

China issues national rules for AI agents with tiered oversight in sensitive sectors

Source copy kept by wheresthe.ai for the story “China issues national rules for AI agents with tiered oversight in sensitive sectors” (<https://wheresthe.ai/d/cn-ai-agents-implementation-opinions-2026-05/>).

Source: <https://rits.shanghai.nyu.edu/ai/china-issues-first-national-policy-framework-dedicated-to-ai-agents>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: e4158ce8f8d5c5bb1be8aae64152cbe2b6a54093231d0de8dec0fa30e89761e3

NYU Shanghai Library > RITS Home > * On May 8, 2026, China’s Cyberspace Administration (CAC), the National Development and Reform Commission, and the Ministry of Industry and Information Technology jointly released the Implementation Opinions on the Standardized Application and Innovative Development of Intelligent Agents — the country’s first dedicated policy framework treating AI agents as a distinct class of system requiring its own governance, rather than as just another application built on top of large language models.

General Audience

Illustration generated by AI What the Document Says

The Implementation Opinions define an AI agent as an “intelligent system capable of autonomous perception, memory, decision-making, interaction, and execution.” That phrasing matters: it pulls agents out of the broader “generative AI” bucket regulated by China’s 2023 generative AI rules and recognizes them as systems whose autonomy creates distinct risks and opportunities.

The document is organized around four pillars:

Foundations: stronger base models, complete agent tool chains (development, testing, deployment, maintenance), and a national standards system covering interfaces, data exchange, safety assurance, and trustworthiness certification.

Safety and security: behavior containment technology, algorithmic governance, supply chain protections, and frameworks for assessing risks like data poisoning, privacy breaches, and system failures.

Application-driven adoption: 19 priority scenarios spanning scientific research, smart manufacturing, transportation, agriculture, financial risk control, healthcare, education, government services, judicial assistance, and public safety.

Innovation ecosystem: open-source frameworks, compatibility with domestic chips and operating systems, industrial collaboration platforms, and active participation in international standards-setting.

Human Oversight and Decision Boundaries

One of the most concrete provisions concerns who gets to decide what. The guidelines require developers to “clarify the reasonable boundaries and required authority for various decision-making methods” and distinguish three tiers: decisions limited to the user, decisions requiring user authorization, and autonomous decisions by the agent itself.

Crucially, the document states that users “have the right to know and the final decision-making power regarding the autonomous decisions made by the intelligent agent, and that the intelligent agent’s actions do not exceed the scope authorized by the user.” This is functionally similar to European discussions of “meaningful human control,” but framed around practical deployment rather than precaution.

Image credit: Cyberspace Administration of China Tiered Governance

The framework adopts a risk-tiered approach. Agents in sensitive sectors and key industries — healthcare, transportation, media, public safety — will face filing requirements, mandatory testing, product recalls, and oversight by both cyberspace regulators and sector-specific authorities. Lower-risk consumer scenarios are expected to rely more heavily on platform governance, third-party evaluation, and industry self-regulation, supported by a credit evaluation system that can penalize violators.

The document also calls out two specific harm vectors: anthropomorphism-driven dependence among minors and

elderly users, and misuse of agents in automated attacks, privacy violations, and fraud schemes.

What This Means

The release marks a notable philosophical contrast with much of the Western debate around agentic AI. Where U.S. and U.K. discussions have leaned heavily on catastrophic loss-of-control scenarios, the CAC document focuses on practical integration into existing institutions — and argues that real-world constraints like compute quotas, credit ceilings, access permissions, and system shutdowns naturally bound agent autonomy. Analysts have summarized the posture as “deploy first, govern along the way.”

The emphasis on indigenous controllability is also strategic. By tying the policy to domestic chips, operating systems, and open-source frameworks — and by signaling intent to “actively participate in international standards-setting” for agent protocols — China is positioning itself to shape, not just follow, the global rules of the road for autonomous AI systems.

For researchers, developers, and institutions working on agentic AI, the document is worth reading as a concrete preview of how a major jurisdiction plans to handle decision-boundary disclosures, registration requirements, and sectoral filings as agents move from demos into production.

Related Coverage

[Agents of Chaos: What Happens When Autonomous AI Agents Get Real Tools](#) — a red-teaming study that motivates much of the safety-baseline language now appearing in agent policy.

[AI Accountability in 2026: State Laws Take Effect as Federal Proposals Compete](#) — for a contrast with the U.S. patchwork of state-level AI regulation.

[China Issues New Regulations on Generative AI](#) — the 2023 generative AI rules that the new agent framework builds on.

[List of Legal AI Models in China](#) — earlier CAC publication on registered models.

This post was drafted with AI assistance and reviewed by RITS staff.

Sources

Cyberspace Administration of China — 智能体规范应用与创新发展实施意见 (original document, May 8, 2026)

Xinhua — China unveils guidelines to regulate, boost innovative development of AI agents

People’s Daily Online — China unveils guidelines to promote development of AI agents

Geopolitechs — China’s First Policy Framework for AI Agents (analysis)

The Register — China’s agentic AI policy wants to keep humans in the loop