution, the document frames its expanding scale, diversified values, and varied interests as sources of “risks and challenges” (风险挑战) for the Party, and urges officials to strengthen private entrepreneurs' “political, ideological, and emotional identification” (政治认同、思想认同、情感认同) with the CCP and core socialist values.⁶¹ It calls on state actors to strengthen political guidance and institutional engagement with the private sector, ensuring that private enterprise serves state interests under the Party's leadership.

In parallel, the Party also built direct leverage within private firms as part of this reconfiguration. In 2021, an investor linked to the CAC acquired a one per cent stake in Beijing ByteDance Technology, ByteDance's key domestic operating entity, giving the state a board seat in the company that holds operating licences for major China-market apps.⁶² Some observers have described these “golden shares” as a tool that allows state agencies to monitor and influence firms at critical junctures without micromanaging daily operations.⁶³

Incorporation into the Governance System

Notably, these measures did not push private firms out of China's technology economy, but redefined the terms on which they could remain central to it. The 2020 United Front Opinions make this logic explicit. Alongside disciplining the private sector and reaffirming the Party's authority, the document also sought to mobilise private-sector capacity for governance, stating that work with the private economy should help “stimulate the enthusiasm and initiative of private entrepreneurs” in “participating in national governance” (激发民营经济人士在[...]参与国家治理的积极性、主动性), while combining the decisive role of the market with a better-functioning government under China's socialist system.⁶⁴ The aim, in other words, was not to suppress private-sector capacity, but to ensure that it could be incorporated into a Party-led governance system.

This logic became more visible in 2025, after several years of reconfiguring state-private sector relations. In February 2025, Xi Jinping hosted a private enterprise symposium (民营企业座谈会) with business leaders from the private sector, including CEOs from major tech firms such as Huawei, Tencent, and DeepSeek — and most notably, Jack Ma from Alibaba.⁶⁵ Ma's appearance was significant because he had been among the most visible targets of the earlier crackdown and was seen by many Chinese private entrepreneurs as a “political bellwether” (政治风向标).⁶⁶ His return therefore signalled the Party's re-engaging approach toward the private sector.

Nevertheless, this did not signal a softening of the Party's political red lines. Instead, the symposium clarified the terms of private-sector participation in post-reconfiguration China. Xi Jinping told entrepreneurs that “whatever is decided by the Party must be firmly implemented without reservation” (凡是党中央定了的就要坚决执行，不能打折扣), while directing them to be the “builders of socialism with Chinese characteristics” (中国特色社会主义的建设者) under a “unified thought” (统一思想).⁶⁷ Similar language was codified two months later in the Private Economy Promotion Law (民营经济促进法), which serves as a rulebook for the private sector. While reiterating entrepreneurs' responsibility to support the Party's leadership and ideology, the law also promises supportive measures and benefits for enterprises that align with state interests and promotes private-sector participation in “major national strategies” (国家重大战略).⁶⁸

This carrot-and-stick framework creates the conditions for political incorporation. Private tech firms retain space to pursue commercial and technical development, but that space is increasingly conditioned on their ability to align with Party priorities and contribute to state-defined objectives.⁶⁹ This is particularly visible in the AI sector, where key actors either emerged within the new political environment, such as DeepSeek, or developed from legacy platforms once subject to the Party's crackdown, such as Alibaba and Tencent.

Standardisation as Political Incorporation

Standardisation gives this incorporation a practical form. Beijing has long treated industrial standardisation as a policy tool for strengthening state control and security while advancing technological development, especially in emerging technologies such as AI.⁷⁰ In the case of generative AI, where standard-setting extends to information control and content security, participation in standardisation allows firms to remain inside state-backed channels of technological development while helping turn political requirements into technical procedures.

Recent policy reinforces this arrangement. The latest Five-Year Plan, released in early 2026, signals Beijing's willingness to give private firms greater autonomy in technical areas, provided that they remain aligned with state priorities.⁷¹ The 2025 Private Economy Promotion Law reflects the same logic by presenting participation in standardisation as a right open to private firms operating within Party-defined boundaries.⁷² For the Chinese tech firms, joining the standard-setting process offers a way to demonstrate alignment while securing more favourable conditions for business development. For the Party-state, it draws firms' technical expertise into the governance system itself.

Beyond participation in standard-setting, private firms' implementation of these standards shows how political incorporation can shape model behaviour. Alibaba provides a clear example. In June 2025, its Qwen (通义千问) model received the highest-level security certification from a research institute under MIIT, after being tested against a range of content security risks under an MIIT industrial standard that Alibaba Cloud reportedly helped draft.⁷³ In December 2025, one month after the national standard “Basic Security Requirements for Generative Artificial Intelligence Services” came into effect, Qwen was also reported to be among the first Chinese models to pass testing under the standard. Qwen's developer publicly praised the pivotal role of standardisation in AI governance and presented the model as a technical reference for China's AI stack.⁷⁴ This alignment with content security requirements has in turn been validated through state adoption, with Qwen now reportedly used at scale in China's public security sector to support policing and governance work.⁷⁵

DeepSeek offers an even sharper example. When it launched in early 2025, the DeepSeek-R1 model reportedly achieved a 100 per cent success rate in providing “reasonable and positive responses” (合理且正向) in the content security test under the Basic Security Requirements for Generative Artificial Intelligence Service.⁷⁶ Notably, while the model demonstrates near-perfect performance under state-defined content security tests, a 2025 report by the U.S. Center for AI Standards and Innovation shows that it responds at a markedly higher rate to prompts designed to elicit prohibited or harmful outputs, such as phishing-related requests, suggesting that the model is more tightly aligned around political content control than around general safety constraints.⁷⁷ The same report also found that DeepSeek models frequently echoed inaccurate and misleading CCP narratives.⁷⁸ This finding is consistent with other observations that DeepSeek does not merely refuse sensitive prompts, but structures “accuracy” around official sources and Party-aligned narratives.⁷⁹ Political incorporation therefore appears not only in firms’ participation in standard-setting, but also in the content-control logic embedded in model behaviour.

Exporting China Model

The expanding control mechanism within Chinese generative AI systems does not affect only China's domestic digital space. As Beijing pursues global technological leadership, the control logic embedded into Chinese models travels beyond the country's borders through the diffusion of the models.

Scale of Adoption

Driven by high customisability, low cost, and increasingly competitive performance, Chinese generative AI models are being adopted at a growing scale outside China.⁸⁰ On Hugging Face, a major global repository where developers share and download AI models, Chinese models have overtaken U.S. counterparts in cumulative downloads since the summer of 2025.⁸¹ By August 2025, a partner at Andreessen Horowitz (a16z), a U.S. venture capital firm focused on technology start-ups, estimated that 80 per cent of American AI start-ups were building on Chinese open-source base models.⁸² Adoption has also reached mainstream platforms, with Airbnb reportedly deploying Alibaba's Qwen models for customer-facing services.⁸³ Major Western cloud providers, including AWS, Azure, and Google Cloud, now offer managed access to Chinese models such as DeepSeek, allowing developers in liberal democracies to integrate these systems into applications without building models from scratch.⁸⁴

Beyond these visible cases, Chinese models are often adopted in less traceable ways, meaning their global use likely extends beyond what available figures capture. Developers can connect to models through application programming interfaces (APIs), access them through Western cloud infrastructure, or build on open-source base models and their derivatives. Platforms such as OpenRouter act as intermediaries that route requests across multiple models, with Chinese systems accounting for a substantial share of usage.⁸⁵ At the same time, model provenance is becoming harder to identify. In February 2026, 70 per cent of newly fine-tuned models on Hugging Face were built on Chinese base models, often without clear attribution, while third-party tools such as Cursor's Composer 2 have been found to rely on Chinese models despite presenting themselves as independent systems.⁸⁶

Persistent Control

As Chinese generative AI models are adopted abroad, the control logic shaped by China's regulatory system can travel with them. As discussed, these models are trained and developed under requirements to align with Party-defined narratives such as "core socialist values". This narrative control is notably not limited to issues directly involving China, but also reproduces Beijing's broader geopolitical worldview, favouring authoritarian partners while taking a more hostile view of liberal democracies. The Estonian Foreign Intelligence Service, for example, found that DeepSeek distorted discussion of Russia's invasion of Ukraine by echoing Kremlin narratives and repeatedly presenting China as a source of peace and stability.⁸⁷ Alibaba's Qwen-3 model has also reportedly softened criticism of Uganda's government and avoided domestic corruption issues that have previously triggered public anger.⁸⁸ Other research similarly finds that Chinese models such as DeepSeek often embed anti-U.S. sentiment in their responses.⁸⁹

When such models are deployed outside China, these requirements do not simply disappear. A China

Media Project study reports that pro-CCP bias remained in some international deployments of DeepSeek models, even after attempts to remove it.⁹⁰ Research by Anthropic has shown that behaviours learned during training can be embedded directly in model weights and persist across deployment environments, meaning outputs remain shaped by training and alignment even when the model is hosted and deployed elsewhere.⁹¹ The extent of that persistence may vary by deployment mode. Models accessed through APIs leave developers dependent on upstream alignment decisions, since they cannot modify the underlying weights. Even where open-weight adopters can fine-tune models, such modifications may only partially overwrite political bias, and whether they can remove more deeply embedded behaviours remains uncertain.⁹²

This creates a form of influence distinct from conventional information manipulation. Rather than removing or amplifying content after it appears, generative AI systems shape information at the point of production. As Chinese models are integrated into tools and applications abroad, they can influence what information is generated, how events are described, and which narratives are treated as acceptable or authoritative. These effects operate through model behaviour rather than overt censorship, making them less visible but potentially broader in reach.

Dependency Risk

The risk grows as adoption deepens. Integration into developer frameworks, APIs, cloud services, and downstream applications creates technological dependency, with systems and workflows built around specific models. Once such dependencies form, substitution becomes costly and complex. The UK's removal of Huawei equipment from its 5G infrastructure, which cost £2 billion and delayed rollout by several years, illustrates how difficult technological dependency can be to reverse once established.⁹³ In generative AI, dependency accumulates less through a single procurement decision than through repeated integration across the software stack. As Chinese models become embedded in the infrastructure through which information is generated and processed, the control logic built into their design can scale with them, shaping information environments abroad without direct intervention from Chinese authorities.

Discussion

This report has shown that political bias and content control in Chinese AI models are not only output-level problems, but the result of a governance system that makes such control technically operational. Driven by anxiety over maintaining ideological security and information control in the age of generative AI, Chinese authorities, in collaboration with private technology firms, are extending information control across the generative AI pipeline through expanding policy efforts and industrial standard-setting. The Party's political requirements are thereby translated into granular technical procedures that shape training data, model behaviour, service testing, and content traceability, embedding information control into the process before content is generated.

This carries implications beyond China's domestic information space. Beyond information operations already identified under the framework of FIMI, Chinese AI systems themselves could become channels of influence in the global information landscape. There is currently limited evidence to determine whether such external influence is a deliberate objective of the Chinese authorities or an unintended byproduct of their quest for content security within Chinese borders. In either case, the issue remains consequential as governments, firms, and developers accelerate AI adoption, and as the low cost and growing competitiveness of Chinese models speed their diffusion into foreign digital infrastructure.⁹⁴

Looking ahead, two areas warrant closer attention:

- **The potential influence of Chinese AI standards on global AI governance.** International regulators and developers should recognise that political and ideological requirements in China are often incorporated under the seemingly neutral language of AI safety and security. In China's governance framework, data security, algorithm security, application security, and content security are closely connected to political reliability and compliance with core socialist values. This differs sharply from dominant understandings of AI safety in liberal democracies, where the concept is more commonly associated with risks such as model reliability, discrimination, misuse, privacy, and systemic harm. The distinction matters as China seeks a greater role in international AI standard-setting alongside the growing global traction of its models and applications.⁹⁵ If regulators and developers in liberal democracies overlook this distinction, they risk treating politically loaded standards as neutral technical safeguards, allowing authoritarian approaches to content control to shape the rules of global AI governance.
- **The diffusion of Chinese AI models in sectors consequential to the information environment.** Although Chinese models and applications are gaining traction worldwide, the actual picture of that diffusion remains difficult to assess with precision. Public indicators such as model downloads, platform availability, web traffic, and token usage may suggest broader adoption, but they reveal little about the actual use cases of these Chinese models, and the extent to which they are being integrated into sectors consequential to the information environment, such as media, education, public communication, and civic services. This uncertainty is itself a governance issue. Greater scrutiny is therefore needed in these sectors, including clearer disclosure of model origins and assessment of whether embedded content-control logic affects system outputs.

Annex A

Selected Official Documents in relation to China's Information Control

DATE	DOCUMENT TITLE	ISSUING DEPARTMENT	TYPE
1 July 2015	National Security Law of the People's Republic of China (中华人民共和国国家安全法)	National People's Congress	Law
17 March 2016	The 13th Five-Year Plan for National Economic and Social Development (国民经济和社会发展第十三个五年规划纲要)	National Development and Reform Commission (NDRC)	National development plan
1 June 2017	Cybersecurity Law of the People's Republic of China (中华人民共和国网络安全法)	National People's Congress Standing Committee	Law
1 March 2020	Provisions on the Governance of the Online Information Content Ecosystem (网络信息内容生态治理规定)	CAC	Administrative rule
15 September 2020	Opinions on Strengthening United Front Work with the Private Sector in the New Era (关于加强新时代民营经济统战工作的意见)	General Office of the CCP Central Committee	Party policy document
10 October 2021	National Standardisation Development Outline (国家标准化发展纲要)	CCP Central Committee and State Council	National standard
1 March 2022	Provisions on the Management of Algorithmic Recommendations in Internet Information Services (互联网信息服务算法推荐管理规定)	CAC, MIIT, MPS, SAMR	Administrative rule
15 February 2022	Cybersecurity Review Measures (网络安全审查办法)	CAC	Administrative measure
10 January 2023	Provisions on the Administration of Deep Synthesis for Internet Information Services (互联网信息服务深度合成管理规定)	CAC, MIIT, MPS	Administrative rule
15 August 2023	Interim Measures for the Administration of Generative Artificial Intelligence Services (生成式人工智能服务管理暂行办法)	CAC, NDRC, Ministry of Education, MOST, MIIT, MPS, NRTA	Administrative measure
6 August 2023	Information Security Technology – Assessment Specification for Security of Machine Learning Algorithms (信息安全技术 机器学习算法安全评估规范)	SAMR and SAC	National standard
5 June 2024	Guidelines for the Development of a Comprehensive Standardisation System for the National Artificial Intelligence Industry (国家人工智能产业综合标准化体系建设指南)	MIIT and other departments	Technical guidance
9 August 2024	Security Benchmark Dataset for Generative Artificial Intelligence Services (生成式人工智能服务安全基准测试集)	China Electronics Standardisation Institute and Fudan University	Official benchmark
September 2024	Artificial Intelligence Safety Governance Framework (人工智能安全治理框架)	National Technical Committee 260 on Cybersecurity of the Standardization Administration of China, CAC	Policy framework
25 April 2025	Cybersecurity Technology – Generative Artificial Intelligence Data Annotation Security Specification (网络安全技术 生成式人工智能数据标注安全规范)	SAMR and SAC	National standard

25 April 2025	Cybersecurity Technology – Security Specification for Generative Artificial Intelligence Pre-Training and Fine-Tuning Data (网络安全技术 生成式人工智能预训练和优化训练数据规范)	SAMR and SAC	National standard
25 April 2025	Cybersecurity Technology – Basic Security Requirements for Generative Artificial Intelligence Service (网络安全技术 生成式人工智能服务安全基本要求)	SAMR and SAC	National standard
30 April 2025	Law of the People's Republic of China on the Promotion of the Private Economy (中华人民共和国民营经济促进法)	National People's Congress; Ministry of Justice	Law
18 July 2025	Guidelines for the Development of a Standardisation System for Artificial Intelligence Security Governance in the Industrial and Information Technology Sectors (工业和信息化领域人工智能安全治理标准体系建设指南)	Technical Committee on Artificial Intelligence Standardisation and MIIT	Technical guidance
September 2025	Practical Guide to Cybersecurity Standards – Guidelines for Security Incident Response in Generative AI Services (网络安全标准实践指南：生成式人工智能服务安全应急响应指南)	National Technical Committee 260 on Cybersecurity of SAC	Technical guidance
28 February 2026	Cybersecurity Technology – Labelling Method for Content Generated by Artificial Intelligence (网络安全技术 人工智能生成合成内容标识方法)	State Administration for Market Regulation and National Standardisation Administration; TC260	National standard
12 March 2026	The 15th Five-Year Plan for National Economic and Social Development (中华人民共和国国民经济和社会发展第十五个五年规划纲要)	National People's Congress; NDRC	National development plan
17 June 2026	More Just and Equitable Global Governance: China's Principles, Proposals and Actions (更加公正合理的全球治理：中国的理念、主张与行动)	State Council Information Office of the People's Republic of China	White paper

Annex B

Industrial Standards Related to Content Security of Generative AI

For details, see Part II of this report.

STANDARD	IDENTIFIER	GOVERNANCE AREA	ISSUED	EFFECTIVE
Information Security Technology – Assessment Specification for Security of Machine Learning Algorithms(信息安全技术 机器学习算法安全评估规范)	GB/T 42888-2023	Algorithm & Model Security (算法模型安全)	6 August 2023	1 March 2024
Cybersecurity Technology – Labelling Method for Content Generated by Artificial Intelligence(网络安全技术 人工智能生成合成内容标识方法)	GB 45438-2025	Governance Capacity (治理能力)	28 February 2025	1 September 2025
Cybersecurity Technology – Security Specification for Generative Artificial Intelligence Pre-Training and Fine-Tuning Data(网络安全技术 生成式人工智能预训练和优化训练数据规范)	GB/T 45652-2025	Data Security (数据安全)	25 April 2025	1 November 2025
Cybersecurity Technology – Basic Security Requirements for Generative Artificial Intelligence Service(网络安全技术 生成式人工智能服务安全基本要求)	GB/T 45654-2025	Application Security (应用安全)	25 April 2025	1 November 2025
Cybersecurity Technology – Generative Artificial Intelligence Data Annotation Security Specification(网络安全技术 生成式人工智能数据标注安全规范)	GB/T 45674-2025	Data Security (数据安全)	25 April 2025	1 November 2025

Annex C

Major Private Tech Firms Participating in China's AI Standard-Setting Related to Content Security

For details, see Part III of this report.

COMPANY	SIGNIFICANCE IN CHINA'S AI STACK	STANDARDS CONTRIBUTED TO DRAFTING*
Alibaba	Full-stack AI provider spanning semiconductors, cloud infrastructure, enterprise AI services, and foundation-model development through Qwen. Its AI chip development, including the Zhenwu series, supports training and inference, while Alibaba Cloud provides cloud platforms, databases, security resources, and model-service infrastructure for public- and private-sector use cases.	› Cybersecurity Technology – Generative Artificial Intelligence Data Annotation Security Specification › Cybersecurity Technology – Security Specification for Generative Artificial Intelligence Pre-Training and Fine-Tuning Data › Cybersecurity Technology – Basic Security Requirements for Generative Artificial Intelligence Service › Cybersecurity Technology – Labelling Method for Content Generated by Artificial Intelligence
Huawei	AI infrastructure provider combining Ascend chips, cloud services, AI software platforms, and the Pangu model family. Its role is concentrated in hardware-software integration, industrial AI deployment, and China's broader push for technological self-reliance.	› Cybersecurity Technology – Generative Artificial Intelligence Data Annotation Security Specification › Cybersecurity Technology – Security Specification for Generative Artificial Intelligence Pre-Training and Fine-Tuning Data › Cybersecurity Technology – Basic Security Requirements for Generative Artificial Intelligence Service
DeepSeek	Foundation-model developer focused on cost-efficient large language models. Its models have strengthened China's position in frontier AI development and helped narrow the perceived performance gap between Chinese and leading U.S. AI systems.	› Cybersecurity Technology – Basic Security Requirements for Generative Artificial Intelligence Service
Baidu	Full-stack AI provider spanning Kunlun chips, PaddlePaddle, Baidu AI Cloud, search infrastructure, and the ERNIE model family. Its role is strengthened by the integration of foundation models into search, cloud services, and downstream AI applications.	› Cybersecurity Technology – Generative Artificial Intelligence Data Annotation Security Specification › Cybersecurity Technology – Security Specification for Generative Artificial Intelligence Pre-Training and Fine-Tuning Data › Cybersecurity Technology – Basic Security Requirements for Generative Artificial Intelligence Service
Xiaomi	Consumer-technology firm deploying AI across smartphones, smart-home devices, home appliances, electric vehicles, and edge-AI applications. Its XRING chips, MiMo model family, and "Human x Car x Home" ecosystem place AI within connected consumer environments.	› Cybersecurity Technology – Security Specification for Generative Artificial Intelligence Pre-Training and Fine-Tuning Data › Cybersecurity Technology – Generative Artificial Intelligence Data Annotation Security Specification
RealAI	AI security provider focused on model testing, adversarial robustness, red-teaming, and security tools for AI systems. Its products support vulnerability detection, cybersecurity testing, and protection of systems such as facial-recognition applications.	› Cybersecurity Technology – Generative Artificial Intelligence Data Annotation Security Specification › Cybersecurity Technology – Security Specification for Generative Artificial Intelligence Pre-Training and Fine-Tuning Data › Cybersecurity Technology - Basic Security Requirements for Generative Artificial Intelligence Service
Ant Group	Fintech platform applying AI to payments, SME services, merchant operations, and agentic business tools. Its Ling model family, Alipay AI Pay, and Antom Copilot support the integration of generative AI into financial and commercial-service infrastructure.	› Cybersecurity Technology – Generative Artificial Intelligence Data Annotation Security Specification › Cybersecurity Technology – Security Specification for Generative Artificial Intelligence Pre-Training and Fine-Tuning Data › Cybersecurity Technology – Basic Security Requirements for Generative Artificial Intelligence Service

* Includes only standards discussed in Part II of this report that are directly related to information control and content security.

Endnotes

FOREWORD

- 1 Pope Leo XIV, Magnifica Humanitas: On Safeguarding the Human Person in the Time of Artificial Intelligence, Holy See, 15 May 2026, <<https://www.vatican.va/content/leo-xiv/en/encyclicals/documents/20260515-magnifica-humanitas.html>> [accessed 29 June 2026].
- 2 Gerrit De Vynck and Isaac Arnsdorf, “The U.S. Government Will Decide Who Gets to Use the Latest American AI Technology”, The Washington Post, 26 June 2026, <<https://www.washingtonpost.com/technology/2026/06/26/openai-says-us-government-will-vet-users-its-latest-ai-model/>> [accessed 29 June 2026].
- 3 The State Council Information Office of the People’s Republic of China, “More Just and Equitable Global Governance: China’s Principles, Proposals and Actions”, State Council Information Office, 17 June 2026, <https://web.archive.org/web/20260629103230/http://english.sio.gov.cn/whitepapers/2026-06/17/content_118552858_5.html> [accessed 29 June 2026].
- 4 Yaqiu Wang, “The Missing Resistance in China’s AI Debate”, Lawfare, 25 June 2026, <<https://www.lawfaremedia.org/article/the-missing-resistance-in-china-s-ai-debate>> [accessed 29 June 2026].

REPORT

- 5 Jennifer Collins, “What Does DeepSeek Mean For Ai’s Environmental Impact?”, DW, 30 January 2025, <<https://archive.ph/skRhy>> [accessed 29 May 2026].
- 6 Valentin Weber, “Why DeepSeek Is So Dangerous”, *Journal of Democracy*, March 2025, <<https://archive.ph/nPPw1>> [accessed 29 May 2026].
- 7 Fergus Ryan and others, “The Party’s AI: How China’s New AI Systems Are Reshaping Human Rights”, ASPI, 1 December 2025, <<https://www.aspi.org.au/report/the-partys-ai-how-chinas-new-ai-systems-are-reshaping-human-rights/>> [accessed 29 May 2026].
- 8 Brett Goldstein and Brett Benson, “The GoLaxy Documents: Inside One Chinese Company’s AI-Driven Influence Machine”, Wicked Problems Lab, 5 August 2025, <<https://www.vanderbilt.edu/national-security/wicked-problems-lab/golaxy/>> [accessed 29 May 2026].
- 9 Julian E. Barnes, “China turns to A.I. in Information Warfare”, *The New York Times*, 6 August 2025, <<https://www.nytimes.com/2025/08/06/us/politics/china-artificial-intelligence-information-warfare.html>> [accessed 29 May 2026]; Doublethink Lab, Digital Intelligence Team and Athena Tong, “The Rise of AI in PRC Influence Operations: Nine Takeaways from the GoLaxy Documents”, Medium, 4 March 2026, <<https://medium.com/doublethinklab/the-rise-of-ai-in-prc-influence-operations-nine-takeaways-from-the-golaxy-documents-2d6617a753e5>> [accessed 29 May 2026].
- 10 Felix Simon, Rasmus Kleis Nielsen, and Richard Fletcher, “Generative AI and News Report 2025: How People Think About AI’s Role in Journalism and Society”, Reuters Institute and University of Oxford, October 2025, <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2025-10/Gen_AI_and_News_Report_2025.pdf> [accessed 29 May 2026]; Oliver Fenton, Jacob Bellamy, and Eva Radukic, “AI Skills for Life and Work: General Public Survey Findings”, UK Government, 29 January 2026, <<https://www.gov.uk/government/publications/ai-skills-for-life-and-work-general-public-survey-findings/ai-skills-for-life-and-work-general-public-survey-findings>> [accessed 29 May 2026]; Colleen McClain and others, “How Teens Use and View AI”, Pew Research Center, 24 February 2026, <<https://www.pewresearch.org/internet/2026/02/24/how-teens-use-and-view-ai/>> [accessed 29 May 2026].
- 11 Xue Er Shi Xi Studio, “What are the key priorities for building a technologically advanced nation? General Secretary Xi Jinping has highlighted five points” [建设科技强国，重点抓什么？习近平总书记强调了五点], *Qiushi*, 6 April 2025, <<https://web.archive.org/web/20260513161429/https://www.qstheory.cn/20250406/a53629be63794d65aac29ac1f87ef4bd/c.html>> [accessed 29 May 2026].
- 12 OpenRouter, “AI Model Rankings”, <<https://openrouter.ai/rankings?view=month>> [accessed 29 May 2026].
- 13 People’s Daily Online, “Upholding Marxism as Our Guiding Principle, and Firmly Maintaining Leadership, Control and the Right to Define the Discourse in Ideological Work” [坚持以马克思主义为指导，牢牢掌握意识形态工作领导权、管理权、话语权], CPC News, 4 December 2018, <<https://archive.ph/Z4Dos>> [accessed 4 June 2026]; Jidong Zhu, “Reject the Tendency to be Overly Lenient and Maintain a Fighting Spirit” [拒绝好人主义永葆斗争精神], *People’s Tribune*, 3 June 2022, <<https://archive.ph/ICV66>> [accessed 4 June 2026].
- 14 Government of the People’s Republic of China, “The 13th Five Year Plan: For Economic and Social Development (Full Text)” [国民经济和社会发展第十三个五年规划纲要(全文)], 12371.cn, 17 March 2016, <<https://archive.ph/QLNp2>> [accessed 4 June 2026].
- 15 Bill Bishop, “Engineers of the Soul: Ideology in Xi Jinping’s China by John Garnaut”, Sinocism, 17 January 2019, <<https://sinocism.com/p/en-gineers-of-the-soul-ideology-in>> [accessed 4 June 2026].
- 16 Rogier Creemers, “Communiqué on the Current State of the Ideological Sphere (Document No. 9)”, DigiChina, 22 April 2013, <<https://digi.hina.stanford.edu/work/communique-on-the-current-state-of-the-ideological-sphere-document-no-9/>> [accessed 4 June 2026].
- 17 Creemers, “Communiqué on the Current State of the Ideological Sphere (Document No. 9)”.
- 18 “Consolidate and strengthen mainstream ideology and culture in cyberspace” [巩固壮大网络空间主流思想文化], *Qiushi*, 9 June 2025, <<https://web.archive.org/web/20250624015121/http://www.qstheory.cn/20250609/c96934d24e2e41e38c31de29e8a7e862/c.html>> [accessed 4 June 2026].
- 19 Creemers, “Communiqué on the Current State of the Ideological Sphere (Document No. 9)”.

- 20 National People's Congress of the People's Republic of China, "National Security Law of the People's Republic of China" [中华人民共和国国家安全法], NPC News, 10 July 2015, <<https://web.archive.org/web/20251023094138/http://npc.people.com.cn/n/2015/0710/c14576-27285049.html>> [accessed 4 June 2026].
- 21 Resilience Innovation Lab, "An Unbreakable Wall? Challenges and Prospects of Defending Freedom of Information in China", *Resilience Innovation Lab*, 17 April 2025, <https://resilienceinnovationlab.org/wp-content/uploads/2025/04/RIL_Unbreakable_Wall.pdf> [accessed 4 June 2026].
- 22 Ryan and others, "The Party's AI: How China's New AI Systems Are Reshaping Human Rights"; Barnes, "China turns to A.I. in Information Warfare".
- 23 Bochum Yin, "A Think Tank's Proposal for Managing Public Opinion Risks Associated with Generative Artificial Intelligence" [生成式人工智能舆论风险治理的智库方案], China Social Sciences Net, 25 November 2025, <https://web.archive.org/web/20251215091556/https://www.cssn.cn/skgz/bwyc/202511/t20251125_5952089.shtml> [accessed 4 June 2026].
- 24 National Technical Committee 260 on Cybersecurity of Standardization Administration of China (TC260), "Artificial Intelligence Safety Governance Framework" [人工智能安全治理框架], Cyberspace Administration of China, September 2024, <<https://web.archive.org/web/20251206065533/https://www.cac.gov.cn/rootimages/uploadimg/1727568303900999/1727568303900999.pdf>> [accessed 26 June 2026].
- 25 ScaDS.AI, "Cracking the Code: The Black Box Problem of AI", ScaDS.AI, 19 July 2023, <<https://scads.ai/cracking-the-code-the-black-box-problem-of-ai/>> [accessed 4 June 2026].
- 26 "Usage Statistics of Content Languages for Websites", W3 Techs, <https://w3techs.com/technologies/overview/content_language> [accessed 4 June 2026].
- 27 Wang Jinying, Zhang Yuantong, and Ning Yali, "Challenges and Responses to the Use of Generative Artificial Intelligence in Innovating Models of Ideological and Political Education in Higher Education Institutions" [生成式人工智能赋能高校思想政治教育模式创新的挑战与应对], *Journal of Zhengzhou University of Light Industry (Social Science Edition)*, 2025, <<https://archive.ph/NTUyy>> [accessed 2 July 2026].
- 28 Chen Shihua, "Ideological Security Risks in Large-Scale Model Corpora and Approaches to Their Governance" [大模型语料库的意识形态安全风险及其治理路径], *China Cyberspace*, 29 September 2025, <<https://web.archive.org/web/20251117053456/https://www.secrss.com/articles/83603>> [accessed 4 June 2026].
- 29 National Technical Committee 260 on Cybersecurity of SAC, "Practical Guide to Cybersecurity Standards— Guidelines for Security Incident Response in Generative AI Services" [网络安全标准实践指南——生成式人工智能服务安全应急响应指南], September 2025, <<https://web.archive.org/web/20260213090430/https://www.tc260.org.cn/upload/2025-09-15/1757913511246034883.pdf>> [accessed 4 June 2026].
- 30 Qiao Zhe, "Risks and Countermeasures of Artificial Intelligence Generated Content Technology in Content Security Governance" [人工智能生成内容技术在内容安全治理领域的风险和对策], *Telecommunications Science*, 10 November 2023, doi: 10.11959/j.issn.1000-0801.2023190.
- 31 Cyberspace Administration of China, "Cybersecurity Law of the People's Republic of China" [中华人民共和国网络安全法], Cyberspace Administration of China, 7 November 2016, <https://web.archive.org/web/20260521114122/https://www.cac.gov.cn/2016-11/07/c_1119867116.htm> [accessed 29 June 2026].
- 32 Cyberspace Administration of China, "Provisions on the Governance of the Online Information Content Ecosystem" [网络信息内容生态治理规定], Cyberspace Administration of China, 20 December 2019, <https://web.archive.org/web/20260429112535/https://www.cac.gov.cn/2019-12/20/c_1578375159509309.htm> [accessed 29 June 2026].
- 33 Cyberspace Administration of China, "Provisions on the Management of Algorithmic Recommendations in Internet Information Services" [互联网信息服务算法推荐管理规定], Cyberspace Administration of China, 4 January 2022, <https://web.archive.org/web/20260430133710/https://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm> [accessed 29 June 2026].
- 34 Cyberspace Administration of China, "Cybersecurity Review Measures" [网络安全审查办法], Cyberspace Administration of China, 4 January 2022, <https://web.archive.org/web/20260521051919/https://www.cac.gov.cn/2022-01/04/c_1642894602182845.htm> [accessed 29 June 2026].
- 35 National People's Congress (2016), "Cybersecurity Law of the People's Republic of China", Cyberspace Administration of China, 7 November 2016, <https://web.archive.org/web/20260412155837/https://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm> [accessed 4 June 2026].
- 36 Cyberspace Administration of China and others, "Interim Measures for the Administration of Generative Artificial Intelligence Services" [生成式人工智能服务管理暂行办法], Cyberspace Administration of China, 27 July 2023, <https://web.archive.org/web/20260517183958/https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm> [accessed 4 June 2026].
- 37 Kim Martineau, "A Toxic Language Filter Built for Speed", IBM Research, 9 September 2024, <<https://research.ibm.com/blog/efficient-llm-hap-detector>> [accessed 4 June 2026]; "End-to-End Generative AI Architecture Explained", Impressico Business Solutions, 16 February 2026, <<https://www.impressico.com/blog/end-to-end-generative-ai-architecture/>>, [accessed 4 June 2026].
- 38 The National Standardisation Administration, also known as the Standardisation Administration of China, retains the abbreviation SAC.
- 39 Ministry of Industry and Information Technology of the People's Republic of China and others, "Guidelines for the Development of a Comprehensive Standardisation System for the National Artificial Intelligence Industry" [国家人工智能产业综合标准化体系建设指南], Government of the People's Republic of China, 5 June 2024, <https://web.archive.org/web/20251109164914/https://www.gov.cn/zhengce/zhengceku/202407/content_6960720.htm> [accessed 4 June 2026].

- 40 Technical Committee on Artificial Intelligence Standardisation and Ministry of Industry and Information Technology, “Guidelines for the Development of a Standardisation System for Artificial Intelligence Security Governance in the Industrial and Information Technology Sectors” [工业和信息化领域人工智能安全治理标准体系建设指南], 18 July 2025, <<https://web.archive.org/web/20260419092248/https://caict-ilm-portal-storage.oss-cn-beijing.aliyuncs.com/6153dd34-d7fc-4d24-97b5-d40fc48105c5>> [accessed 4 June 2026].
- 41 State Administration for Market Regulation and National Standardisation Administration, “Cybersecurity Technology – Generative Artificial Intelligence Data Annotation Security Specification” [网络安全技术 生成式人工智能数据标注安全规范], 25 April 2025, <https://web.archive.org/web/20260319025539/https://content.mlex.com/Attachments/2025-05-23_CZT297CHX8R49J68/SAMR_NSA_GenAI_data_annotation_security_specification.pdf> [accessed 4 June 2026].
- 42 State Administration for Market Regulation and National Standardisation Administration, “Cybersecurity Technology – Security Specification for Generative Artificial Intelligence Pre-Training and Fine-Tuning Data” [网络安全技术 生成式人工智能预训练和优化训练数据规范], 25 April 2025, <https://web.archive.org/web/20260319025539/https://content.mlex.com/Attachments/2025-05-23_CZT297CHX8R49J68/SAMR_NSA_Security_Specification_GenAI_pre-training_fine-tuning_Data.pdf> [accessed 5 June 2026].
- 43 Technical Committee on Artificial Intelligence Standardisation and Ministry of Industry and Information Technology, “Guidelines for the Development of a Standardisation System for Artificial Intelligence Security Governance in the Industrial and Information Technology Sectors”.
- 44 State Administration for Market Regulation and National Standardisation Administration, “Information Security Technology – Assessment Specification for Security of Machine Learning Algorithms” [信息安全技术 机器学习算法安全评估规范], 6 August 2023, <<https://web.archive.org/web/20250719173517/https://www.iscn.org.cn/uploadfile/2025/0702/2025070203223913.pdf>> [accessed 5 June 2026].
- 45 Technical Committee on Artificial Intelligence Standardisation and Ministry of Industry and Information Technology, “Guidelines for the Development of a Standardisation System for Artificial Intelligence Security Governance in the Industrial and Information Technology Sectors”.
- 46 State Administration for Market Regulation and National Standardisation Administration, “Cybersecurity Technology – Basic Security Requirements for Generative Artificial Intelligence Service” [网络安全技术 生成式人工智能服务安全基本要求], 25 April 2025, <https://web.archive.org/web/20260503014803/https://content.mlex.com/Attachments/2025-05-23_CZT297CHX8R49J68/SAMR_NSA_Basic_Security_Requirements_GenAI_service.pdf> [accessed 5 June 2026].
- 47 See State Administration for Market Regulation and National Standardisation Administration, “Cybersecurity Technology – Generative Artificial Intelligence Data Annotation Security Specification”; State Administration for Market Regulation and National Standardisation Administration, “Cybersecurity Technology – Security Specification for Generative Artificial Intelligence Pre-Training and Fine-Tuning Data”.
- 48 ESI and Fudan University, “Security Benchmark Dataset for Generative Artificial Intelligence Services” [生成式人工智能服务安全基准测试集], Fudan University, 9 August 2024, <<https://web.archive.org/web/20251030185957/http://47.103.21.113:8012/>> [accessed 5 June 2026].
- 49 Technical Committee on Artificial Intelligence Standardisation and Ministry of Industry and Information Technology, “Guidelines for the Development of a Standardisation System for Artificial Intelligence Security Governance in the Industrial and Information Technology Sectors”.
- 50 State Administration for Market Regulation and National Standardisation Administration, “Cybersecurity Technology – Labeling Method for Content Generated by Artificial Intelligence” [网络安全技术 人工智能生成合成内容标识方法], 28 February 2026, <<https://web.archive.org/web/20250327124551/https://www.tc260.org.cn/upload/2025-03-15/1742009439794081593.pdf>> [accessed 5 June 2026].
- 51 Technical Committee on Artificial Intelligence Standardisation and Ministry of Industry and Information Technology, “Guidelines for the Development of a Standardisation System for Artificial Intelligence Security Governance in the Industrial and Information Technology Sectors”.
- 52 Wu Zhenglong, “Combining an Efficient Market with an Active Government”, *Qiushi*, 17 November 2025, <<https://web.archive.org/web/20251211231512/https://www.qstheory.cn/20251117/306d2756cbda47128ed7e48bb5c8474a/c.html>> [accessed 5 June 2026]; Li Chao, “August press conference of the National Development and Reform Commission”, National Development and Reform Commission, 28 August 2025, <<https://web.archive.org/web/20250907105810/https://www.ndrc.gov.cn/xwdt/wszb/2025yxwfbh/>> [accessed 5 June 2026].
- 53 Wency Chen, “Alibaba Holds Wide Lead over Rivals Bytedance, Huawei, Tencent in China’s AI Cloud Market”, *SCMP*, 10 September 2025, <<https://archive.ph/DjvqG>> [accessed 5 June 2026].
- 54 National Development and Reform Commission, “Lhasa, Tibet: Empowering Public Services with Large-Scale AI Models to Set a New Benchmark for Smart Governance on the Plateau” [西藏拉萨：以人工智能大模型赋能政务服务 打造高原智慧政务新标杆] <https://web.archive.org/web/20260528192202/https://www.ndrc.gov.cn/xwdt/ztl/xhyshj/dfdt/202601/t20260122_1403407_ext.html> [accessed 5 June 2026]; Eleanor Olcott and Wenjie Ding, “DeepSeek Spreads Across China with Beijing’s Backing”, *Financial Times*, 27 February 2025, <<https://www.ft.com/content/5684fb1f-1a84-4542-8fe9-2fcae9653f87?syn-25a6b1a6=1>> [accessed 5 June 2026].

- 55 Zhou Jingjie, "Tencent releases Q1 financial results: Hy3 preview calls remain top of the OpenRouter rankings, WorkBuddy leads in daily active users, and core businesses see across-the-board efficiency gains" [腾讯发布Q1财报：Hy3 preview调用量稳居OpenRouter榜首，WorkBuddy日活领跑，核心业务全面提效], *Xinhua*, 13 May 2026, <<https://archive.ph/wip/qKfFz>> [accessed 5 June 2026]; Tencent, "Tencent Announces First-Quarter Results for 2026 Significant Performance Improvements in the Hunyuan Iteration; Notable Achievements in Intelligent Agent Applications AI Empowers Steady Growth in Core Businesses" [腾讯公布二零二六年第一季业绩 混元迭代性能跃升 智能体应用成果凸显AI赋能主业稳健增长], Tencent, 13 May 2026, <<https://web.archive.org/web/20260528192038/https://static.www.tencent.com/uploads/2026/05/13/aaeed028d215a9612da4b41aca17708b.pdf>> [accessed 5 June 2026].
- 56 Daniel Fu, "Political Drivers of China's Private Sector Demise", *Jamestown China Brief*, 23.16, 9 August 2023, <<https://jamestown.org/political-drivers-of-chinas-private-sector-demise/>> [accessed 5 June 2026].
- 57 Bien Perez, "Now it's our turn, Not the Government's: Alibaba's Jack Ma Says This is One Chinese Economic Slowdown with a Silver Lining", *SCMP*, 25 September 2025, <<https://www.scmp.com/tech/enterprises/article/1861409/now-its-our-turn-not-governments-alibaba-jack-ma-says-one-chinese>> [accessed 5 June 2026].
- 58 "Outspoken Billionaire Sun Dawu Jailed for 18 years in China", *BBC News*, 12 July 2021, <<https://www.bbc.com/news/world-asia-china-58007515>> [accessed 5 June 2026]; Chris Buckley, "This Chinese Businesswoman Was a Model of Success. Then She Angered the Party", *The New York Times*, 8 February 2021, <<https://www.nytimes.com/2021/02/08/world/asia/china-geng-xiaonan-trial.html>> [accessed 5 June 2026]; Chris Buckley, "China Sentences Tycoon Who Criticized Xi to 18 Years in Prison", *The New York Times*, 22 September 2020, <<https://www.nytimes.com/2020/09/22/world/asia/china-ren-zhiqiang-tycoon.html>> [accessed 5 June 2026].
- 59 State Administration for Market Regulation, "The State Administration for Market Regulation has imposed administrative penalties on Alibaba Group Holding Limited for engaging in 'forced exclusivity' practices in the online retail platform services market within China, in accordance with the law." [市场监管总局依法对阿里巴巴集团控股有限公司在中国境内网络零售平台服务市场实施"二选一"垄断行为作出行政处罚], 10 April 2021, <https://web.archive.org/web/20250430005311/https://www.samr.gov.cn/zt/qhldzf/art/2021/art_a10f74fa09cd49ee8db7804ba834db2a.html> [accessed 5 June 2026]; Cyberspace Administration of China, "Announcement by the Cybersecurity Review Office on the Initiation of a Cybersecurity Review of 'Didi Chuxing'", 2 July 2021, <https://web.archive.org/web/20260522050131/https://www.cac.gov.cn/2021-07/02/c_1626811521011934.htm> [accessed 5 June 2026]; State Administration for Market Regulation, "The State Administration for Market Regulation has, in accordance with the law, imposed penalties on Tencent Holdings Limited, including ordering the company to terminate exclusive online music licensing agreements" [市场监管总局依法对腾讯控股有限公司作出责令解除网络音乐独家版权等处罚], 24 July 2021, <https://web.archive.org/web/20260117131447/https://www.samr.gov.cn/xw/zj/art/2023/art_7fe51000dd5c47d993d775b504621805.html> [accessed 5 June 2026].
- 60 Government of the People's Republic of China, "Xi Jinping chaired the ninth meeting of the Central Financial and Economic Affairs Commission" [习近平主持召开中央财经委员会第九次会议], 15 March 2021, <https://web.archive.org/web/20260515105631/https://www.gov.cn/xinwen/2021-03/15/content_5593154.htm> [accessed 5 June 2026].
- 61 Government of the People's Republic of China, "The General Office of the Central Committee of the Communist Party of China has issued the 'Opinions on Strengthening United Front Work with the Private Sector in the New Era'" [中共中央办公厅印发《关于加强新时代民营经济统战工作的意见》], 15 September 2020, <<https://archive.ph/6LN93>> [accessed 5 June 2026].
- 62 Yang Yingzhi and Brenda Goh, "Beijing Took Stake and Board Seat in Key ByteDance Domestic Entity This Year", *Reuters*, 17 August 2021, <<https://www.reuters.com/world/china/beijing-owns-stakes-bytedance-weibo-domestic-entities-records-show-2021-08-17/>> [accessed 5 June 2026].
- 63 Lizzi C. Lee and Jing Qian, "China's Private Sector Pivot: How Beijing Is Encouraging Entrepreneurs Without Giving Up Control", *Foreign Affairs*, 20 February 2026, <<https://www.foreignaffairs.com/china/chinas-private-sector-pivot>> [accessed 5 June 2026].
- 64 Government of the People's Republic of China, "The General Office of the Central Committee of the Communist Party of China has issued the 'Opinions on Strengthening United Front Work with the Private Sector in the New Era'".
- 65 Zhang Qiaosu, "Xi Jinping: The private sector has a bright future and great potential; now is the time for private enterprises and entrepreneurs to make their mark" [习近平：民营经济发展前景广阔大有可为 民营企业和民营企业家大显身手正当其时], *Xinhua*, 17 February 2025, <<https://web.archive.org/web/20250824034031/http://www.news.cn/politics/leaders/20250217/8bccf23cac99480bbb1f4b23cb209824/c.html>> [accessed 5 June 2026]; "Analysis: Xi Jinping meets with private sector representatives for the first time in over six years – what else is in the spotlight besides Jack Ma's 'clearance'?" [观察：习近平时隔六年多再晤民企，除了马云“过关”还有哪些关注点], *BBC News*, 18 February 2025, <<https://www.bbc.com/zhongwen/articles/cj3e3l3r2exo/simp>> [accessed 5 June 2026].
- 66 "Analysis: Xi Jinping meets with private sector representatives for the first time in over six years – what else is in the spotlight besides Jack Ma's 'clearance'?"
- 67 Zhang Qiaosu, "Xi Jinping: The private sector has a bright future and great potential; now is the time for private enterprises and entrepreneurs to make their mark".
- 68 National People's Congress, "Law of the People's Republic of China on the Promotion of the Private Economy" [中华人民共和国民营经济促进法], Ministry of Justice of the People's Republic of China, 30 April 2025, <https://www.moj.gov.cn/pub/sfbgw/gwxw/xwyw/202504/t20250430_518418.html> [accessed 5 June 2026].
- 69 Christopher Nye and Charles Sun, "PRC Tech Firms Practice Proactive Alignment", *Jamestown China Brief*, 20 March 2026, <<https://jamestown.org/prc-tech-firms-practice-proactive-alignment/>> [accessed 5 June 2026].
- 70 Government of the People's Republic of China, "National Standardisation Development Outline", 10 October 2021, <https://web.archive.org/web/20260520174239/https://www.gov.cn/zhengce/2021-10/10/content_5641727.htm> [accessed 5 June 2026].
- 71 National People's Congress, "The 15th Five-Year Plan for National Economic and Social Development of the People's Republic of China" [中华人民共和国国民经济和社会发展第十五个五年规划纲要], 12 March 2026, <<https://web.archive.org/web/20260525131437/https://www.ndrc.gov.cn/fggz/fzlggh/gjfgzh/202603/U020260317369114704096.pdf>> [accessed 5 June 2026].
- 72 National People's Congress, "Law of the People's Republic of China on the Promotion of the Private Economy".

- 73 Ali Cloud, “Top-tier! Tongyi Qianwen has obtained large-model security certification” [最高等级！通义千问获得大模型安全认证], 6 June 2025, <<https://archive.ph/wip/myOJm>> [accessed 5 June 2026].
- 74 Ali Cloud, “First of their kind! Tongyi Qianwen and Tongyi Wanshang have passed the national safety standard compliance verification for generative AI” [首批！通义千问、通义万相通过生成式AI安全国标符合性验证], 16 June 2025, <<https://archive.ph/wip/ekeYl>> [accessed 5 June 2026].
- 75 Zhang Xiaobo and Fang Tong, “Hebei Province has released 35 exemplary cases of ‘AI+’” [河北省發布35個“人工智能+”典型案例], *People’s Daily*, 17 March 2022, <<https://archive.ph/wip/rL2tD>> [accessed 5 June 2026].
- 76 “To enhance DeepSeek’s security, a joint team from Tsinghua Ruilai has released the RealSafe-R1 large language model” [增强DeepSeek安全性，清华瑞莱联合团队发布RealSafe-R1大模型], BSIA, 25 February 2025, <<https://archive.ph/wip/ZKEYx>> [accessed 5 June 2026].
- 77 “CAISI Evaluation of DeepSeek AI Models Finds Shortcomings and Risks”, Center for AI Standards and Innovation, 16 September 2025, <https://www.nist.gov/system/files/documents/2025/09/30/CAISI_Evaluation_of_DeepSeek_AI_Models.pdf> [accessed 5 June 2026].
- 78 Center for AI Standards and Innovation, “Evaluation of DeepSeek AI Models”, National Institute of Standards and Technology, 30 September 2025, <https://www.nist.gov/system/files/documents/2025/09/30/CAISI_Evaluation_of_DeepSeek_AI_Models.pdf> [accessed 29 June 2026].
- 79 Alex Colville, “Deepseeking Truth”, China Media Project, 10 February 2025, <<https://chinamediaproject.org/2025/02/10/deepseeking-truth/>> [accessed 5 June 2026].
- 80 Jasmine Cui and Jared Perlo, “More of Silicon Valley is Building on Free Chinese AI”, *NBC News*, 30 November 2025, <<https://www.nbcnews.com/tech/innovation/silicon-valley-building-free-chinese-ai-rcna242430>> [accessed 5 June 2026].
- 81 Nathan Lambert, “The ATOM Project: Towards Fully Open Models for US Research & Industry”, 4 August 2025, <<https://atomproject.ai/>> [accessed 5 June 2026].
- 82 “China is Quietly Upstaging America with its Open Models”, *The Economist*, 21 August 2025, <<https://www.economist.com/business/2025/08/21/china-is-quietly-upstaging-america-with-its-open-models>> [accessed 5 June 2026].
- 83 Samantha Subin, “Airbnb CEO Chesky Says ChatGPT isn’t ‘Quite Robust Enough’ to Integrate into Travel App”, *CNBC*, 22 October 2025, <<https://www.cnn.com/2025/10/22/airbnb-chatgpt-ai-chesky.html>> [accessed 5 June 2026].
- 84 “Qwen3 and DeepSeek-V3.1 Models Now Available Fully Managed on AWS”, *Amazon News*, 18 September 2025, <<https://www.aboutamazon.com/news/aws/alibaba-qwen3-deepseek-v3-amazon-bedrock>> [accessed 5 June 2026]; “Foundry Models”, Microsoft Azure, <<https://azure.microsoft.com/en-us/products/ai-foundry/models>> [accessed 5 June 2026]; “DeepSeek Models”, Google Cloud, 3 June 2026, <<https://docs.cloud.google.com/vertex-ai/generative-ai/docs/maas/deepseek>> [accessed 5 June 2026].
- 85 “AI Model Rankings”, OpenRouter, <<https://openrouter.ai/rankings?view=week>> [accessed 5 June 2026].
- 86 Nathan Lambert, “The ATOM Project: Towards Fully Open Models for US Research & Industry”; Anthony Ha, “Cursor Admits its New Coding Model was Built on Top of Moonshot AI’s Kimi”, 22 March 2026, <<https://techcrunch.com/2026/03/22/cursor-admits-its-new-coding-model-was-built-on-top-of-moonshot-ais-kimi/>> [accessed 5 June 2026].
- 87 Estonian Foreign Intelligence Service, *International Security and Estonia 2026*, 10 February 2026, <<https://www.valisluureamet.ee/doc/report/2026-en.pdf>> [accessed 5 June 2026].
- 88 Alex Colville, “The Chinese Core of ‘Uganda’s ChatGPT’”, China Media Project, 17 December 2025, <<https://chinamediaproject.org/2025/12/17/the-chinese-core-of-ugandas-chatgpt/>> [accessed 5 June 2026].
- 89 Huang Peihsuan and others, *Analysis of LLM Bias (Chinese Propaganda & Anti-US Sentiment) in DeepSeek-R1 vs. ChatGPT o3-mini-high*, June 2025, DOI: [10.48550/arXiv.2506.01814](https://doi.org/10.48550/arXiv.2506.01814) [accessed 5 June 2026].
- 90 Alex Colville, “Rehabilitating DeepSeek”, China Media Project, 4 March 2025, <<https://chinamediaproject.org/2025/03/04/rehabilitating-deepseek/>> [accessed 5 June 2026].
- 91 Evan Hubinger and others, *Sleeper Agents: Training Deceptive LLMs that Persist Through Safety Training*, arXiv, 17 January 2024, <<https://doi.org/10.48550/arXiv.2401.05566>> [accessed 5 June 2026].
- 92 Audrey Tang, “Invite the Frontier Model onto Your Macbook”, 2026, <<https://pi.audreyt.org/>> [accessed 5 June 2026].
- 93 Oliver Dowden, “UK Telecommunications. Volume 678: Debated on Tuesday 14 July 2020”, UK Parliament, 14 July 2020, <<https://hansard.parliament.uk/commons/2020-07-14/debates/9467A1F1-B87B-4E16-AD0C-DFF02A51981A/UKTelecommunications>> [accessed 5 June 2026].
- 94 Secretary of State for Science, Innovation and Technology of the United Kingdom, *AI Opportunities Action Plan*, UK Government, 13 January 2025, <<https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan>> [accessed 5 June 2026]; European Commission, “Shaping Europe’s Leadership in Artificial Intelligence with the AI Continent Action Plan”, 10 April 2026, <https://commission.europa.eu/topics/competitiveness/ai-continent_en> [accessed 5 June 2026].
- 95 Government of the People’s Republic of China, “National Standardisation Development Outline” [《国家标准化发展纲要》], 10 October 2021, <https://web.archive.org/web/20260430035006/https://www.gov.cn/zhengce/2021-10/10/content_5641727.htm> [accessed 5 June 2026].

ANNEX

- 96 First Financial, “Alibaba’s New-Generation Training-Inference Integrated AI Chip Makes First Appearance” [阿里新一代训推一体AI芯片首次亮相], Securities Times, 20 May 2026, <<https://web.archive.org/web/20260630093757/https://www.stcn.com/article/detail/3917836.html>> [accessed 1 July 2026]; National Radio and Television Administration Monitoring Centre, “Award Announcement for the Procurement Project for Private-Network-Side Software for the Radio and Television Industry Cybersecurity Protection Management System” [国家广播电视总局监管中心广播电视行业网络安全保护管理系统专网侧软件采购项目中标公告], China Government Procurement Network, 16 December 2024, <https://web.archive.org/web/20260403134121/https://www.ccgp.gov.cn/cggg/zygg/zbagg/202412/t20241216_23883714.htm> [accessed 1 July 2026].
- 97 Huawei Developer Forum, “Hardcore Innovation and Full-Domain Collaboration: The Evolution of Huawei’s Technology System and Its Industrial Value” [硬核创新与全域协同：华为技术体系演进与产业价值], Huawei Developer Forum, <<https://archive.ph/Le0qW>> [accessed 1 July 2026].
- 98 Xinhua, “DeepSeek, a ‘Breath of Fresh Air’ Stirring the AI Industry” [DeepSeek，激荡AI行业的“一股清流”], Xinhua, 21 February 2025, <<https://web.archive.org/web/20260630094248/https://www.news.cn/tech/20250221/6269c30fe30c4960a5450bdbc1b65a5e/c.html>> [accessed 1 July 2026].
- 99 Baidu AI Cloud, “Baidu AI Foundation” [百度AI大底座], Baidu AI Cloud, <<https://web.archive.org/web/20260410231956/https://cloud.baidu.com/solution/aif.html>> [accessed 1 July 2026].
- 100 Xiaomi, “Human x Car x Home” [人车家全生态], Xiaomi, <<https://web.archive.org/web/20251211140451/https://www.mi.com/human-car-home>> [accessed 1 July 2026]; Xiaomi, “Xiaomi Unveils Xiaomi XRING Chips, Xiaomi 15S Pro, Xiaomi Pad 7 Ultra and Multiple AIoT Products in Beijing Launch Event”, Xiaomi, 22 May 2025, <<https://archive.ph/MFlnq>> [accessed 1 July 2026].
- 101 RealAI, “Solutions” [解决方案], RealAI, <<https://archive.ph/OA6xO>> [accessed 1 July 2026].
- 102 Antom, “Antom Copilot 2.0 Helps Ant International Expand and Optimize Merchant Payment AI Functions” [Antom Copilot 2.0 助力蚂蚁国际拓展优化商户支付AI], Antom, 2025, <<https://archive.ph/VJlW>> [accessed 1 July 2026].

