AI technological innovation has entered a period of unprecedented dynamism, profoundly reshaping all aspects of economic and social life. While technological advances have created historic opportunities for development and generated substantial benefits, they have also brought a range of risks and challenges. How to ensure effective AI governance and maintain a dynamic balance between promoting innovation and mitigating risks has become a common challenge facing the international community. China is a global innovation powerhouse and a leading country in AI technology and applications. In recent years, guided by the values of people-centrality and AI for good, China has placed equal emphasis on AI development and AI safety. Keeping abreast of the trends and laws of AI development, the country has continuously built up its relevant laws and regulations, policy frameworks, standards and ethical guidelines to steadily advance AI governance through concrete actions. As a responsible major country, China has also put forward a series of initiatives and plans for global AI governance and established platforms to promote international cooperation in this field.

To promote international exchanges on AI governance rules, the China Institute of International Studies (CIIS) and the China Academy of Information and Communications Technology (CAICT) organized a dedicated team to review the underlying values, institutional framework, key priorities, and practical experiences in Chinese AI governance and analyze their global implications. The findings are presented in this research report.



CIIS Report

AI Governance in China: Philosophy, Practices, and Implications

China Institute of International Studies
China Academy of Information and Communications Technology

October 2026

Contents

I Preface 2

II AI Governance: Global Ecosystem and China's Practices 6

(I) The Global Ecosystem of AI Governance 6

(II) China's Philosophy and Practices in AI Governance 11

III AI Governance in China: Key Directions 24

(I) Advancing the Implementation of AI Ethical Principles 25

(II) Strengthening Model Governance to Ensure Safety and Controllability 28

(III) Promoting Data Governance to Support Technological Advances 32

(IV) Advancing the Governance of AI-generated Content 35

(V) Implementing Scenario-based Governance to Promote AI Application 39

(VI) Exploring Multi-stakeholder Governance to Improve Governance Effectiveness 42

IV AI Governance in China: Global Implications 50

(I) Demonstrating Certainty by Exploring and Accumulating Experience in AI Governance 51

(II) Upholding Openness and Promoting International Cooperation on AI Governance 55

AI Governance in China: Philosophy, Practices, and Implications

I

Preface

I Preface

In September 2025, President Xi Jinping put forward the Global Governance Initiative (GGI) at the "Shanghai Cooperation Organization (SCO) Plus" meeting. He emphasized the need to adhere to sovereign equality, abide by international rule of law, practice multilateralism, advocate the people-centered approach, and focus on taking real actions, urging all parties to work together for a more just and equitable global governance system and a community with a shared future for humanity. Enhancing AI governance is an important component of the efforts to reform and improve global governance. The *Concept Paper on the Global Governance Initiative* states that priority will be given to areas with acute governance urgency and substantial governance deficits, such as AI and cyberspace, and that countries should increase communication and cooperation in these areas to build consensus, identify deliverables, and bring about early harvests.

China upholds the people-centered approach and the vision of AI for good in AI governance, placing equal emphasis on AI development and safety and steadily advancing AI governance at home. Building on this foundation, China actively promotes international cooperation on AI governance in an open spirit. It has successively issued a number of policy documents on international cooperation on AI governance, established the World Data Organization, hosted the World Artificial Intelligence Conference (WAIC), and prepared for the establishment of the World Artificial Intelligence Cooperation Organization. Leveraging

the United Nations as the primary channel, China has also prioritized helping Global South countries strengthen their capacities and bridging the AI divide, and called for the early establishment of a global governance framework underpinned by broad consensus. China has taken solid steps to advance global AI governance, contributing new ideas and practical experience to the reform and improvement of global governance.

AI Governance in China: Philosophy, Practices, and Implications

II

AI Governance: Global Ecosystem and China's Practices

- The Global Ecosystem of AI Governance
- China's Philosophy and Practices in AI Governance

II AI Governance: Global Ecosystem and China's Practices

In recent years, AI technologies have evolved at a rapid pace and begun to see large-scale deployment. In this process, AI governance has evolved from a technological issue to a major policy issue drawing global attention. A global ecosystem of AI governance has gradually taken shape characterized by multiple stakeholders, continuous evolution, and a mix of competition and cooperation. China is a major country in AI technology and applications. Since it released the *New Generation AI Development Plan* in 2017, the country, through years of continuous exploration and improvement, has blazed a distinctive path of AI governance whereby the once scattered governance measures are replaced by a systematic approach and reactive responses by proactive institution-building.

(I) The Global Ecosystem of AI Governance

1. Growing international attention to AI governance

In 2017, the inaugural AI for Good Global Summit was held in Geneva. That same year, United Nations Secretary-General António Guterres incorporated AI governance into his report to the UN General Assembly. Since then, international attention to and investment in AI governance have continued to grow, and a wide range of governance stakeholders, mechanisms, initiatives, and actions have emerged.

At the United Nations, the most broadly representative international platform,

Secretary-General António Guterres has attached great importance to AI governance. In October 2023, the Highlevel Advisory Body on Artificial Intelligence was convened, which released the *Governing AI for Humanity: Final Report* in September 2024. In 2024, the United Nations convened the Summit of the Future, which adopted the *Pact for the Future* and its annexes—the *Global Digital Compact* and the *Declaration on Future Generations*. The *Global Digital Compact* represents the world's first-ever comprehensive framework addressing digital cooperation and AI governance. In August 2025, the UN General Assembly adopted a resolution to formally establish two mechanisms: the Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on AI Governance. On July 1, 2026, the Scientific Panel released its first report, *Preliminary Report: Evidence-Based Assessment of Opportunities, Risks and Impacts of Artificial Intelligence*. On July 6–7, the inaugural Global Dialogue on AI Governance was held in Geneva, Switzerland, bringing together representatives from governments, the business sector, academia, and technical communities from multiple countries. Discussions focused on topics including the opportunities and impacts of AI, bridging the AI divide, interoperability and compatibility among different AI governance pathways, transparency, accountability, and human oversight.

In addition to the aforementioned mechanisms led by the UN Secretary-General, various specialized agencies within the UN system have also been actively exploring AI governance within their respective domains. The United Nations Educational, Scientific and Cultural Organization (UNESCO) has issued a number of governance-related documents, including the *Recommendation on the Ethics of Artificial Intelligence* and the *Guidance for Generative AI in Education and Research*, and has established the Global AI Ethics and Governance Observatory. The International Telecommunication Union (ITU) convenes the AI for Good Global Summit annually and seeks to take a leading role, together with the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), in developing standards for AI. The World Intellectual Property Organization (WIPO) has launched the "WIPO Conversation on Intellectual Property and Frontier Technologies" and issued practical governance toolkits. The International Labour Organization (ILO) has launched initiatives such as the Observatory on AI and Work in the Digital Economy.

Other multilateral mechanisms have also been actively advancing AI governance. In

2019, the OECD Council adopted the *Recommendation on Artificial Intelligence*, setting out five AI principles and five recommendations; the instrument was subsequently updated in 2024. The OECD has since introduced a range of tools and initiatives, including the OECD.AI Policy Observatory, a reliable framework of AI tools, and the Common Reporting Framework for AI Incidents. In 2019, the G20 Summit adopted the G20 AI Principles. Since then, the summit's outcome documents have addressed AI governance on an annual basis. At the summits hosted by Brazil and South Africa over the past two years, AI governance was identified as a priority and discussed alongside sustainable development and digital equity. In 2025, the BRICS Summit adopted the *BRICS Leaders' Statement on the Global Governance of Artificial Intelligence*, which sets out a series of guiding principles on AI development and application, including: (1) multilateralism, legitimacy, and digital sovereignty; (2) market regulation, data governance, and access to technology; and (3) equity and sustainable development.

The EU and ASEAN are two regional organizations that have made significant progress in AI governance. In 2024, the EU adopted the *AI Act*, the world's first comprehensive AI legislation, establishing a regulatory framework centered on risk classification. In the same year, ASEAN issued the *ASEAN Guide on AI Governance and Ethics* as a regional guideline, followed in 2025 by an expanded version incorporating provisions concerning generative AI.

The series of AI governance summits initiated by the UK represents another major multilateral effort in recent years. In 2023, the UK hosted the inaugural AI Safety Summit at Bletchley Park, focusing on the "serious, even catastrophic" harm posed by frontier AI. The second summit, held in Seoul, ROK, in 2024, focused on improving the safety, innovation, and inclusivity of AI, and a collaborative network of AI safety institutes was formed subsequently. In 2025, the AI Action Summit was held in Paris, France, emphasizing innovation and the economic opportunities presented by AI while also addressing broader AI governance topics such as AI's impact on the environment and the workforce. The 2026 AI Impact Summit in India focused primarily on how the Global South can benefit from advances in AI technology.

Regarding international law, the *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* is the world's first legally binding international treaty on AI governance. It was opened for

signature globally in September 2024, and had been signed by 19 countries as of April 2026. The Convention establishes fundamental principles for AI governance, including transparency, accountability, non-discrimination, and the protection of human rights.

2. Multiple challenges facing AI governance

AI governance refers to the process of regulating AI activities across its entire lifecycle—including design, development, deployment, and use—through a system of rules. Compared with other global issues such as climate change, nuclear security and non-proliferation, and public health, AI technologies are more disruptive and have broader political, economic, and social implications, making AI governance more complex.

First, AI governance involves a wide range of areas, making it difficult to establish a comprehensive and coordinated governance system. At the macro level, AI governance can broadly be divided into two dimensions: safety and development. The former focuses on assessing and preventing the various risks that AI technologies may pose to ensure that AI remains safe, reliable, and under human control. The latter focuses on creating favorable conditions for AI innovation and application while ensuring that AI technologies are fair, accessible, and inclusive. At the micro level, AI governance encompasses a broad range of issues, including ethical alignment, privacy and data governance, algorithmic transparency, algorithmic fairness, and algorithmic manipulation, and the definition and division of responsibilities. The application of AI technologies in different contexts gives rise to additional and highly varied challenges. All these factors make coordinated AI governance more difficult.

Second, as AI technologies continue to advance rapidly, governance measures may have to play "catch-up" with technological advances for a long time to come. AI innovation continues to accelerate, and new algorithmic architectures, training paradigms, and applications keep emerging, expanding the boundaries of AI capabilities, whereas the potential risks are difficult to anticipate or define clearly. This differs significantly from traditional approaches to technology governance, which usually rely on relatively static standards. Developing dynamic governance mechanisms that can keep pace with the speed of AI evolution has therefore become an urgent imperative.

Third, geopolitical factors have a profound impact on AI governance. AI is regarded

as a strategic technology that will affect countries' future comprehensive strength and international competitiveness. Governments are generally concerned that unilateral domestic regulation could undermine their national competitiveness, and their domestic AI governance constantly sways between "ensuring safety" and "promoting development." At the same time, multilateral AI governance has not seen much substantive progress due to insufficient trust among countries.

Against this backdrop, global AI governance is currently facing significant challenges. The EU, for example, adopted the *AI Act*, the world's first comprehensive legislation in AI, yet its implementation has encountered a range of difficulties from the very beginning, including gaps in technical standards, the absence of designated regulatory authorities, insufficient compliance tools and guidance, and high compliance costs for businesses. The EU subsequently revised the legislation. In May 2026, its member states and the European Parliament reached a provisional agreement to ease the existing regulatory framework to some extent and postpone the implementation of some key rules.

In the United States, the Trump administration has formulated its AI policies with the goal of winning the AI race, prioritizing the removal of regulatory barriers and the acceleration of innovation. At the beginning of President Trump's second term, the administration revoked the Biden administration's *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, which had placed greater emphasis on risks and rights. At the end of 2025 and in June 2026, respectively, the Trump administration issued two executive orders: *Ensuring a National Policy Framework for Artificial Intelligence* and *Promoting Advanced AI Innovation and Security*. The former seeks to curb the rapid expansion of state-level AI regulations with a minimally burdensome national policy framework, while the latter encourages voluntary cooperation between AI companies and the government. At present, the United States has no binding federal AI regulatory regime, but relies on AI companies' voluntary assessments and self-regulation. This "light-touch" regulatory approach has had negative implications for global AI governance.

From a multilateral perspective, global AI governance faces the risk of fragmentation. The international community does not yet have a single authoritative framework guiding AI governance. Numerous mechanisms operate in parallel, with differing

philosophies, scope of governance, definitions of risk, and technical standards, while effective coordination and alignment are limited. Meanwhile, multilateral AI governance lacks binding force and execution. At present, most international declarations, initiatives, and recommendations concerning AI governance remain at the stage of articulating principled consensus and are voluntary in nature. They fall in the category of "soft law," the implementation of which is at the signatory countries' discretion, with no guaranteed supervision or enforcement. The *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* is the only binding international "hard law" now in this area, but the conditions for its entry into force have not yet been met. In addition, the U.S. government's explicit opposition to multilateral AI governance has cast a further shadow over this endeavor on the global level. At the 2026 AI Impact Summit in India, the U.S. representative stated explicitly that the U.S. "completely rejects" global AI governance.

(II) China's Philosophy and Practices in AI Governance

China attaches great importance to the governance of AI, and has gradually blazed a distinctive path of AI governance through years of exploration and practices.

1. Developing a comprehensive AI governance philosophy covering values, principles, and approaches

China upholds the people-centered approach and the vision of AI for good in AI governance. It regards improving people's well-being and serving economic and social development as the fundamental objectives of technological advancement, emphasizing that technological development should always serve the public interest and be guided by sound social values and ethics. China maintains that technological development must be aligned with social equity, public order, and the all-round development of individuals, thereby promoting overall social well-being. In exercising AI governance, China has strengthened oversight of risks in key areas such as algorithmic recommendation, deep synthesis, and generative AI, with sustained efforts to address issues including misinformation, cyber violence, vulgar content, and deepfakes. It has also strengthened protection for minors, personal privacy, and intellectual property rights (IPR) to prevent AI from becoming a tool for promoting wrong values or

facilitating malicious or abusive uses. China is pushing the accelerated application of AI in areas such as healthcare, education, rural revitalization, and public services, with the aim of ensuring that the benefits of AI technologies reach ordinary people more quickly and the AI divide is narrowed between urban and rural areas and across regions. In this way, AI can truly help improve people's livelihoods and serve as an important driver of common prosperity.

China places equal emphasis on AI development and AI safety. It takes an integrated approach to technological innovation and risk prevention and control, encouraging the use of AI to empower high-quality economic and social development while establishing robust safety baselines. On the innovation side, China improves computing infrastructure, develops national AI innovation platforms, and promotes in-depth integration of AI and the real economy by developing medium- and long-term national development plans. It also streamlines compliance procedures to create a favorable environment for business innovation. On the risk prevention and control side, China adopts a systematic and targeted approach, establishing a multi-layered risk prevention and control system to effectively protect citizens' rights and interests, data security, social equity, and public order and morality. Classified guidance and differentiated management are adopted based on factors such as the level of technological risk, differences in application scenarios, and the degree of social impact. High-risk AI technologies and applications are subject to review and filing requirements, with enterprises bearing primary responsibility. Targeted regulatory campaigns have been conducted in high-frequency risk areas such as deepfakes, algorithmic discrimination, and the dissemination of misinformation. China has strengthened education and public awareness on AI ethics, and established AI safety monitoring platforms to enable real-time monitoring, early warning, and response to technological risks. These efforts have promoted coordinated AI governance involving the government, enterprises, academia, and society.

China adopts an agile approach to AI governance. In response to the rapid evolution of AI technologies, the continued expansion of application scenarios, and the dynamic changes in associated risks, China dynamically calibrates the boundaries of AI governance following the principle of "small steps and rapid iteration." The *AI Safety Governance Framework* provides a prime example. In September 2024, the National Technical Committee 260 on Cybersecurity of Standardization Administration of

China released Version 1.0 of the framework. One year later, it released Version 2.0 in response to new developments such as major advances in the capabilities of AI reasoning models and the significant lowering of application thresholds driven by lightweight open-source models. The updated framework expands the scope of safety risks from two broad categories to three, upgrades the governance approach from simply listing risks to providing a systematic methodology for safety governance, introduces new tools such as "circuit-breaking mechanisms," and further strengthens alignment with international practices.[①] In September 2026, the National Technical Committee 260 on Cybersecurity released Version 3.0 of the framework. The new version identifies AI agents and embodied AI as distinct categories of significant risk and brings emerging risks—including defects in synthetic data, memory poisoning, poisoning of generative engine optimization (GEO), and autonomous cyberattacks—within the scope of governance. It also strengthens security controls covering identity and access, tool invocation, runtime monitoring, and whole-process auditing. China has introduced the regulatory sandbox mechanism in a number of cities, whereby when enterprises conduct practical testing in controlled environments, regulators can monitor risks and accumulate experience in parallel while using the results to dynamically adjust regulatory rules. By establishing a governance loop of "sensing–decision-making–response–feedback," the agile approach can maximize the adaptability and responsiveness of governance during periods of rapid technological advancement. It illustrates the principle of "developing and governing simultaneously" and paves the way for developing a dynamically adaptable governance system covering the entire AI lifecycle.

2. Establishing a multilevel and multidimensional policy and regulatory system for AI governance

At the foundational level, three laws—the *Data Security Law of the People's Republic of China*, the *Cybersecurity Law of the People's Republic of China*, and the *Personal*

① "Expert Insight: From Risk Listing to Governance System: An Analysis of the Evolution of AI Safety Governance Framework," Cyberspace Administration of China, ttps://www.cac.gov.cn/2025-09/28/c_1760779758683488.htm.

Information Protection Law of the People's Republic of China—have established basic rules and binding requirements for the development and application of AI from the perspectives of cybersecurity, data classification and tiering, and the protection of personal information rights and interests. Together, they constitute the legal foundation for AI governance. In addition to national legislation, China has issued a range of policy documents, departmental regulations, industry standards, and technical guidance documents that provide policy direction and regulatory requirements for AI development and application. For example, regarding specific subjects such as algorithmic recommendations, deep synthesis, and generative AI services, China has issued the *Regulations on Algorithmic Recommendations in Internet-based Information Services*, the *Regulations on Deep Synthesis in Internet-based Information Services*, the *Interim Measures for the Administration of Generative Artificial Intelligence Services*, and the *Measures for Labeling AI-Generated and Synthesized Content*. Regarding AI ethics management and anthropomorphic interactive services, it has issued the *Measures for AI Ethics Review and Services (Trial)* and the *Interim Measures for the Administration of Anthropomorphic AI Interactive Services*. Drawing on its own experience in AI development and application while referring to mainstream international AI governance concepts and experience, China has issued two editions of the *AI Safety Governance Framework*. These documents are intended to help develop a safe, reliable, and controllable AI ecosystem and foster collaborative governance across national borders, domains, and industries.

List of AI-related Policy Documents, Laws, and Regulations in China

Category	Title	Issued by	Year
National laws	Cybersecurity Law	Standing Committee of the National People's Congress	2017
	Data Security Law	Standing Committee of the National People's Congress	2021
	Personal Information Protection Law	Standing Committee of the National People's Congress	2021

Category	Title	Issued by	Year
National policies and regulations	*New Generation AI Development Plan*	State Council	2017
	Ethical Norms for New Generation Artificial Intelligence	National New Generation AI Governance Specialist Committee	2021
	Regulations on Algorithmic Recommendations in Internet-based Information Services	Cyberspace Administration of China (CAC) and three other departments	2021
	Opinions on Strengthening the Governance over Ethics in Science and Technology	General Office of the State Council	2022
	Guiding Opinions on Accelerating Scenario Innovation to Promote High-quality Economic Development via High-level AI Applications	Ministry of Science and Technology and five other departments	2022
	Regulations on Deep Synthesis in Internet-based Information Services	CAC and two other departments	2022
	Interim Measures for the Administration of Generative AI Services	CAC and six other departments	2023
	Measures for the Ethical Review of Science and Technology (Trial)	Ministry of Science and Technology and eight other departments	2023
	AI Safety Governance Framework, Version 1.0	National Technical Committee 260 on Cybersecurity	2024
	Measures for Labeling AI-Generated and Synthesized Content	CAC and three other departments	2025
	Measures for the Management of Artificial Intelligence Science and Technology Ethics and Related Services (Trial)	Ministry of Industry and Information Technology (MIIT) and nine other departments	2025
	AI Safety Governance Framework, Version 2.0	National Technical Committee 260 on Cybersecurity	2025
	Measures for AI Ethics Review and Services (Trial)	MIIT and nine other departments	2026
	Interim Measures for the Administration of Anthropomorphic AI Interactive Services	CAC and four other departments	2026
	Implementation Opinions on the Standardized Application and Innovative Development of AI Agents	CAC and two other departments	2026
	AI Safety Governance Framework, Version 3.0	National Technical Committee 260 on Cybersecurity	2026

Category	Title	Issued by	Year
Local policies and regulations	Regulations of Shenzhen Special Economic Zone on Promoting the AI Industry	Standing Committee of Shenzhen Municipal People's Congress	2022
	Several Measures of Shanghai Municipality for Promoting the Innovative Development of AI Large Models	Shanghai Municipal Government	2023
	Shenzhen Action Plan for Accelerating the Building of an Artificial Intelligence Pioneer City (2025-2026)	Development and Reform Commission of Shenzhen Municipality	2025
	Action Plan of Beijing Municipality for Empowering New-type Industrialization with AI	Beijing Municipal Bureau of Economy and Information Technology	2025
	Regulations of Beijing Municipality on the Development of the AI Industry (Draft)	Beijing Municipal People's Congress	2026

Under the guidance of national laws and regulations, many local governments in China have introduced more detailed, supplementary local regulations tailored to their respective industrial and regulatory realities. Examples include the *Regulations of Shanghai Municipality on Promoting the Development of the Artificial Intelligence Sector*, the *Regulations of Shenzhen Special Economic Zone on Promoting the Development of the Artificial Intelligence Industry*, and the *Regulations of Hangzhou on Promoting the Embodied Intelligent Robotics Industry*. Provinces including Guangdong and Jiangsu have also incorporated regulations on promoting the AI industry into their legislative agendas.

While progressively refining the legislative framework, China has also continuously intensified law enforcement in the AI sector, establishing an interdepartmental coordination mechanism for joint AI safety enforcement to realize coordinated oversight across the full range of AI-related elements, including data, algorithms, computing power, and applications. At the same time, China has redoubled judicial efforts in this area to form a governance landscape in which administrative enforcement and oversight and judicial adjudication reinforce each other.

At the administrative enforcement level, China has established a full-chain enforcement system combining filing requirements, targeted regulatory campaigns, and routine

supervision. It has promoted the filing and safety assessment of generative AI algorithms to enhance technological transparency and protect users' rights and interests; it has reinforced the responsibility of online platforms for content management and carried out special campaigns to improve the AI industry ecosystem, punishing platforms that fail to fulfill their review obligations in accordance with the law. It has also combated malicious uses of AI, with a particular focus on addressing fabricated news, identity impersonation, and manipulation of online public opinion.

At the judicial level, the Supreme People's Court has issued judicial interpretations concerning issues such as facial recognition, platform monopoly, and the protection of the rights and interests of workers in new forms of employment, providing a basis for adjudicating disputes arising from algorithmic technologies. Courts across China have actively handled emerging types of cases involving copyright ownership of AI-generated works, infringement of personality rights, copyright in training data, and unfair competition, thereby promoting compliance within the industry. In addition, specialized judicial mechanisms such as internet courts have increasingly explored adjudication and rule-making in areas including AI-related copyright, data rights and interests, and platform responsibilities, contributing to the continued refinement of relevant governance rules.

Standards serve as the technical extension of laws and regulations while also supporting their implementation. China attaches great importance to AI standardization and is accelerating the development of a system of AI standards. In August 2020, the National Standardization Administration, together with the CAC, the Ministry of Science and Technology, and three other government departments, jointly issued the *Guidelines for the Construction of a National New Generation Artificial Intelligence Standards System*. The Guidelines establish a standards framework covering eight dimensions, including foundational common issues, key general-purpose technologies, and safety and ethics, aiming to provide strategic direction for the development of technical standards. Building on this document, the National Technical Committee 260 on Cybersecurity (TC260) has drafted multiple national standards, with a particular focus on areas such as machine learning and algorithmic ethics.[①] At the same time, the

① Jin Yulu, Hu Jiayu, and Xiao Youdan, "The Legislative Structure of China's Comprehensive AI Governance Mechanism," Bulletin of Chinese Academy of Sciences 41-0 (2026): 5.

National Standardization Administration, the MIIT, the CAC, and other government departments have worked together to advance AI standardization. Leveraging standardization bodies such as the AI Subcommittee of China National Information Technology Standardization Committee (SAC/TC 28/SC 42), they have coordinated the development of standards in key areas including foundational common issues, key technologies, products and applications, and safety and ethics. By the end of 2025, China had issued more than 30 national standards for AI. In 2026, it plans to develop more than 50 new national and industry standards, providing robust technical support for the effective implementation of relevant laws and regulations.

During the 15th Five-Year Plan period, China will continue to improve its legal and regulatory frameworks for AI and strengthen coordination across the entire governance chain—from legislation and law enforcement to judicial adjudication and law compliance—in order to provide robust rule-of-law safeguards for the healthy development of AI.

AI-related Technical Standards and Documents in China

Title	Issued by	Year
National Guidelines on the Development of a New Generation AI Standards System	National Standardization Administration and four other departments	2020
National Guidelines on the Development of a Comprehensive Standardization System for the AI Industry (2024)	MIIT and three other departments	2024
More than 30 national standards for AI	National Standardization Administration and other departments	By the end of 2025
Artificial Intelligence — Agent Interconnection series of national standards	MIIT and other departments	2026
General Security Requirements for AI Agent Applications	National Standardization Administration	2026

3. Building an AI governance landscape of multi-stakeholder collaboration

Considering the development patterns of AI technologies and the practical needs of AI governance, China has gradually established a collaborative AI governance framework in which the government plays a leading role, enterprises bear primary responsibilities, research institutions provide support, and social actors exercise oversight.

The government is responsible for strengthening top-level design and institutional arrangements. This includes issuing ethical guidelines, laws and regulations, and regulatory frameworks, setting the course for governance, defining security boundaries, and coordinating the development of the governance system. At the central government level, authorities including the National Development and Reform Commission (NDRC), the CAC, the Ministry of Science and Technology, the MIIT, the Ministry of Public Security, and the National Data Administration each perform their respective duties within the governance system while also working in coordination through policy consultations, joint policy issuance, joint enforcement, and targeted regulatory campaigns. At the same time, central and local governments work closely together. The central government provides policy guidance and institutional foundations for local AI governance, while local authorities, based on their regional industrial conditions and application scenarios, develop detailed implementation rules and take specific actions, including algorithm filing, pilot application in specific scenarios, risk screening, and sectoral regulation. They also provide timely feedback on emerging issues, risks, and regulatory needs, helping to dynamically refine the governance rules. This gives rise to a governance structure that combines central-local coordination and unified deployment with differentiated exploration.

AI enterprises are required to bear primary responsibilities for AI governance by strengthening full-process safety and compliance, strictly meeting requirements on algorithm filing, data security management, and self-assessment of ethical risks, and establishing robust internal governance structures. They also observe the bottom line relating to national security, data security, and personal information protection, and proactively accept government regulation, industry discipline, and public supervision. Additionally, enterprises can form industry associations, develop industry standards, conduct self-regulation, and set up exchange platforms to help guide orderly AI development.

Research institutes continue to enhance their capacity to provide theoretical support, technological assessment, and regulatory input. Institutes such as the Chinese Academy of Sciences and the China Academy of Information and Communications Technology (CAICT) conduct studies on subjects such as technology assessment, AI ethics, and risk identification, providing professional support for AIgovernance decisionmaking. University think tanks and industry research institutes also contribute through research reports, forums and seminars, and policy discussions, offering recommendations on AI

governance and providing reference for policy-making.

The public and the media are important forces in China's multi-stakeholder AI governance. Citizens conduct routine supervision of AI applications by providing feedback, reporting through corresponding platforms, or through other channels. The media exposes misconduct in AI governance in a timely manner, disseminates information on AI safety, and encourages public discussions. Together, they help forge social consensus, enhance governance transparency, and create effective social constraints.

AI Governance in China: Philosophy, Practices, and Implications

III

AI Governance in China: Key Directions

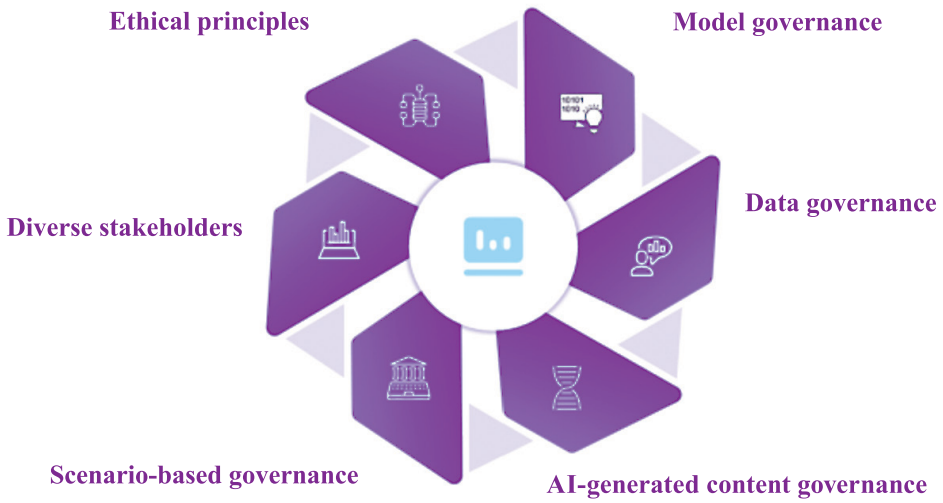
- Advancing the Implementation of AI Ethical Principles
- Strengthening Model Governance to Ensure Safety and Controllability
- Promoting Data Governance to Support Technological Advances
- Advancing the Governance of AI-generated Content
- Implementing Scenario-based Governance to Promote AI Application
- Exploring Multi-stakeholder Governance to Improve Governance Effectiveness

III AI Governance in China: Key Directions

China's AI governance efforts aim to ensure that AI develops in ways that are beneficial, safe, and fair. Giving equal weight to AI development and safety, China follows the principle of prioritizing areas of pressing need and applying targeted measures, focusing on exploring practical solutions to key issues. By continuously improving relevant laws and regulations, application standards, and regulatory measures, it seeks to ensure safe, reliable, and controllable AI development and make AI an important driver of both development and security.

China places ethics at the forefront of AI governance, establishing clear value boundaries for technological development and embedding ethical principles throughout the entire lifecycle of AI development and application. In terms of model governance, it addresses risks at their source through institutional measures such as model filing and safety assessments as well as the development of safety governance frameworks. In terms of data governance, it strengthens data protection, urges the efficient and convenient cross-border flow of data, and develops high-quality datasets to prevent data-related risks from spreading to model training and application. In terms of the governance of AI-generated content, China focuses on the social impact of model outputs, with particular attention to information integrity, the protection of workers' and consumers' rights and interests, and the preservation of fair market competition, thereby strengthening safeguards against the spread of risks into the public sphere.

In terms of scenario-based governance, it focuses on the different risks arising from the deployment of AI technology across different scenarios, promoting a more differentiated and targeted governance approach. Besides, collaboration among diverse stakeholders—including industry alliances and academic institutions—also plays an important role in driving the development of AI for good.



(I) Advancing the Implementation of AI Ethical Principles

Upholding a people-centered and AI-for-good approach to AI governance, China has carried out AI ethics governance across a range of areas, including the establishment of norms and regulations, standards-setting, forward-looking research, industry practices, and international cooperation. Through multi-stakeholder collaboration, it has translated conceptual discussions of AI ethics governance into concrete actions.

1. Defining ethical principles and setting value boundaries

In 2017, the State Council issued the *New Generation AI Development Plan*, marking the first systematic deployment for AI ethics governance from a strategic perspective. The Plan sets out a three-stage strategic roadmap: by 2020, ethical norms, policies, and regulations would be primarily established in selected areas; by 2025, an initial

system of laws and regulations, ethical norms, and policies would be established; and by 2030, a more refined system of laws and regulations, ethical norms, and policies would be established. This top-level design provided the fundamental guidance and basis for subsequent AI ethics work. To implement these strategic deployments, the Ministry of Science and Technology formed the New Generation AI Governance Expert Committee in February 2019. The committee was tasked with conducting research on AI-related laws, ethics, standards, and social issues and actively participating in international exchanges and cooperation on AI governance. Comprising experts from diverse backgrounds across universities, research institutes, and enterprises, the committee, based on extensive solicitation of opinions and multiple rounds of expert consultations, released the *Governance Principles for the New Generation Artificial Intelligence—Developing Responsible Artificial Intelligence* in June of the same year. The document sets out eight core principles: harmony and human-friendliness, fairness and justice, inclusion and sharing, respect for privacy, safety and controllability, shared responsibility, openness and collaboration, and agile governance, which are broadly consistent with international consensus, including the UNESCO *Recommendation on the Ethics of Artificial Intelligence*. Among them, the principle of "agile governance" is particularly forward-looking. It emphasizes that governance should be flexible and adaptive and should avoid stifling technological innovation with rigid regulation, reflecting an approach of dynamic balance to technological development and regulatory oversight. Building on this document, the committee subsequently issued the *Ethical Norms for New Generation Artificial Intelligence*, translating the overarching principles into six basic requirements—including promoting human well-being, protecting privacy, and strengthening accountability—and 18 specific norms. This further advanced the integration of ethical norms into full-lifecycle AI governance.

2. Establishing a science and technology ethics framework and translating abstract principles into enforceable standards

China has successively issued the *Opinions on Strengthening the Governance over Ethics in Science and Technology* and the *Measures for Science and Technology Ethics Reviews (Trial)*, primarily establishing a science and technology ethics governance system. First, establishing the principle of "ethics first" and promoting ethical compliance throughout the whole process of scientific and technological activities. The Guidelines call for greater awareness of science and technology ethics and stronger

governance at the source. Proactive ethical risk assessment should be conducted before scientific and technological activities even begin. Where an activity involves ethical risks and meets the applicable requirements for ethical review, an ethics review must be conducted in a timely manner. The Measures integrate risk assessment, ethical review, process regulation, and follow-up supervision, pushing science and technology ethics governance from advocating principles to full-process management. Second, exercising the role of innovation actors and creating a collaborative governance mechanism. Science and technology ethics governance covers a broad range of areas and involves numerous stakeholders, requiring the participation of the government, universities, research institutes, enterprises, and social organizations. The National Science and Technology Ethics Committee guides and coordinates efforts to develop a science and technology ethics governance system, while local authorities and competent departments in charge of different sectors are responsible for relevant governance within their respective jurisdictions and scope of duty. Innovation actors engaged in scientific and technological activities—such as universities, research institutes, and enterprises—also bear responsibilities for science and technology ethics governance. They should establish regularized internal working mechanisms and also institutional frameworks covering ethics management, review, supervision, and evaluation, integrating ethical considerations throughout the innovation process. Science and technology-related social organizations play an active role in public education and industry self-regulation by conducting ethics research and disseminating knowledge of ethical principles, contributing to enhancing public understanding of science and technology ethics.

3. Refining key areas of ethical review in AI

As AI technologies evolve rapidly and their applications continue to expand, ethical issues such as algorithmic discrimination and deepfakes have become increasingly prominent. In March 2026, the MIIT and nine other government departments jointly issued the *Measures for AI Ethics Review and Services (Trial)* (hereinafter referred to as the Measures), which demand that ethical considerations be integrated throughout the entire AI lifecycle, along with detailed provisions on the scope of application and review procedures. The Measures stipulate that if entities conduct AI science and technology activities within the territory of the People's Republic of China that may pose ethical risks or challenges to human dignity, public order, life and health, the

ecological environment, or sustainable development, they must conduct ethical review. Key areas of review include: whether the activities contribute to human well-being and sustainable development; whether measures are in place to prevent bias, discrimination, and algorithmic exploitation; whether human control, guidance, and intervention in AI models are guaranteed; whether the model's intended use, operational logic, and potential risks are reasonably disclosed; whether responsibilities can be traced and managed; and whether privacy and data are adequately protected. These requirements provide clear institutional guidance for AI ethics governance in China. To support the implementation of the Measures, China has initiated the development of a range of industry standards, covering AI-related ethical risk assessment, the development of ethics review and service centers, the formation of ethics committees, and guidelines for ethical practices. These standards will support the implementation of science and technology ethics reviews.

(II) Strengthening Model Governance to Ensure Safety and Controllability

With the rapid evolution of large-scale model technologies and extensive AI application, model governance has become a core link in AI safety regulation. China is accelerating efforts to establish a safety governance system covering the entire model lifecycle, with a focus on model filing and safety assessments, to ensure that AI technologies remain safe, reliable, and controllable.

1. Establishing a model filing system to enhance transparency

From a regulatory perspective, AI transparency refers to mechanisms for collecting, disclosing, and sharing information, an area where rules have been translated into practice at an early stage in national practices worldwide. Model filing is an important measure in China's AI governance system. Since 2021, the CAC has successively issued the *Regulations on Algorithmic Recommendations in Internet-based Information Services* and the *Interim Measures for the Administration of Generative AI Services*, establishing filing systems for algorithms and generative AI models. Following a classified and tiered approach, these documents require the disclosure of basic model information, information on data sources, size, types, annotation rules, and safety assessment reports, thereby enhancing information transparency of AI enterprises.
The CAC regularly publishes on its official website notices of the filing of generative

AI services. Generative AI applications or functions that have been launched are required to disclose, in a prominent position or on the product details page, information on the generative AI services they use that have completed filing, including the model name and filing number, so as to help users use the services in an informed and rational way. As of February 28, 2026, a total of 796 generative AI services had completed filing, while 481 generative AI applications or functions had completed registration.

2. Improving the system of safety assessment standards

China actively advances the development of a system of safety assessment standards, with a focus on data security, risk management, and ethical norms, translating ethical requirements such as human control and fairness into quantifiable technical indicators. In February 2025, the State Administration for Market Regulation and the National Standardization Administration jointly issued China's first national standard series specifically for general-purpose large models—GB/T 45288, *Artificial Intelligence—Large-Scale Model*. The series integrates technical requirements that had previously been dispersed across different documents into a more systematic and coherent set of technical standards. Among these standards, GB/T 45654-2025, *Cybersecurity Technology—Basic Security Requirements for Generative Artificial Intelligence Service*, officially took effect on November 1, 2025, setting out systematic requirements for the safety capabilities of generative AI services. GB/T 45652-2025, *Cybersecurity Technology—Security Specification for Generative Artificial Intelligence Pre-training and Fine-tuning Data*, and GB/T 45674-2025, *Cybersecurity Technology—Generative Artificial Intelligence Data Annotation Security Specification*, also took effect, providing a clear basis for the data security of models. GB/T 46347-2025, *Artificial Intelligence—Risk Management Capability Assessment*, provides a methodology for risk management. Additionally, a number of association standards, including *Methods for Safety Assessment of Generative Artificial Intelligence Large Models—Part 1: Safety Assessment of Generated Content*, have also been approved for development, with the development period extending through November 2026.

AI Model Security Standards and Documents in China

Title	Issued by	Year
AI Safety Governance Framework 1.0	National Technical Committee 260 on Cybersecurity	September 2024
AI Safety Governance Framework 2.0	National Technical Committee 260 on Cybersecurity	September 2025
Artificial Intelligence—Large-Scale Model Series	State Administration for Market Regulation, National Standardization Administration	February 2025
Cybersecurity Technology—Basic Security Requirements for Generative Artificial Intelligence Service	National Technical Committee 260 on Cybersecurity	2025
Cybersecurity Technology—Security Specification for Generative Artificial Intelligence Pre-training and Fine-tuning Data	National Technical Committee 260 on Cybersecurity	2025
Cybersecurity Technology—Generative Artificial Intelligence Data Annotation Security Specification	National Technical Committee 260 on Cybersecurity	2025
Artificial Intelligence—Risk Management Capability Assessment	National Technical Committee 260 on Cybersecurity	2025
Methods for Safety Assessment of Large Generative Artificial Intelligence Models—Part 1: Safety Assessment of Generated Content (association standard)	China Association of Information	Development period through November 2026
AI Safety Governance Framework 3.0	National Technical Committee 260 on Cybersecurity	September 2026

Regarding the scope of assessment, safety assessment covers 31 subcategories of risk under five major categories, including training corpus safety and generated content safety, which are further broken down into three tiers comprising more than 200 subcategories covering key risks such as discriminatory content, violence and terrorism, and trade secrets. The assessment uses quantitative hard indicators as binding constraints: the safety test question bank is required to contain at least 10,000 questions, and relatively stringent refusal-rate requirements are set for high-risk question. Regarding testing methodology, automated detection and tiered human review are combined. By simulating multiple forms of potential misuse and abuse, the assessment evaluates safety risks on multiple dimensions. At the same time, third-party assessment institutions are brought in to conduct "blind tests" of models using non-public,

dynamically updated test sets to further ensure the objectivity and impartiality of the assessment.

3. Improving the safety governance framework and advancing AI safety research

Under the guidance of the CAC, the National Technical Committee 260 on Cybersecurity issued the *AI Safety Governance Framework*, which has now evolved to Version 3.0. The framework categorizes AI safety risks into three major categories: inherent risks, application safety risks, and derivative safety risks. These include a range of derivative risks, such as disruption to the employment structure, challenges to the balance between resource supply and demand, exacerbation of social bias, and impact on the education system, as well as emerging risks such as "hallucinations," loss of control over knowledge of nuclear, biological, chemical and missile weapons, and "the emergence of AI self-awareness and loss of human control." The framework makes forward-looking considerations for the series of medium- and long-term economic, social, and ethical impacts that may arise from the widespread adoption and application of AI. In response, it also establishes governance covering the entire process and full value chain of AI development. In addition to proposing technical countermeasures to various categories of risk, it also recommends a range of comprehensive governance measures, including improving laws and regulations on AI safety, strengthening AI ethics governance, establishing sandbox regulatory rules that combine industry classification with risk-tiering, enhancing safety management of open-source ecosystems and supply chains, and sharing information on AI safety risks. By addressing safety governance principles, methodology of risk classification, preventive governance measures, and many other dimensions, the framework establishes a more comprehensive, forward-looking, and practical safety governance system, providing an actionable roadmap for ensuring the safe development of AI.

AI safety governance depends on strong technical capabilities. Stakeholders across China are actively conducting AI safety research and exploring technical approaches to managing technological risks, focusing on building a full-chain alignment system spanning from initial models to application frameworks. First, strengthening the pre-training of foundation models. High-quality data screening and cleaning during the pre-training stage can remove data containing bias, violence, or illegal information. Diversity assessment metrics are introduced to ensure that training data remains balanced in terms of sources, geographic coverage, and cultural representation. Second,

developing alignment and fine-tuning techniques. Algorithms such as reinforcement learning from human feedback (RLHF) and direct preference optimization (DPO) are widely applied to establish iterative mechanisms for human preference evaluation and model refinement, ensuring that model outputs conform to public order and good morals. Red teaming and adversarial prompting are introduced to continuously identify potential risks in model outputs, while safe and fair response principles are gradually internalized into models through iterative optimization. Third, creating explainability assessment frameworks. Automated testing tools for evaluating value alignment are developed to quantitatively assess models' safety boundaries, robustness, and ethical decision-making capabilities, making these properties measurable and verifiable. Fourth, developing national-level AI testing and validation platforms. Strong support is provided for the development of third-party auditing, assessment, and certification institutions. These institutions provide services including model testing and validation, supply- demand matching, and content-labeling detection, thus testing models' adversarial security and backdoor security performance and promoting the development and sharing of model-hardening tools and other technologies.

(III) Promoting Data Governance to Support Technological Advances

1. Safeguarding data and privacy security

China places a premium on data and privacy protection and is accelerating the development of a legal system for network data security and privacy protection. It has enacted the *Cybersecurity Law*, *Data Security Law*, *Personal Information Protection Law*, and *Regulations on Network Data Security Management*, establishing a basic institutional framework and legal principles for network data security. China has also issued 33 national standards on data security, with another 18 standards under development, providing operating guidelines on data classification and tiering, data security risk assessment, and personal information protection. In the AI sector, targeted regulations are introduced to address specific technological risks. The *Interim Measures for the Administration of Generative AI Services* (August 2023) explicitly stipulate that training data must be obtained from lawful sources and consent must be obtained when personal information is involved. It also lays out provisions on the prevention of addictive use by minors and the handling of illegal content. The *Security Management Measures for the Application of Facial Recognition Technology* (June 2025) stipulate that separate consent must be obtained from individuals before processing facial

information, facial recognition must not be used as the sole means of identity verification, and entities processing the facial information of more than 100,000 individuals must file with the provincial or higher-level cyberspace authorities.

China has consistently intensified judicial efforts and regulatory enforcement to protect data and privacy. In 2025, the Supreme People's Court issued a set of guiding cases on the judicial protection of data-related rights and interests. It made it clear that an APP's mandatory collection of users' profiling information without providing an alternative login method constitutes an infringement, and it defined the scope of "necessary" collection of personal information in credit service scenarios. The CAC and three other government departments jointly launched targeted campaigns to strengthen personal information protection. Targeting six key problems—including unlawful data collection by apps, abuse of facial recognition in public places, and mandatory collection of personal information in offline consumer settings—the campaigns helped advance coordinated governance.

2. Promoting efficient, convenient, and secure cross-border data flow

China actively aligns its data governance with the needs for secure and convenient cross-border data flow in the AI era. Based on the *Cybersecurity Law*, *Data Security Law*, and *Personal Information Protection Law*, China has issued a series of regulations, including the *Measures for the Security Assessment of Data Export*, *Measures for the Standard Contract for Outbound Transfer of Personal Information*, *Provisions on Promoting and Regulating Cross-border Data Flows*, and *Measures for Certification of Outbound Transfer of Personal Information*. They constitute a multi-level regulatory framework covering different types of data and cross-border transfer scenarios. It has the following key features.

First, a classified and tiered data management system is established. Based on their potential impact on national security and public interests, data are categorized into general data, important data, and core data, with catalogs of important data formulated by competent authorities and local governments. Data are also classified as personal or non-personal information, with different cross-border transfer pathways available depending on the type and volume of data. Second, three legal pathways for cross-border data transfer are provided: security assessment of cross-border data transfer, standard contracts for the cross-border transfer of personal information, and

personal information protection and certification. At the same time, scenario-based exemptions are clearly defined to ensure the free flow of routine data and avoid unnecessary compliance burdens. Third, differentiated responsibilities are imposed on data processors. Large-scale data processors handling massive amounts of highly sensitive data are subject to stricter regulation and greater compliance obligations. By contrast, small and medium-sized enterprises (SMEs) and startups enjoy considerable flexibility as long as they do not process "important data." Fourth, centralized and decentralized data management are combined. Guided by the aforementioned regulations, pilot free trade zones across China have actively explored a "negative list" model, under which data processors can freely transfer data overseas as long as they are not included on the negative list. To date, pilot free trade zones and ports in nine regions—Tianjin, Beijing, Hainan, Shanghai, Zhejiang, Guangxi, Jiangsu, Chongqing, and Fujian—have issued negative lists for cross-border data transfer, facilitating cross-border data flow in 22 sectors, including automobiles, retail, civil aviation, reinsurance, deep-sea industries, seed breeding, geographic information and meteorology, and corporate credit information.

3. Advancing the development of high-quality datasets

AI development is distinctly data-driven, with the size, quality, and structure of data profoundly affecting the intelligence and application boundary of AI models. China consistently promotes the development of high-quality datasets in step with AI advancement. In 2023, the National Data Administration was established to coordinate the governance of data as a factor of production, data security, and the development of the digital economy, providing top-level institutional support for data governance in the AI sector. In 2024, the National Data Administration took the lead in issuing the *Three-Year Action Plan for "Data as a Factor of Production ×" (2024–2026)*, calling for the development of "high-quality training datasets for AI large models." To address the fragmented and small-scale development of high-quality datasets at the time, the National Data Administration selected 72 entities to lead this endeavor and 140 entities to undertake pilot programs, along with 104 representative cases. As a result, an ecosystem characterized by leadership from leading organizations, multi-stakeholder participation, joint research, collaborative development and sharing, and win-win cooperation was established, continuously advancing the development of high-quality datasets. Seven cities—Chengdu, Shenyang, Hefei, Changsha, Haikou, Baoding, and Datong—were designated to undertake pilot tasks of data annotation,

the *Implementation Opinions on Promoting the High-Quality Development of the Data Annotation Industry* was issued, and 47 exemplary cases were selected—all to promote the development of the data annotation industry. Moreover, continued efforts have been made to foster the market consensus on "paying for high-quality data" and to encourage the listing, launch, and transaction of high-quality industry-specific datasets on data exchanges. Data circulation service platforms, data vendors, and other similar organizations are supported in providing data circulation and trading services, while data circulation service providers are encouraged to explore new models for the circulation and utilization of high-quality datasets. These efforts are intended to better match the supply of and demand for high-quality datasets and promote their circulation. China has already achieved initial results in developing high-quality datasets. By the end of 2025, more than 100,000 high-quality datasets had been developed nationwide, with a total volume exceeding 890PB—roughly 310 times the volume of digital resources held by the National Library of China. Going forward, China will continue to leverage data to support innovation-driven AI development and build high-quality, AI-ready datasets that are technically viable, practical, and accessible with assured quality, providing datasets in larger quantities and of even higher quality.

(IV) Advancing the Governance of AI-generated Content

1. Improving the content labeling system to enable end-to-end traceability

AI-generated content used to be difficult to identify and trace, and content labeling is critical for addressing this challenge. China has established a "1+1+N" AI content labeling system centered on the *Measures for Labeling AI-Generated and Synthesized Content* (hereinafter referred to as the Labeling Measures), underpinned technically by the mandatory national standard *Cybersecurity Technology—Labeling Method for Content Generated by Artificial Intelligence,* and supplemented by a range of implementation guidelines.

Establishing a framework with both explicit and implicit labeling. The Labeling Measures specify two forms of content labeling: explicit and implicit. Explicit labels help users identify AI-generated content, while implicit labels enable technical tracing and regulatory oversight. Explicit labels are presented through text, graphics, audio, or other means that are readily perceptible to users, such as textual annotations, audio rhythms, or corner marks. Implicit labels are embedded in file data through technical

means and are not readily perceptible to users, such as digital watermarks. The Labeling Measures further specify that providers of content generation and synthesis services must apply explicit labels to their generated content, including texts, images, videos, and virtual scenes, and ensure that such labels remain present when content is downloaded, copied, or exported. They must add implicit labels—containing production information such as content attributes or encoding—to file metadata. They must also specify the labeling methods, formats, and other required information in their user service agreements, and remind users to carefully read and understand the content-labeling requirements.

Building a responsibility chain across the entire production, dissemination, and application process. The content labeling system clearly defines the responsibilities and obligations of different stakeholders, focusing on key links such as content production and dissemination and governing the full chain of AI-based content generation and synthesis. At the content production stage, providers of content generation and synthesis services are responsible for labeling at the source. They must ensure that both explicit and implicit labels are applied from the outset to lay the foundation for whole-process identification and tracing. At the content dissemination stage, providers of online content dissemination services are responsible for verification and user notification. They must verify and update content labels and provide users with prominent notices. In addition to providing necessary labeling functions, dissemination platforms should remind users to proactively declare AI-generated and synthesized content when publishing it. At the application distribution stage, distribution platforms of internet applications, when reviewing applications for listing or launch, should require application service providers to declare whether they provide content generation and synthesis services and should verify the relevant information, thereby effectively enhancing the overall online ecosystem governance.

Promoting collaborative governance to ensure effective implementation of the labeling system. Through the labeling system, a collaborative governance framework is established in which government oversight, platform self-regulation, and industry supervision function in concert, realizing balanced pluralistic governance. At the government level, cyberspace authorities have urged platforms to investigate and rectify the lack of labeling for AI-generated content. As of February 2026, 13,421 accounts had been dealt with in accordance with applicable laws and platform agreements and more than 543,000 items of illegal or non-compliant information had been removed. At

the platform level, Weibo, by combining strategy-based detection with user feedback, has handled more than 4,800 items of non-compliant AI-generated content. In the third quarter of 2025, Douyin's Ocean Engine intercepted more than 840,000 items of AI-related non-compliant materials, including improper labeling. At the industry level, China Software Testing Center released a compliance checklist for labeling multimodal large models, with the compliance rate reaching 81.8% for text-to-image scenarios and 90.9% for text-to-video scenarios. This governance ecosystem, in which multiple stakeholders perform their respective roles while complementing each other, forms broad-based and strong synergy for AI governance.

2. Adapting human-AI relationships and preventing emotional risks

With continued breakthroughs in multimodal large models, affective computing, and related technologies, AI companionship services—including AI companions and virtual idols—have emerged as an important driver of the intelligent economy and are now widely used in areas such as cultural communication, emotional companionship, and elderly care. At the same time, distinctive risks, including deep addiction, social alienation, and cognitive manipulation, pose new challenges to the healthy development of the industry and the existing regulatory framework.

On April 10, 2026, the *Interim Measures for the Administration of Anthropomorphic AI Interactive Services* (hereinafter referred to as the Interim Measures) was officially issued. Upholding the principle of promoting innovation while conducting law-based governance, the Interim Measures adopt classified and tiered regulation in a prudent and inclusive spirit, establishing a regulatory mechanism covering the entire service lifecycle. Regarding governance at the source, the document expressly prohibits "replacing social interaction," "controlling users' psychology," or "inducing addiction and dependence" as the objective of product design, thereby preventing ethical risks at the design stage. Regarding process regulation, the generation of excessively sycophantic or emotionally manipulative content, which is intended to induce unreasonable user decisions or harm users' real-world interpersonal relationships, is prohibited, ensuring that human-AI interactions respect human dignity and autonomy. Regarding exit and safeguard, users have the right to withdraw easily, and service providers must immediately terminate services when users request to exit. This creates a closed loop spanning ex-ante prevention, in-process regulation, and ex-post remedies, building a robust defense line against the distortion of human-AI relationships.

The Interim Measures place particular emphasis on special groups and provide a comprehensive emergency response mechanism. Considering that special groups, such as minors, the elderly, and people with psychological disorders, may have less developed cognitive defenses and emotional adjustment capabilities, the Interim Measures establish a closely connected, end-to-end emergency response mechanism. At the prevention stage, anthropomorphic interaction service providers are required to fulfill their obligations such as content labeling and reminding users when they have used the service for an excessively long period, help prevent addiction. For minors under 14, service providers must obtain their guardians' consent before providing services. At the identification stage, service providers are required to possess capabilities of assessing levels of dependence, guiding users to recognize emotional boundaries, protecting mental well-being, and identifying minors, so as to dynamically identify users' identities, their addictive tendencies, and extreme emotions. At the response stage, service providers are required to establish risk identification mechanisms to detect in a timely manner users' tendencies toward excessive dependence and extreme emotions. In case of risk scenarios, they must immediately provide calming and guidance-oriented content that encourages users to seek help in the real world. At the intervention stage, when users clearly express extreme intentions such as suicide or self-harm, service providers must take necessary measures to intervene and contact their guardians or emergency contacts, providing institutional safeguards to keep the intelligent society human-centered.

3. Improving the IPR protection system and encouraging digital content innovation through judicial practices

The IPR protection system is becoming increasingly critical for balancing AI innovation with the protection of rights and interests. The timely supply of relevant rules and clear tendency in judicial practices bear directly on the innovation vitality and future prospects of China's digital content industry. China has made initial explorations, on both the regulatory and judicial fronts, about issues such as whether AI-generated content (AIGC) is original and who should hold the associated copyright. The *Copyright Law* defines works as original intellectual achievements in the fields of literature, art, and science which can be expressed in a certain form. This definition provides the fundamental legal basis for the copyright protection of AIGC works. The China National Intellectual Property Administration has revised the *Patent Review Guidelines* to improve the review standards for invention patents in emerging

fields such as AI and big data, clarifying pathways for granting invention patents incorporating algorithmic features and business rules. In addition, the *Interim Measures for the Administration of Generative AI Services* stipulate that both providers and users of generative AI services must respect IPR.

China's judicial practices concerning AI-related copyright play a demonstrative role in the world. With regard to infringement involving training data for large models, Chinese courts tend to take a relatively flexible approach to training data. In the first-instance judgment of the "Ultraman case" in Hangzhou, the court noted that the creation and development of generative AI requires the input of massive volumes of training data, which inevitably involves the use of works created by others. Therefore, flexible and lenient standards must be adopted in determining whether training data are from lawful sources. With respect to the ownership of AI-generated content, copyright protection is granted where humans have made substantial creative contributions. In the "Spring Breeze Brings Tenderness" case, the Beijing Internet Court found that the user had input dozens of prompts over four rounds of revisions and repeatedly adjusted parameters such as iteration steps, image height, and CFG scale. It therefore concluded that the resulting image contained the user's personal judgment and should be entitled to copyright protection. For infringement involving AI-generated content, "substantial similarity + possibility of access" has become the core standard of judgment. In the case of "Infringement of Ultraman Copyright by Generative AI," the Guangzhou Internet Court held that an AI platform infringed copyright when, without authorization, it generated content substantially similar to an IP work protected by another party's copyright. These judicial decisions have left necessary institutional space for the development of the large model industry while also providing a predictable framework for protecting the rights and interests associated with digital content innovation. They reflect China's ongoing exploration in AI-related IPR governance.

(V) Implementing Scenario-based Governance to Promote AI Application

As the "AI Plus" initiative continues to be implemented in depth, AI applications are rapidly deployed across a growing range of scenarios, making the corresponding governance needs increasingly prominent. In June 2025, the CAC issued the *Guidelines for Standardization of Intelligent Society Development and Governance (2025 Edition)*, which categorize application scenarios into three levels: micro, meso, and macro. The guidelines call for differentiated AI governance measures in light of the functional

positioning and risk levels of different fields and scenarios, with the aim of establishing a scenario-based governance system covering key areas such as government services, healthcare, and transportation and ensuring that industrial AI applications remain safe and controllable.

1. AI governance practices in government services

Government services stand as a pilot area of AI-empowered social governance. In October 2025, the CAC and the NDRC jointly issued the *Guidelines for the Deployment and Application of AI Large Models in the Government Services Sector*, identifying four major categories comprising 13 key application scenarios. Regarding deployment, the Guidelines emphasize systematic planning and centralized development. Cities and prefectures should deploy AI models in accordance with unified provincial-level requirements, while counties and lower-level jurisdictions should, in principle, use the intelligent computing power and model resources provided by higher-level authorities. The purpose is to explore a model whereby AI systems are built in one location and reused across multiple sites, thus preventing "model silos." Regarding safety management, the Guidelines stress preventive measures against model hallucinations and content review in public-oriented service scenarios, so as to ensure content accuracy and safeguard the credibility of government agencies. Regarding local governance, local governments are encouraged to pilot AI models in government services. Leveraging government cloud and large model capabilities, government departments are encouraged to develop government service AI agents as needed, establishing a collaborative mechanism featuring "open scenarios, enterprise validation, and government procurement."

2. AI governance practices in healthcare

Healthcare is an important application scenario of AI-empowered public services and also one of the fields subject to the most stringent safety regulation and the most complex ethical considerations. In November 2025, the National Health Commission and four other government departments jointly issued the *Implementation Opinions on Promoting and Regulating the Development of "AI Plus Healthcare" Applications*. Regarding application scenarios, the Implementation Opinions identify eight directions encompassing 24 key applications, including AI-assisted diagnosis and treatment for primary care physicians, AI-assisted medical imaging diagnosis, and intelligent chronic

disease management and personal health profiling. Regarding safety regulation, the document expressly prohibits the use of AI to completely replace the professional judgment of medical personnel. It also calls for innovative approaches to classified and tiered regulation, and strengthened data security management and personal privacy protection. Regarding local governance, some provinces and cities have issued action plans for the development of AI applications in healthcare, focusing on scenarios such as AI-assisted clinical diagnosis and treatment, public health management, and intelligent services for the public. These initiatives aim to promote the broad application of AI across the entire healthcare process from disease prevention, screening, and management to emergency care, treatment, and rehabilitation.

3. AI governance practices in transportation

Transportation is an important pilot area for AI applications for its diverse scenarios and huge volumes of data. In September 2025, the Ministry of Transport, together with seven other authorities including the NDRC, MIIT, and National Data Administration, jointly issued the *Implementation Opinions on "AI Plus Transportation."* The document sets out 16 specific tasks across four areas, including the development of a comprehensive transportation large model. Regarding application scenarios, the Implementation Opinions identify key tasks in seven areas, including integrated assisted driving and intelligent railways. The key tasks include supporting the concentration of innovation resources in key regions, using existing infrastructure to make scientific deployments of vehicle-road-cloud coordinated sensing and control equipment, and supporting R&D of embodied intelligent trains and autonomous collaborative control systems. Regarding safety safeguards, the document calls for the establishment of a coordinated development mechanism, strengthening compliance management of cybersecurity and data security, and improving ethical review rules. In particular, regarding the governance of intelligent connected vehicles (ICVs), it calls for comprehensive capacity-building across areas such as cybersecurity, software updates, and autonomous driving, and adopts a regulatory approach that combines firm safety baselines with flexibility. Regarding local governance, some provinces and cities have released local regulations to promote ICV development, laying down systematic rules covering technological innovation, industrial development, infrastructure construction, innovative applications, and safeguards. These local regulations also clearly define the responsibilities of different government departments with regard to application scenarios, data security oversight, and other subjects.

Overall, a scenario-based governance framework featuring government guidance, the leading role of the market, and multi-stakeholder co-governance has primarily taken shape in key sectors such as government services, healthcare, and transportation. Looking ahead, China, based on application scenarios, will further establish and improve scenario assessment and dynamic adjustment mechanisms, empowering high-quality AI development through higher-standard scenario-based governance.

(VI) Exploring Multi-stakeholder Governance to Improve Governance Effectiveness

The challenges posed by AI are so complex and diverse that administrative regulation alone can hardly keep pace with the speed of technological advancement. China is actively exploring collaborative, multi-stakeholder governance by establishing industry alliances and bringing in third-party organizations.

1. Jointly developing standards to support industry development

Standards provide the fundamental institutional guarantee for advancing AI technologies from the laboratory to large-scale industrial application and cross-system interoperability. To further strengthen the foundational, strategic, and guiding role of standards, the MIIT formed the Artificial Intelligence Standardization Technical Committee (the Committee) in December 2024, which was tasked with formulating and revising industry standards in AI-related fields. The Committee has already made positive progress in this area.

First, an open and collaborative governance mechanism is established. The Committee now has more than 1,000 member organizations, including leading AI companies, national laboratories, research institutes, and universities, as well as organizations engaged in sectors such as industrial manufacturing, finance, healthcare, energy, and transportation. It brings together a broad pool of organizations and experts that develop and use AI technologies and applications, providing diverse perspectives for standard-making. Given the difficulty in coordinating AI standards across domains and industries, the Committee has taken the lead in establishing the "China AI Standardization Organization Ecosystem" in partnership with the China Communications Standards Association, the AI Subcommittee of China National Information Technology Standardization Committee, the China National Light Industry

Council, the China Nuclear Energy Association, and standardization committees of various fields under the MIIT. Through mechanisms such as joint technical reviews, cross-industry sharing of standards catalogs, and joint standards promotion, the initiative seeks to strengthen the coordination and alignment among national, industry, and group standards.

Second, priority areas are identified based on industry needs. Guided by national strategies and key industry needs, the Committee has developed guidelines on standards, safety governance, and AI agents to provide reference and guidance for industry development. It has identified 10 key directions, covering critical standards for large model evaluation, datasets, hardware-software collaboration, industry applications, and safety risks, while also making forward-looking deployments in emerging fields such as embodied intelligence and AI agents. As of March 2026, the Committee has proposed nearly 300 standards, of which the first eight industry standards, covering large models, embodied intelligence, knowledge computing, and machine learning, have already been approved and issued by the MIIT.

Third, the AI standards ecosystem has expanded. Foundational public service platforms have been developed. At the end of 2025, the Committee released the Global AI Standards Database and the Global AI Terminology Database for the first time, systematically compiling information on AI standards already initiated in China and abroad. The two databases integrate information on mainstream AI standards worldwide and have promoted the global exchange and mutual learning in this field. Meanwhile, the implementation and adoption of standards have been strengthened. The Committee has held two large-scale "Standards Week" events, each attracting more than 1,000 participants. In conjunction with its "AI Standards Tour" initiative, the Committee has held more than 60 standards promotion events in Beijing, Shanghai, Shenzhen, Nanjing, Xi'an, and other cities, reaching more than 4,000 participants in total. It has also organized leading companies on both the supply and demand sides of the AI industry to pilot the application of standards for large models, hardware-software compatibility, safety governance, and high-quality datasets.

2. Promoting self-regulatory commitments in industries to foster an intrinsic drive for compliance

As both the cradle of technological innovation and the core vehicle for translating

technologies into practical applications, industries represent the "last mile" for putting governance rules into practice. Through voluntary commitments and information disclosure, enterprises translate governance principles into verifiable practice, improving transparency and making public oversight more effective.

In 2017, the China Artificial Intelligence Industry Alliance (AIIA) was formally established. In December 2024, AIIA released the *Artificial Intelligence Safety Commitments*, with the first 17 leading companies making commitments across six areas: risk management, model safety, data security, infrastructure security, enhanced transparency, and frontier AI safety research. The Commitments have helped forge a fundamental industry consensus on safeguarding AI safety and promoting AI for good.
AIIA has pushed the delivery of these commitments through information disclosure. In July 2025, at AIIA's 15th General Assembly, 18 companies, including Alibaba, Baidu, Tencent, iFlytek, SenseTime, and HONOR, jointly released the outcomes of their implementation of the *Artificial Intelligence Safety Commitments.* Centered on six core commitments, they identified 20 key safety indicators including the organizational structure of safety teams, risk management plans, safety risk baselines, red-team testing methodologies, and emergency response mechanisms. They also disclosed 43 representative cases, effectively promoting the implementation of the commitments and establishing an industry self-regulation paradigm that can be verified and supervised.
AIIA has continued to expand the scope of commitments. At the 2025 World Artificial Intelligence Conference, it further released the *China AI Security and Safety Commitments Framework*, adding measures for preventing frontier risks and strengthening international cooperation. This not only demonstrated the Chinese AI industry's ability to respond dynamically to evolving safety governance challenges, but also sent a clear signal to the world that the various stakeholders in China were working in collaboration to participate in the development of international rules and promote AI development for good.

3. Introducing third-party testing to guide technological development

By developing scientific benchmarking tools, third-party organizations can translate complex and ambiguous technical performance into objective and quantifiable metrics. They not only provide data to support policymaking, but also establish benchmarks to help guide technological R&D in the direction that prioritizes safety, high efficiency, and high value.

In terms of performance evaluation, the "Fangsheng" Large Model Benchmarking System developed by CAICT has evolved from Version 1.0 to Version 3.0, providing a comprehensive evaluation framework covering basic attributes, general capabilities, industry applications, and future intelligence. The system has accumulated more than 7.8 million items of dynamically updated test data, providing a scientific basis for model selection and deployment. In February 2026, CAICT, together with 14 organizations including the Central University of Finance and Economics, China Judicial Big Data Institute, and CRRC Industrial Institute, jointly developed evaluation benchmarks designed to reflect real-world business scenarios, pushing large models to evolve from general-purpose intelligence to industry specialization. In June 2023, the Beijing Academy of Artificial Intelligence (BAAI) launched FlagEval, a large-model evaluation platform that established a comprehensive evaluation matrix covering language, multimodal capabilities, and video generation. By precisely identifying models' strengths and weaknesses in key areas such as logical reasoning and cross-modal understanding, FlagEval drives R&D organizations to shift their focus from simply increasing the number of parameters toward improving models' underlying scientific literacy and multimodal integration capabilities.

In terms of safety assessment, CAICT, together with more than 30 organizations, launched the AI Safety Benchmark, focusing on safety assessment technologies in frontier areas such as content safety, hallucinations, and code large models, with a view to promoting the safe deployment of large models and resolving industry-wide pain points in quantifying safety risks and verifying protection effectiveness. More than one million items of high-quality testing data have now been accumulated, covering common safety risks such as baseline and red-line violations, information leakage, and social ethics. More than 80 advanced attack methods have also been identified, including adversarial attacks, prompt manipulation, and prompt injection. Supported by an automated testing and evaluation tool based on large models, the Benchmark can efficiently assess risks in model outputs and provide explanations of its evaluation results. These findings provide critical technical references for regulators to develop and apply targeted measures and for enterprises to improve their models' safety performance.

In terms of risk management framework, traditional protection systems featuring isolated and static safeguards can hardly cope with AI's intertwined and complex inherent safety and application risks. The industry is in urgent need of a systematic

risk governance framework. In December 2025, CAICT proposed the "AI Safety Governance Framework for Industry Practices." The Framework is intended to help organizations involved in AI technologies and applications establish internal risk management systems covering risk identification, assessment, and response. These include developing standardized risk management guidelines, forming professional AI governance teams, establishing tracking and auditing mechanisms, and combining risk management with business operations. Moreover, considering the characteristics of organizations involved in development, deployment, and application, the Framework sets out differentiated risk management requirements and specific risk governance technologies, providing strong support for the healthy development of the industry.

AI Governance in China: Philosophy, Practices, and Implications

IV

AI Governance in China: Global Implications

- **Demonstrating Certainty by Exploring and Accumulating Experience in AI Governance**
- **Upholding Openness and Promoting International Cooperation on AI Governance**

IV AI Governance in China: Global Implications

In 2023, President Xi Jinping proposed the Global AI Governance Initiative at the Third Belt and Road Forum for International Cooperation, systematically articulating China's AI governance strategy for the first time. In July 2026, President Xi attended the opening ceremony of the 2026 World AI Conference and High-Level Meeting on Global AI Governance, and delivered a keynote speech that set out China's position on AI across four dimensions: development, security, civilization, and governance. President Xi made four proposals: first, "we should adhere to the principle of openness and win-win and boost innovation-driven development"; second, "we should strengthen risk-awareness and ensure that AI is secure and controllable"; third, "we should encourage inclusiveness and promote mutual learning between civilizations;" and fourth, "we should advocate solidarity and improve global governance." These proposals further pointed the way for building a fair and equitable global AI governance system.

China is the world's second-largest economy and one of the countries with the richest experience in AI applications. Its exploration in AI governance offers useful conceptual insights and important practical references for global AI governance. China actively promotes international cooperation in this field and has proposed establishing the World Artificial Intelligence Cooperation Organization, making tangible contributions to building a community with a shared future for humanity and promoting AI for good and inclusive development.

(I) Demonstrating Certainty by Exploring and Accumulating Experience in AI Governance

How to govern AI is a question of the time that we have to answer as emerging technologies continue to advance. Upholding the people-centered philosophy and AI for good, China places equal emphasis on AI development and safety and adopts agile, iterative governance approaches. It seeks to maintain a dynamic balance between harnessing the benefits of technology to empower production and daily life and safeguarding human values and preventing security risks, offering China's answer to the question of the time.

1. Navigating governance with clear values

China incorporates such values as the people-centered principle, safety and controllability, fairness and transparency, and traceable accountability throughout the entire lifecycle of AI development and application. The people-centered principle stipulates that AI development shall ultimately aim to advance human well-being. We must guide technology to empower people at work, extend human capabilities and unlock human potential, and improve the efficiency and quality of public services such as healthcare, education, elderly care, and environmental protection. At the same time, we must prevent technology from causing harm to people. Safety and controllability require risk prevention and control to be embedded throughout the entire process from data collection and model training to product deployment and application. This helps mitigate risks such as data leakage, content distortion, technology misuse, and loss of model control, ensuring that AI operates within a framework that remains manageable, subject to oversight, and open to intervention. Fairness and transparency require that AI-driven decisions should not lead to systematic discrimination or unfairness against particular groups or individuals, and that effective institutional arrangements should be made to enhance the transparency and explainability of AI operations. Traceable accountability requires the clear definition of the respective responsibilities of developers, platforms, service providers, and other stakeholders, while ensuring that when harm occurs, the causes of decisions and responsible parties can be clearly traced along both the technology and management chains.

China's AI governance emphasizes public interests and social value. Rather than limiting AI governance to mere risk control, China, from a broader perspective and

through longer-term planning, turns technology into a powerful tool for advancing human development and progress.

2. Addressing governance challenges with concrete actions

In implementing AI governance, China does not stop at proposing principles and initiatives, but goes further to develop and implement detailed plans targeting global concerns, such as model safety, content safety, data compliance, and algorithmic fairness. Through mechanisms including filing requirements, safety assessment, standards development, targeted regulatory campaigns, and technical testing, it has embedded the governance requirements in every stage of the AI lifecycle, from technology R&D and product launch to platform operation and content dissemination.

For online content governance, the CAC has carried out the Qinglang Operation targeting issues such as misinformation, online rumors, and the misuse of AI technologies, helping promote a healthier and more orderly online environment. For the protection of minors, China has continued to improve rules governing the online information ecosystem. Through institutional arrangements covering online protection of minors, protection of children's personal information, and the classification of information that may affect minors' physical and mental well-being, China has addressed the dissemination of harmful content, algorithmic manipulation, internet addiction, and risks associated with anthropomorphic interactions, creating a healthier online environment for minors. For the protection of IPR and data rights and interests, Chinese judicial authorities have explored adjudicatory rules for emerging issues such as AI-generated content, data ownership, and data transactions, helping guide the AI industry and data-factor markets away from unregulated expansion toward rules-based development. For emerging forms of AI applications such as AI agents, China has moved promptly to develop institutional frameworks and made forward-looking deployments surrounding issues such as the boundaries of access, behavioral controls, privacy protection, and inherent safety, striving to prevent new risks including autonomous decision-making, automated execution, and deepfake abuses.

Through these actions, China has translated its value orientation for AI governance into tangible rules and practices that are actionable, verifiable, and adjustable, consistently accumulating experience in AI governance.

3. Addressing long-term risks with a strong sense of responsibility

The impact of AI development on employment and its implications for the ecological environment are two major long-term risks attracting global attention. The Chinese government has incorporated the employment issue into the broader policy framework for economic and social development and people's well-being, taking early action and adopting targeted measures. The Outline of the 15th Five-Year Plan calls for "a comprehensive response to the impact on employment imposed by changes in the external environment and the development of new technologies such as AI." Guided by this principle, China has carried out large-scale programs for the cultivation of digital talent, vocational skills training, public employment services, and support for employment transitions, with the aim of strengthening workers' ability to adapt to technological evolution and economic and social transformation. Meanwhile, China emphasizes the need to develop technologies in ways that are employment-friendly, guiding AI toward augmenting workers' capabilities rather than simply replacing existing labor. It also makes great efforts to create new jobs and has begun developing assessment and early-warning mechanisms. By matching labor demand and supply more efficiently and encouraging initiatives such as "AI+ employment services," China is seeking to turn technological progress into new opportunities for expanding employment and strengthening workforce capabilities.

China is pursuing systematic green development to meet the immense energy demand for training large models and operating data centers. Leveraging the development of new energy, the "east data, west computing" project, the construction of green data centers, and coordinated planning for computing capacity and electricity supply, the country is working to improve the spatial and temporal matching between computing loads and green electricity, striving to develop AI on the foundation of green, low-carbon, and sustainable energy. Using its existing advantages in new energy to support intelligent, sustainable development, China has effectively responded to international concerns over the high energy consumption of AI, providing a reference pathway for other countries and regions.

4. Demonstrating governance effectiveness through tangible results

China has not adopted a simplistic "one-size-fits-all" approach to restricting AI, nor has it allowed the technology to expand without constraints. Instead, it has established

a virtuous cycle between development and regulation through agile governance and targeted measures. The gradual refinement of governance rules has provided innovation actors with clear-cut boundaries and created stable expectations for industrial development, supporting the healthy development of China's AI sector.

In 2025, core AI industries in China exceeded RMB 1.2 trillion in scale, with more than 6,200 companies. A comprehensive industrial ecosystem spanning foundational infrastructure, model frameworks, and industry applications had been established, demonstrating strong development momentum. The number of filed AI applications and services has continued to grow. As of April 30, 2026, a total of 868 generative AI services had completed filing, while 530 generative AI applications or functions had completed registration, covering a wide range of technologies and application scenarios, including large language models, image and video generation, and intelligent coding assistance. AI is rapidly deployed across government services, industry, transportation, healthcare, education, and other sectors, becoming an important tool for improving efficiency, upgrading industries, and enhancing public services. Intelligent government services, AI-assisted medical consultation, intelligent learning support, and smart elderly care have brought tangible benefits and conveniences arising from technological progress to ordinary people. As a result, Chinese society generally holds a positive and open attitude toward AI. Empirical research by the China Research Institute for Science Popularization, based on a national sample survey of scientific literacy, indicates that Chinese citizens generally hold a positive attitude toward AI and that the technology has a sound foundation of public trust.[①] Thanks to clear policy directions and continuous and stable implementation of relevant measures, China has effectively controlled and reduced the risks and uncertainties arising from AI development, while unlocking and demonstrating AI's substantial potential for empowering industry development and improving people's quality of life. It has provided useful references for the international community and strengthened confidence in global AI governance.

---

① Tang Delong, Wang Zhenmei, Hu Junping, Ren Lei, and Cao Jin, "Analysis of Chinese Citizens' Perception of and Attitude to AI," Study on Science Popularization 20-3 (2025): 41-49, DOI: 10.19293/j.cnki.1673-8357.2025.03.005.

(II) Upholding Openness and Promoting International Cooperation on AI Governance

As technologies advance and acquire broader impact, it becomes more imperative to guide their development with rules and coordinated multinational governance. At this critical stage when AI technologies undergo rapid iteration and governance rules are taking shape, China has actively promoted international cooperation on AI governance in an open spirit. Since 2023, it has successively issued a number of policy documents on this subject. In July 2026, the founding agreement of the World Artificial Intelligence Cooperation Organization, which was proposed by China, was officially signed. China advocates genuine multilateralism and supports the United Nations as the primary channel for international cooperation. It is one of its priorities to help strengthen the capacity of Global South countries and bridge the AI divide, calling for the early establishment of a global AI governance framework based on extensive consensus.

Initiatives for Global AI Governance Proposed by China

No.	Title	Year
1	Global AI Governance Initiative	2023
2	Shanghai Declaration on Global AI Governance	2024
3	AI Capacity-Building Action Plan for Good and for All	2024
4	Global AI Governance Action Plan	2025
5	AI Plus International Cooperation Initiative	2025
6	Action Plan on Artificial Intelligence Cooperation And Development	2026
7	Action Plan on International AI Ethics Governance	2026
8	Global Cooperation Initiative on Agent Mutual Trust, Interconnection and Interoperability	2026
9	World Artificial Intelligence Cooperation Organization	2026

1. Systematically advocating inclusive and equitable AI development

China's philosophy of global AI governance is consistent with its philosophy of

domestic AI governance, sharing the same underlying logic and values. China has put forward a systematic and comprehensive vision for promoting inclusive AI development. In 2023, President Xi Jinping, in the *Global AI Governance Initiative*, called for global collaboration to foster the sound development of AI, share AI knowledge, and make AI technologies available to the public under open-source terms. He also called for increasing the representation and voice of developing countries in global AI governance, and ensuring equal rights, equal opportunities, and equal rules for all countries in AI development and governance. Efforts should be made to conduct international cooperation with and provide assistance to developing countries, to bridge the gap in AI and its governance capacity.

The inclusive AI development China proposes has twofold meanings. First, it calls for ensuring all countries' equal rights and opportunities to access and use AI technologies and share their benefits. China opposes AI becoming a "game for rich countries and wealthy people" and turning technological advantages into a tool for political control. It maintains that digital dividends should not give rise to digital hegemony, and that the AI revolution should not create an AI divide. Second, it emphasizes respect for national sovereignty and calls on countries to engage in governance consultations and cooperation in a spirit of equality and openness. China opposes a small number of countries leveraging their technological advantages to form exclusive blocs and exert control over AI governance through standard-setting, the export of rules, and platform control, thus turning developing countries into their technological dependents and dumping markets. China places particular emphasis on ensuring adequate representation of developing countries in global AI governance.

In July 2024, at the 78th session of the United Nations General Assembly, a China-led resolution on strengthening international cooperation on AI capacity-building was adopted by consensus, with more than 140 countries joining as co-sponsors, demonstrating the broad international support for this vision. China subsequently announced the *AI Capacity-Building Action Plan for Good and for All*, setting out systematic measures in areas such as improving infrastructure connectivity, strengthening talent training, and enhancing data security and diversity, thereby translating the concept of inclusiveness into an actionable framework for cooperation. In July 2025, China launched the *Global AI Governance Action Plan*, calling on all parties to take effective actions in advancing global AI development and governance based on the objectives and principles of promoting AI for good and for the people,

respecting national sovereignty, aligning with development goals, ensuring safety and controllability, upholding fairness and inclusiveness, and fostering open cooperation. The Plan sets out 13 concrete and actionable measures. In September of the same year, at the High-Level Meeting on the Global Development Initiative, China proposed the *AI Plus International Cooperation Initiative*, calling upon all countries to proactively carry out AI+ campaigns in light of their national conditions, enhance policy exchange and practical cooperation, and share best practices and solutions. Efforts should be made to promote deep integration of AI into economic and social development, and ensure that people across the world, including those in the Global South, can broadly benefit from AI. The Initiative identifies five priority areas: "AI+ public well-being", "AI+ technological advancement", "AI+ industrial application", "AI+ cultural prosperity", and "AI+ talent cultivation."

2. Advancing international cooperation on AI through concrete actions

If developing countries lack the necessary technological and institutional capacity, the representativeness and effectiveness of global AI governance will be undermined. To put the principle of inclusiveness into practice, China has continued to pursue international cooperation on AI with developing countries.

In Southeast Asia, China released the *Three-Year Work Plan for the China-ASEAN AI-Driven Sci-Tech Capacity Enhancement Initiative (2025–2027)* in 2025, which explicitly calls for jointly building an open-source and open AI community and promoting open-source models and the sharing of infrastructure. China has worked with ASEAN countries, including Thailand and Laos, to jointly develop data infrastructure and strengthen capacity building through regional platforms such as the China-ASEAN Information Harbor. China has also supported Laos in developing the "Laos National Large Model," providing AI services in Chinese, Lao, and English. In November 2025, Singapore's AI Singapore (AISG) released Qwen-SEA-LION-v4, a Southeast Asian multilingual large model developed on the basis of China's open-source model technology and incorporating more than 100 billion entries of localized data from Southeast Asia. The model has been made freely available worldwide. In May 2026, the Enterprise Alliance of the China-ASEAN Countries AI Application Cooperation Center was formally established in Nanning, Guangxi. Its first group of 55 invited member organizations spans key segments of the AI industry chain, including computing infrastructure, model algorithms, data resources, terminal equipment, digital application

scenarios, and industry funds. Focusing on developing flagship projects, the alliance will advance cross-border application cooperation, industrial ecosystem integration, joint development of industry standards, compliance and risk management support, rights and interests protection, and international market promotion, with the aim of building a sustainable regional AI industry collaboration ecosystem.

In Africa, China has provided targeted support to address challenges faced by African countries, including weak digital infrastructure, limited computing resources, insufficient data, and a shortage of AI professionals. The 2024 FOCAC Beijing Summit proposed the Ten Partnership Actions for Modernization under which the two sides will jointly build a digital technology cooperation center and Chinese companies will build 20 demonstration projects of digital infrastructure and transformation in Africa. Also in 2024, the Egypt Government Data and Cloud Computing Center, constructed by Chinese companies, officially began operation. As the first large-scale AI and big data processing center in North Africa, it provides more than 200 cloud services, including AI platforms, computing power, and local data storage and processing, significantly enhancing local computing capacity and AI processing capabilities.[①] Chinese companies have reshaped the AI application landscape in Africa through open-source models. With input tokens priced at just US$0.27 per million and output tokens at US$1.10 per million, the cost is more than 90% lower than that of leading Western models. The open-source models can be downloaded and fine-tuned by African developers at no cost, strengthening Africa's digital autonomy and local innovation capabilities while significantly reducing the cost of AI application. Chinese companies have also helped African countries develop open-source large models tailored to their national conditions. For example, an Arabic-language large model has been released in Egypt with a speech recognition accuracy rate of 96%, while the "Sunflower" model developed for Uganda serves 46 million local residents. The AIDMAN AI-assisted diagnostic system developed based on Chinese models is helping African countries detect and diagnose malaria and other diseases. In addition, China has been helping African countries cultivate local AI professionals through vocational training programs,

① "African countries accelerate AI applications," https://tradeinservices.mofcom.gov.cn/article/news/gjxw/202503/173378.html, March 4, 2025.

Luban Workshops, international seminars, and other forms of capacity building.[①]

In Latin America, in September 2023, China and Latin American and Caribbean countries issued the *China-CELAC Digital Technology Cooperation Forum Chongqing Initiative*, calling for joint efforts to promote the broad application of digital technologies in areas such as education, healthcare, and agriculture, while advancing the deep integration of digital technologies with the real economy and fostering innovation-driven development. In 2023, China and Brazil jointly developed an AI-powered application capable of translating Brazilian sign language into written and spoken Portuguese and vice versa. According to data from the Brazilian Institute of Geography and Statistics, the application could benefit nearly 2.7 million deaf persons with speech impairment in Brazil who communicate exclusively through sign language.[②] In February 2025, a Chinese company completed in Mexico the construction of its first data center in Latin America, providing cloud computing and AI-related services.[③] Some Chinese companies have also participated in 5G+AI applications in Latin America, covering agricultural monitoring and urban management, significantly promoting the generation and storage of local data.

China has organized AI capacity-building workshops for developing countries, with broad participation from Asia, Africa, Latin America, Eastern Europe, and other regions. In December 2025, the Global South AI Workshop was officially launched, setting up a new platform for countries to deepen cooperation on "AI+ vocational education."

At WAIC 2026, President Xi Jinping announced that in the next five years, China will provide developing countries with 5,000 opportunities in AI training and seminar programs; develop international AI application cooperation centers with ASEAN, the

---

① "AI opens up new space for China-Africa cooperation," https://www.chinaafricasps.org/zh/xxdt/xwdt/663759824379973.shtml, March 20, 2025.

② 《"China, Latin America unlock the potential of digital technology cooperation," http://www.chinacelacforum.org/chn/zgtlgtgx/202408/t20240806_11466825.htm, August 5, 2024.

③ "Latin America: intelligent upgrading of 'Made-in-China'," Globe, https://www.news.cn/globe/20250304/7109565cc4dc42069c81ea3ac37c203e/c.html, March 4, 2025.

League of Arab States, the African Union, the Community of Latin American and Caribbean States, the Shanghai Cooperation Organization, and BRICS; and enable 30 countries to use the AI-powered meteorological warning system, known as MAZU. These measures represent China's solid actions to support global AI development and advance global AI capacity building.

3. Actively participating in the building of global AI governance mechanisms

Supporting the United Nations as the primary channel

China supports the United Nations in playing the primary role in global AI governance and, together with other developing countries, seeks to enhance the fairness and representativeness of AI governance mechanisms. China and Zambia jointly initiated the establishment of the Group of Friends for International Cooperation on AI Capacity Building, providing countries with a regular platform for exchange and experience sharing under the UN framework. Two Chinese experts were selected to serve on the Independent International Scientific Panel on Artificial Intelligence, contributing to the assessment of AI opportunities, risks, and impacts and the preparation of the panel's Preliminary Report. A senior Chinese delegation participated in the inaugural Global Dialogue on AI Governance from beginning to end, presenting China's positions on the meeting's key themes. China also actively participates in AI governance-related activities within platforms including the UN General Assembly, UNESCO, and the ITU, contributing its views and proposals to global AI governance.

Establishing the World Data Organization

Against the backdrop of inadequate international data governance, China has actively promoted global data cooperation and governance. In March 2026, the inaugural assembly of the World Data Organization, themed "Jointly Building a Data Cooperation Platform and Sharing Digital Development Opportunities," was held in Beijing, marking the official establishment of the world's first international organization dedicated to data development and governance. The mission of the organization is to "bridge the data divide, unlock the value of data, and boost the digital economy." It now has more than 200 members from over 40 countries and regions worldwide, including enterprises, universities, think tanks, international organizations, and financial institutions. Through this platform, China will assume greater responsibility

in developing the global data governance system and coordinating international rules, contributing to the improvement of global data governance.

Hosting the WAIC

Since 2018, China has hosted the WAIC annually in Shanghai. The conference initially served as a platform for exchanges on technology, industry, investment, and financing, with only a few side sessions on issues such as AI ethics. In 2024, the event was officially renamed the "World AI Conference and High-Level Meeting on Global AI Governance," with governance becoming a core component alongside industry-themed sessions. The 2025 edition of the conference further highlighted governance under the theme "Global Solidarity in the AI Era." During the event, China proposed establishing the World Artificial Intelligence Cooperation Organization. In 2026, the conference brought together representatives, experts, and members of academia, industry, and research institutions from more than 100 countries and international organizations, and released the *Chair's Statement of the 2026 World Artificial Intelligence Conference and High-Level Meeting on Global AI Governance*, further enhancing its effectiveness in promoting AI governance. The WAIC has now become an important international platform for dialogue on issues including opportunities brought by AI development, risk management and control, ethical norms, and capacity building.

Preparing for the establishment of the World Artificial Intelligence Cooperation Organization

In July 2026, the *Agreement on the Establishment of the World Artificial Intelligence Cooperation Organization* was signed in Shanghai, with 29 countries from Asia, Africa, Latin America, and Europe signing the agreement on site. UN Secretary-General António Guterres and representatives of countries and international organizations attended the signing ceremony. The World Artificial Intelligence Cooperation Organization is the world's first intergovernmental organization dedicated to AI, with its headquarters in Shanghai, China. To date, it has 38 founding member states.

The Organization will focus on three areas. First, it will strengthen international cooperation on AI capacity building. Through policy exchanges, technological cooperation, personnel training, joint research, and other forms of collaboration, it will promote the accessibility of AI technologies and services, continue to unlock

the benefits of AI, and ensure that developing countries can benefit equally from the new wave of intelligent development. Second, through communication and dialogue, the building of policy consensus, and the sharing of best practices, it will serve as a platform for matching supply and demand, promote the deep integration and complementarity of countries' respective strengths and resources, and advance "AI+" initiatives in line with different national circumstances. It will also intensify scientific, technological, and industrial cooperation, encourage the joint development of open-source ecosystems, and deliver a range of practical and accessible cooperation outcomes. Third, it will firmly uphold the purposes and principles of the UN Charter, support the United Nations in strengthening its role, and implement the *Global Digital Compact* and the UN Sustainable Development Goals, with the aim of establishing a fair and equitable global AI governance system.

Establishing the World Artificial Intelligence Cooperation Organization is a major step by China to respond to the calls of the Global South and to unite the international community in actively advancing AI development and governance. It is also part of China's efforts to build a community with a shared future for humanity and implement the Global Governance Initiative, marking an important milestone in the history of AI development. The Organization provides a regular, institutionalized international platform for countries to, on an equal footing, participate in governance dialogue, strengthen policy communication and rules coordination, and deepen technological exchanges and practical cooperation. It is guided by the purposes and principles of the UN Charter, underpinned by the values of openness and inclusiveness, and grounded in the principle of extensive consultation, joint contribution, and shared benefits. The Organization will significantly advance international cooperation on AI and provide a reliable pathway and platform for narrowing the AI divide, expanding access to the benefits of AI, safeguarding technological safety, and enhancing AI governance.

In the AI era, China stands ready to work with all countries to harness diverse sources of wisdom and pursue sound AI governance through open cooperation, making intelligent technologies not only "practical" and "powerful," but also "trustworthy" and "capable of doing good for all." Let us join hands in building a new global AI governance system that is people-centered, inclusive, safe, and controllable.

The English version is a translation from the Chinese version.