
Your
China
Strategy.
Chasing Artificial
General Intelligence:
China Between Breakthroughs
and Bottlenecks
sinolytics

Table of Contents.

Contact & Main Author:
Mengying Tao,
Senior Consultant
mengying.tao@sinolytics.de

Contributing Authors & Researchers:
Bowen Han
Camille Chabot
Jingwen Tong
Shengyi Zhang
Sishi Xie
Weiqing Zhu

sinolytics

#1

Introduction

AGI (Artificial General Intelligence), simply put, is AI with human-level cognitive capabilities. It is capable of learning, reasoning, and solving problems in various domains like a human. So far, AGI remains a theoretical concept, with estimates of when it will be achieved ranging from as early as 2026 to many decades later. The pursuit of AGI is a great aspiration for many in the global AI community. The technological competition between the U.S. and China is partly reflected in the progress towards AGI. While the U.S. remains the leader, China has emerged as a serious challenger.

Several leading Chinese AI companies have openly declared AGI as their goal, although their understanding of the path to AGI differs. The Chinese central government has not formally made AGI a policy goal, but AGI theory is listed as a topic for exploratory research in the country's 2017 overarching AI development plan. The central Politburo meeting in April 2023, chaired by Xi Jinping, highlighted the importance of AGI development for the first time. As a result, local governments such as Guangdong and Anhui have issued policy guidelines aimed at building AGI.

This whitepaper presents AGI as an ecosystem of interconnected elements — chips, computing, data, models, and applications — and assesses China's strengths and weaknesses in each of these areas. It focuses on the dominant technological path to AGI, namely Large Models, while intentionally leaving out broader debates about AGI risks. Based on research completed in March 2025, the paper aims to guide multinational companies in understanding China's AI ecosystem, highlighting both opportunities and challenges.

At a Glance: China's Pursuit of AGI

Chips	AI chips represent China's biggest vulnerability in the AGI race, constraining the efficient training of complex models. The synergy between state-led industrial policies and strong business drives offers China a chance to make some meaningful progress in chipmaking technologies.
Computing	China has laid out an ambitious blueprint for building a national computing networking coordinated with energy resources, but technical hurdles and institutional inefficiencies remain key obstacles to effective implementation.
Data	China possesses vast data resources but struggles to obtain high-quality training datasets, limiting the reasoning capabilities and accuracy of its AI models. Government-led initiatives to develop and open access to high-quality data could unlock China's data potential.
Models	Dynamic competition has driven model iteration, with disruptive players like DeepSeek leading breakthroughs that could expedite progress toward AGI. Its open-source push has potential to reshape global competition and expand China's AI influence.
Applications	China demonstrates strong application strengths that could be leveraged to overtake competitors on the curve. It looks to replicate the killer app playbook in the AI era and expand from digital AI agent to embodied intelligence.

#2

AI Chips – Most Pressing Vulnerability

Specialized chips such as GPUs, TPUs, FPGAs, and ASICs form the computational backbone of modern AI. By optimizing parallel processing and energy efficiency, these chips unlock the massive computing power needed to train trillion-parameter models. Without advanced AI chips, even the most sophisticated algorithms would struggle to scale to the levels required for AGI.

AI chips lie in the center of U.S. export controls, and U.S. effort to close loopholes further restricts China's access to advanced chips.

U.S. export controls began blocking Huawei's access to TSMC's advanced manufacturing processes in 2020. Fast forward to 2024, a TSMC chip found in Huawei's 910B led to the latest expansion of controls. Foundries like TSMC and Samsung are now held accountable for verifying that their chips are not being diverted to Chinese companies. Baidu and many other Chinese AI chip developers, while not on the U.S. Entity List, now face strict approval processes to gain access to 7nm technology.

Shifting to domestic foundries also poses significant challenges, particularly due to escalating restrictions on access to advanced chip manufacturing equipment since 2019. Huawei's reliance on SMIC's deep ultraviolet (DUV) lithography systems to produce its 7nm Ascend series underscores this struggle. The low yield rate, once estimated at 20% for the most advanced 910C, had caused Huawei to temporarily halt production and hindered

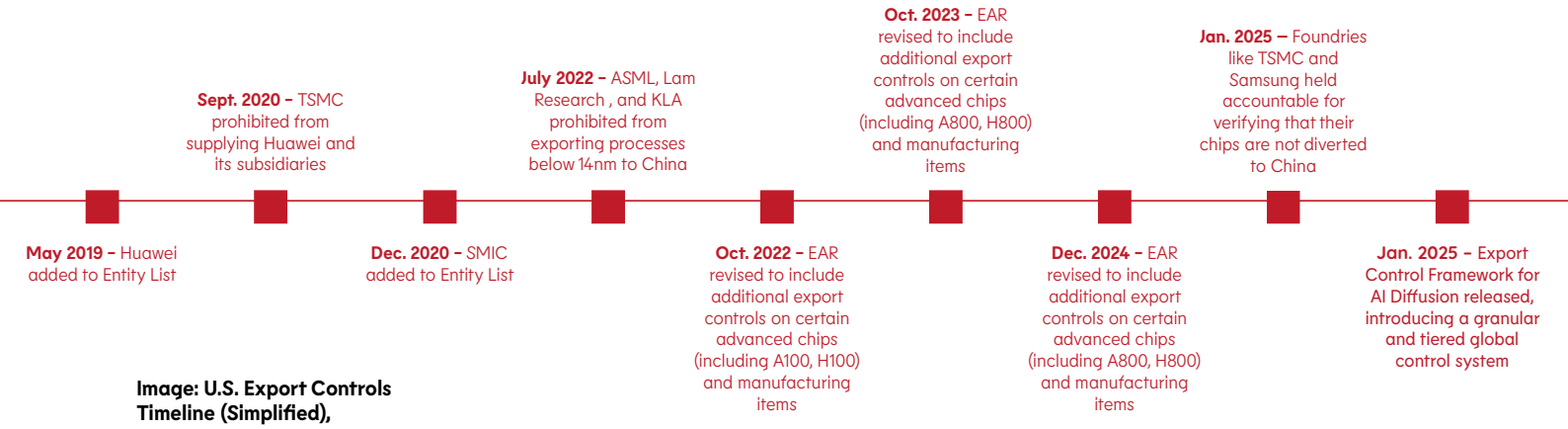


Image: U.S. Export Controls Timeline (Simplified), Source: Sinolytics Research

shipments to customers. However, the latest report indicates that Huawei has improved 910C yield rate to nearly 40%, making the Ascend line profitable for the first time.

China's exploitation of loopholes and circumvention of existing restrictions has prompted the U.S. government to reassess the effectiveness of its control measures. The Export Control Framework for AI Diffusion was unveiled along those lines. Taking a granular and tiered approach, the Framework establishes a global licensing system for AI chips and frontier model weights.

If rigorously enforced, the new rules will effectively crack down on covert purchases of advanced AI chips via third countries and block Chinese AI developers' access to overseas cloud services. In the longer term, unless China achieves disruptive breakthroughs, global AI computing power will become increasingly concentrated in the United States and its allies, forcing other nations to move away from China's sphere of influence and align themselves with U.S. rules and standards.

China's domestic AI chip sector has grown but is still struggling to fill the void left by the banned NVIDIA chips.

With companies running out of chips they had stockpiled in advance of export controls — and with fewer loopholes left to exploit — self-reliance and indigenous innovation are emerging as the only viable options. China established "indigenous innovation" (自主创新) as a strategic policy goal in 2005 and accelerated its pursuit of technological self-reliance with "Made in China 2025" (中国制造) in 2015, well before U.S. export controls were tightened. However, the urgency of this mission has never been more existential than it is today.

China's AI chip sector has grown rapidly in recent years, fueled by a surge in demand for training large-scale models. According to IDC, domestic chips accounted for about 20% of the Chinese market in the first half of 2024 — up from less than 15% in 2022 — driven largely by Huawei's Ascend series. Despite this progress, NVIDIA GPUs continue to dominate, and strong sales of downgraded H20 chips suggest that domestic alternatives are still struggling to replace banned high-end foreign chips, even after heavy investment.

Huawei's Ascend 910 series is seen as the most viable domestic competitor to NVIDIA. According to insiders cited by media outlets, the 910B is reportedly used by companies like ByteDance and Baidu to train LLMs. The more advanced 910C, said to rival NVIDIA' s H200, is expected to enter mass production in 2025. Meanwhile, Baidu has deployed its 7nm Kunlun 2.0 chips for model inference and launched a compute cluster with 100,000 Kunlun 3.0 chips in early 2025. According to media reports, ByteDance is collaborating with Broadcom on the development of a 5nm AI chip. Neither company has publicly confirmed the partnership to date.

Beyond big tech, smaller GPU developers such as Cambricon, Biren, and Enflame are advancing AI chips, with deployments in major cloud platforms and mass production underway. However, the product iteration of leading GPU developers has slowed down after the U.S. expanded control measures in 2022. The latest generation of AI chips from Huawei (910C), Baidu (Kunlun 3), Tencent (Zixiao V2), and Cambricon (MLU 590) remain officially unreleased, fueling widespread market speculation about their performance capabilities and shipping schedules.

China's state-led, investment-driven approach has yet to invigorate the chip sector, but this may change as businesses are increasingly motivated to join the state in pursuit of self-sufficiency.

China pursues a state-led, investment-driven approach to semiconductor development. The state continues to pour money into the chip sector through investment funds and financial support at both the central and local levels. At the central level, the third phase of the China IC Industry Investment Fund ("Big Fund III") has exceeded the combined size of the previous two phases. Local governments have also set up their own funds to support local industries and cultivate regional champions.

A wide range of financial incentives have been introduced to attract investment in the semiconductor industry and stimulate R&D activities. However, unlike its success in other industries, such as EVs, this strategy has yet to deliver significant breakthroughs in the chip sector.

What marks a shift this time is the close alignment of commercial interests with national strategic priorities. As the value of imported chips declines, Chinese companies are now proactively seeking domestic substitutes, whereas in the past they were reluctant to use domestic chips when foreign options were still available and easier to use.

By testing domestic chips and using them in their systems, users provide critical feedback on how to improve hardware design and capabilities, thereby accelerating the pace of development. DeepSeek has sped up the process. Dozens of hardware and computer companies in China have launched "DeepSeek all-in-one" devices powered by Huawei, Hygon, Cambricon and other domestic processors. The collaboration between domestic AI chips and large-scale models is expected to accelerate the adoption of chips in AI training and inference, enabling continuous optimization through real-world use.

What does it mean?

AI chips represent China's most pressing vulnerability in the AGI race. It is also a strategic chokepoint targeted by geopolitical rivals. China's most advanced AI chips still lag behind NVIDIA's mainstream offerings, and the gap in the chip software ecosystem is even wider. The limited capabilities of advanced chips pose a significant risk, potentially limiting innovative technology pathways and hindering China's progress toward AGI.

China can leverage its strong engineering capabilities to maximize the use of available chips, but original breakthroughs in chip manufacturing are essential to overcome its strategic disadvantage in the current U.S.-China AGI race. The synergy between government-led industrial policies and strong business initiatives provides an opportunity for meaningful progress in overcoming chip bottlenecks. In the short term, however, China's chip vulnerabilities are likely to persist, limiting its AGI ambitions.

China's restricted access to advanced AI chips poses major challenges for multinational companies aiming to deploy AI products in the country. They have to either rely on lower-performance NVIDIA chips or adapt to domestic alternatives and develop a China-specific software stack, hindering the development and rollout of AI-powered applications and services.

#3

AI Computing – Brilliant Design, Challenging Execution

Scaling laws suggest that model performance predictably improves as computational investment increases. The experience of AI companies so far confirms this trend, as they race to build superclusters of hundreds of thousands of GPUs to train the next bigger and more powerful model that will bring us closer to AGI

As computing power expands, models can be trained on more parameters at higher speeds, increasing their ability to reason and handle complex tasks. In addition to sheer scale (measured in FLOPS), computational efficiency is equally important, ensuring more cost-effective training through algorithmic optimizations.

China is coordinating national computing resources through a top-down strategy, with the public sector and internet giants investing heavily in AI computing centers.

China aims to boost its national computing capacity to over 300 EFLOPS by 2025 — a 50% increase from 2023. The goal is for 35% of this capacity to be dedicated to AI, enough to power large-scale model development using a standard format (FP32) commonly used in AI training. To coordinate these resources nationwide, the government is pursuing a top-down strategy through the Eastern Data, Western Computing (东数西算) project. Launched in early 2021 as part of President Xi Jinping's "Digital China" agenda, the project aims to build a national integrated computing power network anchored by eight regional hubs and ten data center clusters. The strategy is to develop data centers in the resource-rich western regions-where land and renewable energy are more abundant to support the growing computing needs of the more economically developed eastern areas.

Source: NDRC; State Information Center; MIIT; Sinolytics analysis

sinolytics

By mid-2024, the government had already spent CNY 43.5 billion (~ USD 6.1 billion) on the national project. Direct government funding has kick-started the construction of the new computing hubs and spurred a much larger amount of SOE investment and private funding. China's three major state-owned telecom operators are showing strong momentum in building AI data centers, closely aligning with national strategies and prioritizing domestic technologies. In 2024, China Mobile invested CNY 21.6 billion (~ USD 2.96 billion) in two data centers equipped with 9,200 Huawei Ascend servers.

At the same time, China's tech giants are making their own multi-billion-yuan bets on AI infrastructure. Alibaba reported a 13% year-over-year growth in Alibaba Cloud revenue in the fourth quarter of 2024, driven by public cloud demand and AI adoption. CEO Eddie Wu highlighted AGI as a strategic priority, announcing plans to invest more in cloud and AI infrastructure over the next three years than in the entire previous decade. ByteDance has emerged as one of the most aggressive players. The company has reportedly earmarked CNY 40 billion (USD 5.5 billion) for AI chip purchases in 2025 and plans even larger investments in leasing or building overseas data centers.

More than two years after the "Eastern Data, Western Computing" project, the size and distribution of data centers nationwide has improved. However, the newly built data centers in the western regions are struggling to provide efficient computing services and are functioning more as data storage centers. To achieve the original goal of the project, China still needs to solve bottlenecks in network latency, computing power scheduling, high communication costs, etc.

China may encounter a new wave of computing power demand, driven by the rising demand for inference workloads due to the growing adoption of AI.

In 2023, the rise of large models triggered an explosion in AI computing demand, leading to a period of shortage. However, by 2024, heavy investment boosted compute supply, and the model market consolidated, even leading to idle capacity in some data centers. Rising demand for inference computing — driven by growing commercial adoption — is expected to trigger the next wave of compute demand. Inference now dominates AI workloads, accounting for an estimated 80-90% of total machine learning cloud demand. The release of DeepSeek-R1 has shown that advanced models can be trained with far fewer compute resources, which may temporarily ease pressure on compute infrastructure. However, such innovations are likely to accelerate the adoption of AI, ultimately driving up overall demand. China sees this shift from training to inference as a key opportunity for domestic chipmakers, as inference chips have lower technical barriers compared to NVIDIA's high-performance training chips — making faster deployment of domestic alternatives more feasible.

China's 14 national supercomputing centers also play a key role in meeting the growing demand for AI computing, supporting research, industry, and government projects with powerful systems such as Tianhe, Sunway TaihuLight, and Shenwei. By integrating high-performance computing with AI cloud platforms, these centers enable scalable AI development in fields such as healthcare, smart cities, industrial design, and climate modeling, without requiring companies to build their own infrastructure.

China's data centers have enough power for now, but the government is promoting efficiency and green energy to prepare for future shortages.

Data center energy challenges made headlines in 2024, when the Memphis power grid couldn't support Elon Musk's AI supercluster — prompting the deployment of mobile natural gas turbines. Major U.S. tech companies are also investing in nuclear reactors to power their data centers.

Meanwhile, China's data center development faces fewer energy constraints, as the sector is only the ninth largest energy consumer. Despite an overall state of surplus in electricity supply, some eastern cities with high computing demand may face power shortages, prompting the launch of the "Eastern Data Western Computing" project to address these challenges.

Experts of the Beijing Institute of Technology estimate that data center power consumption will grow at an annual rate of about 20% over the next five years, far outpacing China's overall power consumption growth. Its share of national electricity consumption is expected to rise from 2.4% in 2025 to 4.8% in 2030. In response, the Chinese government is proactively coordinating computing power development and energy planning, with policies focused on reducing consumption and expanding the use of green energy.

PUE (Power Usage Effectiveness) is a measure of how efficiently a data center uses energy. Chinese regulations require large and ultra-large data centers to achieve a PUE of 1.25 or lower, meaning 80% of the energy consumed by the data center should go into computing functions and the remaining 20% is used for infrastructure support. National hub data centers must achieve an even lower PUE of 1.2, meaning they are even more energy-efficient. To improve energy efficiency, China has set technical standards for servers, cooling systems, network equipment and power supplies. In terms of green energy, the March policy update raised the green energy consumption target for newly built national hub data centers beyond the previous baseline of 80%. In regions with abundant renewable resources, many new data centers have incorporated photovoltaic systems and energy storage. While nuclear-powered data centers are emerging in the U.S., China has yet to develop similar projects, although state-owned enterprises such as Qinshan Nuclear Power are exploring the viabilities of nuclear power solutions for data centers.

What does it mean?

China has taken a top-down approach to computing infrastructure development, using policy guidance and financial support to build a nationwide computing network similar to the power grid. This model contrasts with the U.S., where large private companies are leading the decentralized expansion of computing. China hopes that such a computing power network can achieve planned distribution of computing resources, cross-regional scheduling, and coordination with energy resources. However, this project faces significant challenges beyond the limitations of advanced chip hardware. Inadequate network transmission capacity and limited computing power scheduling technologies have hindered progress. In addition, as with other centrally planned projects, local governments tend to invest for political gain, resulting in data centers in western regions that are poorly aligned with market demand and remain underutilized.

Overall, China has laid out an ambitious blueprint for its computing infrastructure, with a well-defined vision and strategic planning. However, technical hurdles and institutional inefficiencies remain key barriers to effective implementation. For multinational companies, their China operations are able to access stable and cost-effective computing services, despite constrained performance due to hardware limitations. When evaluating local computing solutions, multinational companies need to assess not only the performance of service providers' compute resources, but also their infrastructure footprint, technical capabilities, and market experience.

#4

Data – Unlocking Great Potential

Data quantity and quality can directly affect the performance of large models. The scaling law shows that increasing the scale of training data can improve model performance. Massive datasets allow AI models to learn diverse real-world patterns and understand contextual nuances, enabling them to develop inference capabilities and provide more accurate responses.

However, performance gains diminish as training data increases, making data quality critical. High-quality data with high accuracy, diversity, and minimal bias enables models to learn reliable knowledge and develop advanced capabilities such as reasoning and fact-checking.

A data bottleneck, where the availability of large, high-quality, multimodal datasets is limited, can significantly impede progress toward AGI. Overcoming data limitations requires innovative approaches, including synthetic data technologies, algorithmic breakthroughs, and hardware improvements.

Despite the huge amount of data generated, the datasets available for training AI models remain limited in both scale and quality.

China's vast real economy, integrated supply chains, and extensive government infrastructure generate enormous amounts of corporate, industrial, and public data. Its 1.4 billion people also produce significant amounts of internet data — by 2025, China is expected to generate 48.6 zettabytes, making it the world's largest data producer. However, the amount of data available for training AI models remains limited, both in terms of volume and quality. For example, Chinese-language data accounts for only 5.4% of Common Crawl, compared to 43.9% for English, and open-source Chinese training datasets are still insufficient.

Despite a huge Internet user base, Chinese users predominantly rely on multiple mobile apps (such as WeChat, Douyin, Taobao), where content is restricted to individual platforms and not publicly accessible, resulting in a closed and fragmented data landscape. Although various levels of government in China control vast amounts of public data, the country lacks a comprehensive open data law like those in the U.S. and EU, resulting in limited data transparency and accessibility. In addition, a significant number of Chinese books, journals, and academic papers remain in physical form, making it difficult to fully leverage offline data. Companies in vertical industries are often reluctant to open or share data due to concerns about commercial interests and data security.

Investing in higher quality data can lead to greater improvements than simply increasing the volume of data. Chinese AI companies are widely aware of the scarcity of high-quality Chinese language corpora, which manifests itself in several ways: First, a lack of content that reflects Chinese values. While machine translation can address factual gaps in datasets, it falls short in capturing the cultural semantics, implicit value systems, and idiomatic expressions embedded in Chinese language. Second, China's data annotation sector is still developing, with a strong need for annotators who combine model training expertise with deep domain knowledge. Third, the lack of standardized data formats and protocols further hinders the effective use of data.

China is torn between "regulation for security" and "openness for innovation", but providing high quality data for model training is currently a higher priority.

China's approach to data governance follows the Party's strategic thinking of "grasping firmly with both hands," which means the pursuit of two goals with equal determination and vigor. On the one hand, the government enforces strict laws and regulations to protect national security and ensure data privacy. On the other hand, it is promoting data as a new production factor to drive economic growth and innovation. Policymakers aim to achieve the ideal of "secure and free flow of data", but in practice this has proven to be a difficult balance, creating ambiguity and obstacles for businesses.

This dual-goal approach is also evident in the governance of GenAI training data. China is the first country in the world to introduce regulations for GenAI training data and has quickly established a national approval system. Recognizing the limitations of data access for AI development, the state subsequently launched top-down policies to encourage the opening of public and corporate datasets and facilitate data transactions for model training.

A three-year plan to unlock the potential of data was introduced in 2023, and one of its focuses is the development of high-quality corpora and scientific datasets for AI model training. The central government also called for better use of both public and corporate data, urging ministries to start releasing high-quality public datasets and encouraging companies to share more data for AI development. Local governments followed with more detailed incentives, such as Shenzhen offering up to CNY 1 million in subsidies for companies that open corpora and up to CNY 2 million for corpora buyers in data exchanges.

China's data trading market continues to expand, with many data-rich cities setting up government-backed data trading centers and making efforts to build and open datasets. It is also reported by state media that China is in the process of establishing a new state-owned enterprise called the National Data Group, which will be responsible for integrating, operating, and sharing data resources across key sectors and regions.

The Party's ideological control restricts training data, narrows the scope of generated content, and impairs the ability to reason about models.

Ideological risks are at the heart of China's regulatory framework for Gen AI. Standards such as the "Basic Safety Requirements for Gen AI Services" meticulously outline 31 types of content risks related to training data and generated content, and require model developers to build a keyword library and question bank for effective content moderation and risk mitigation. The first and foremost risk group is the violation of socialist core values, such as inciting subversion of state power, endangering national security, damaging national image, etc.

This regulatory approach is consistent with how China has been controlling internet information since 2011. Largely due to censorship, the mass removal of websites, accounts, and user-generated content has sparked fears of "losing collective memory." In terms of AI models, this has led to a lack of high-quality training data that reflects diverse values, limiting the performance and reasoning capabilities of models. Due to limited exposure to diverse perspectives and critical analysis, models may not be able to provide nuanced and in-depth discussions on certain issues.

China has also implemented an approval system for the market entry of GenAI models. To obtain approval, model providers must go through a lengthy and complicated process that includes the preparation of detailed safety reports and verification of a test account by auditors.

Models must undergo a review of both their corpora and generated content to ensure that they do not contain sensitive or illegal data. They are tested on their ability to effectively filter out sensitive keywords during content generation and refuse to answer sensitive questions. The test is based on an interception keyword list of at least 10,000 keywords and a question bank of over 500 questions that models should refuse to answer. Even after approval, model providers must continually update the keyword list and question bank to maintain compliance. Such an approval system increases compliance costs for companies and could potentially slow down research and iteration.

Ideological restrictions could discourage other countries from adopting Chinese technology. The political bias of Chinese models is already a contentious issue in AI discussions outside of China. Such restrictions could further hinder the interoperability of Chinese models with global AI systems, deepening the fragmentation of the AI ecosystem.

What does it mean?

China's data advantages are often overstated. While the Chinese language domain has generated massive amounts of data, high-quality datasets for model training remain scarce. Key constraints include insufficient access to public data, data moats created by mobile platforms, limited digitization of physical texts, and government content removal practices. Chinese AI companies see this scarcity as a critical bottleneck that hinders model progress.

The Chinese government recognizes the data challenge and has initiated policy moves to build and open high-quality datasets. The non-governmental community is learning from the U.S. experience to consolidate and preserve Chinese-language corpora, while AI companies are investing in synthetic data technologies and algorithmic innovations to mitigate data constraints. Given China's proven ability to mobilize resources centrally, its potential to leverage its unique data advantages could accelerate its path to AGI.

Multinational companies launching public-facing AI products in China need to navigate through complex data compliance requirements, including rules on training data, generated content, and mandatory filings with the Cyberspace Administration. If cross-border data transfers are involved, it could trigger additional compliance steps like a security assessment for important data. These raise costs and market entry barriers, making it essential for MNCs to plan ahead with a clear compliance strategy.

#5

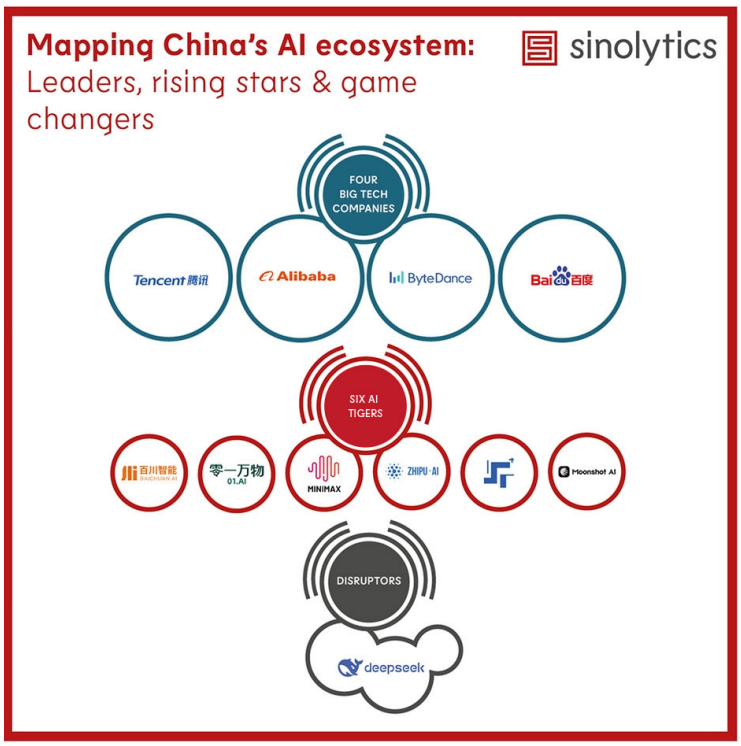
Models – Closing in Fast

Many argue that current large models are essentially data-driven systems ("token generators") that lack core AGI features such as cross-domain reasoning and autonomous consciousness. Yet they continue to dominate research and may represent the most viable path to AGI. Iterative improvements have transformed early LLMs, once limited to simple questions and answers, into models capable of multistep reasoning and complex tasks. Each iteration improves performance, validates AGI approaches, and brings models closer to human-like cognition.

From the "Hundred Model Battle" to a "survival of the fittest" dynamic, China's fierce competition has rapidly driven model iteration and advancement.

China's "Hundred Model Battle" began in 2023, marking fierce competition in the field of GenAI models. Major tech companies, AI startups, and research institutions joined the battle, driven by huge market potential and government policy support. By July 2024, China accounted for 36% of the world's LLM, with a total of 1,328 models worldwide. This ranks China second in AI model development after the United States.

After more than a year of industry consolidation, the landscape of China's large-scale AI sector has become more defined, with the exit of underperforming models and the survival of the fittest. Six unicorns — MiniMax, Zhipu AI, Baichuan Intelligence, Moonshot AI, StepFun, and 01.AI — have each surpassed USD 1 billion in valuation, earning the title of "China's Six AI Tigers." Alongside tech giants like ByteDance, Alibaba, Baidu, and Tencent, as well as emerging disruptors like DeepSeek, the number of dominant players in the industry has shrunk to about a dozen and is expected to decline further



Source: Sinolytics research

For AI startups, the founder/CEO's vision often dictates the company's technological path toward AGI. Technological innovation and model iteration are widely regarded as the key. While most players remain committed to the scaling law and large-scale model training, some voiced concerns over its limitations. DeepSeek sets itself apart from other startups by prioritizing technological research over commercialization. Led by a cash-rich CEO with a passion for AGI, its team of young talents is fully dedicated to advancing models through original innovation.

Chinese AI leaders' views on model iteration and AGI

Leading AI model companies in China

DeepSeek is the most disruptive Chinese model to date, but it is not the only Chinese model keeping up the chase.

In the highly dynamic market of AI models, companies are chasing each other to release even better, faster, and cheaper models. While OpenAI remains the industry leader, Chinese companies have shown significant competitive strength, with some models outperforming OpenAI models in certain tasks and capabilities. Chinese AI companies are also catching up with OpenAI's models at a faster rate, with DeepSeek replicating o1 within a few months. Stanford University's 2025 AI Index Report shows that the performance gap between Chinese and US models has narrowed from 17.5% in 2023 to 0.3 by the end of 2024.

Alibaba's Qwen Series. The Qwen series was developed by Alibaba Cloud to process both natural language and multimodal data. It is a large family with more than hundreds of models of different sizes and for different applications. Alibaba has open-sourced more than 100 models so far. It is one of the Chinese companies that bet on open-source models to contribute to the global technology ecology. As part of its business strategy, it relies on open-source models to expand the customer base for its cloud business. The company also offers proprietary models, which tend to be more advanced than open-source models. Its latest proprietary model, Qwen2.5-Max, an MoE model pre-trained on 2 tn tokens, achieved performance comparable to the world's best models and ranked particularly high in technical domains such as math and coding on Chatbot Arena.

DeepSeek V3 and R1. DeepSeek's V3 and later R1 models have made rapid breakthroughs in AI, challenging U.S. leadership. Built by local talent, the models reduce training costs and offer competitive pricing. The company was one of five in China to purchase more than 10,000 NVIDIA A100 GPUs ahead of U.S. restrictions, originally intended for quantitative trading. Its app also became the fastest growing in the world, reaching over 100 million users in just 20 days. Its paradigm-shifting impact lies in innovations in model architecture, training methods, and reasoning efficiency that significantly reduce the cost of computing. This is expected to drive AI adoption and long-term growth in computing demand, while fueling optimism for an evolving open-source AI ecosystem. These advances mark a critical step in China's pursuit of AGI.

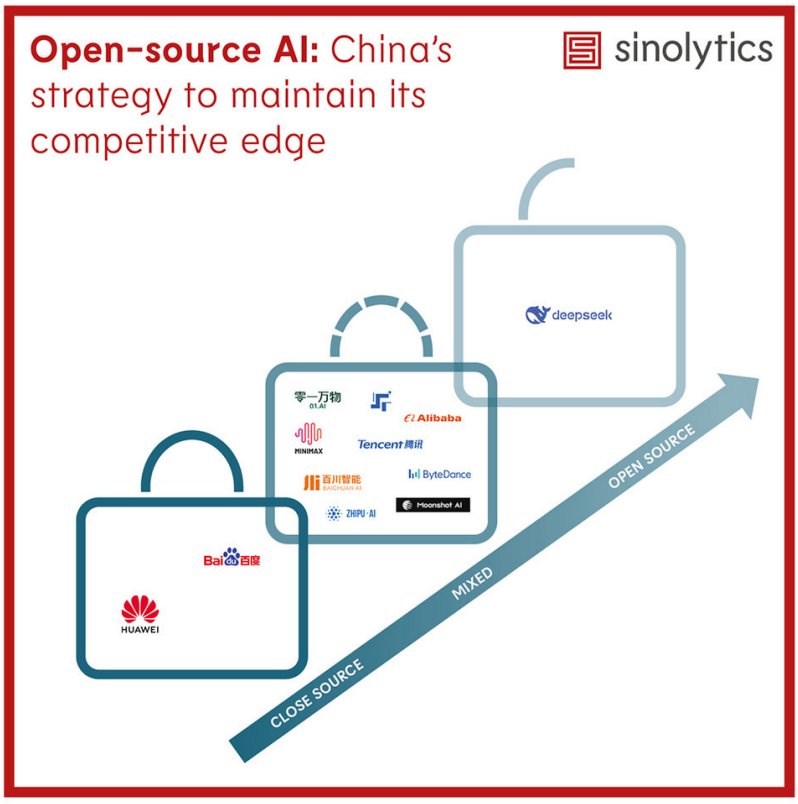
ByteDance's Doubao Series. The Doubao Series powers over 50 ByteDance applications, including Douyin (TikTok), Feishu (Lark Suite), and Jianying (CapCut), leveraging Douyin's massive video data for its industry-leading multimodal capabilities. Doubao-1.5-Pro, the latest release, uses a Mixture of Experts (MoE) approach to improve efficiency by selecting the most relevant parts of the model for tasks. Bytedance claims that it outperforms GPT-4o and Claude 3.5 Sonnet in benchmark tests. ByteDance emphasizes the model's efficient use of computing resources and clarifies that it didn't use data from other models during training, likely in response to earlier speculation. In 2023, OpenAI suspended ByteDance's account for using its API to develop its model.

Moonshot's K Series. Moonshot, an AI startup founded by a group of Tsinghua graduates, was the most highly valued (at USD 3.3 billion) AI startup in China before DeepSeek shook up the AI scene. Its flagship product, Kimi Chat, is a popular AI chatbot in China with over 36 million monthly active users. The company's k-series consists of reinforcement learning (RL) models, including k0 math, k1 visual reasoning model, and the latest k1.5, an o1-level multimodal model. Although overshadowed by DeepSeek V3 and R1 upon its release, k1.5 made similar algorithmic innovations and remains a top-tier model. It excels in long context processing, deep reasoning, and multimodal capabilities, with its Long-CoT (Chain of Thought) performance rivaling OpenAI o1 across multiple modalities. K1.5 was mentioned alongside DeepSeek for independent CoT research in an OpenAI paper on large reasoning models published in February 2025.

The pursuit of open-source AI, led by DeepSeek, will change the competitive dynamic and expand the scope of China's AI influence.

There has been a heated debate among Chinese tech companies about the merits of open-source versus closed-source models. Baidu initially insisted on a closed-source approach, with CEO Robin Li remarking that "open-source models will increasingly be left behind." Alibaba has adopted a hybrid strategy, combining both open-source and closed-source approaches, while being very active on the open-source front, releasing more than 100 models at once in September 2024. Moonshot believes that the closed-source model is the only way to achieve AGI.

DeepSeek is currently the only Chinese AI company fully committed to open-source. Founder Liang Wenfeng noted that the moats created by closed-source models are temporary, and that even OpenAI's closed-source approach cannot prevent others from catching up. He emphasizes that the company's true value lies in its team, whose growth and accumulated expertise foster an innovative culture. Earlier this year, DeepSeek open-sourced the weights of its V3 and R1 models, demonstrating that open-source models are rapidly closing the gap with their closed-source counterparts. This commitment was further reinforced by the release of several AI infrastructure-related code repositories. DeepSeek's open-source initiative has influenced the Chinese AI community, prompting startups such as Minimax, Stepfun, and Zhipu to open-source their latest models. Even Baidu announced plans to open-source its next-generation ERNIE model in May, and Moonshot unexpectedly released its lightweight Moonlight model.



Source: Sinolytics research

There is a clear trend that AI companies are shifting from a closed-source strategy to a hybrid strategy that combines both open-source and closed-source approaches. They showcase technical capabilities through open-source offerings to attract users and developers, while pursuing commercialization through closed-source solutions that provide differentiated, customized services in vertical domains or niche markets.

The Chinese government has been consistently supportive of open-source, mentioning the promotion of open-source technology and ecosystem development in several central policies. China hopes that more companies like DeepSeek will emerge and take the lead in fostering China's own open-source AI ecosystem. Such an ecosystem can not only accelerate innovation toward AGI but also expand China's sphere of influence in other countries, especially developing countries.

What does it mean?

In the rapidly evolving Large Language Model competition, every day brings new surprises. U.S. companies continue to lead, while Chinese model companies, especially the leading ones, although having a strong focus on application layers, maintain the AGI narratives and are catching up with their US competitors at an accelerating pace.

The dynamic competition in China's model scene has proven to be helpful for model iteration and advancement. Disruptors like DeepSeek have challenged the perception that China is merely replicating U.S. models. Through algorithmic innovation, engineering optimization, and open-source collaboration, DeepSeek brought new technological methods that could accelerate the pursuit of AGI.

DeepSeek also brought hope for China to build a participatory open-source AI ecosystem. By harnessing the collective intelligence of global developers, China aims to accelerate technological innovation and model iteration and avoid redundant efforts, potentially accelerating progress toward AGI. This contrasts with the U.S.'s predominantly closed-source approach, adding a fascinating dimension to the U.S.-China AI competition.

The intense competition and rapid development of China's AI model market offer multinational companies a wide range of potential local partners for deploying AI products. These model service providers can help MNCs achieve efficient localization. When forming partnerships, MNCs should assess the model providers' technical capabilities, compliance track record, ecosystem maturity, and ability to provide ongoing service and support.

#6

Applications – China's Strategic Edge

Applications are not an inherent part of AGI, but they are an important part of the AGI ecosystem and play a critical role in driving its development. Real-world needs, whether for industrial automation or personal companionship, drive developers to push technological boundaries to meet those needs. Applications in different scenarios generate valuable feedback that can be used to iteratively improve models. In addition, commercial returns from successful applications can be reinvested in R&D, creating a virtuous cycle that drives AGI exploration.

Chinese tech industry bets on application strength to leapfrog competitors on the curve.

The Chinese tech industry generally favors an application-oriented approach to large-scale AI models over purely technological pursuits. This inclination stems in part from the practical need to generate cash flows, and in part from a playbook that built the tech giants' success in the mobile internet era.

Chinese private investors also favor products that can be commercialized quickly. MiraclePlus, a Chinese venture capital firm and startup incubator, was the most active investor in the AI sector in 2024, with 34 deals, particularly targeting application companies. Commercialization pressure from investors is one of the main reasons why many large model companies have shifted their focus to applications.

Source: Sinolytics research; IMD Center forDigital Transformation

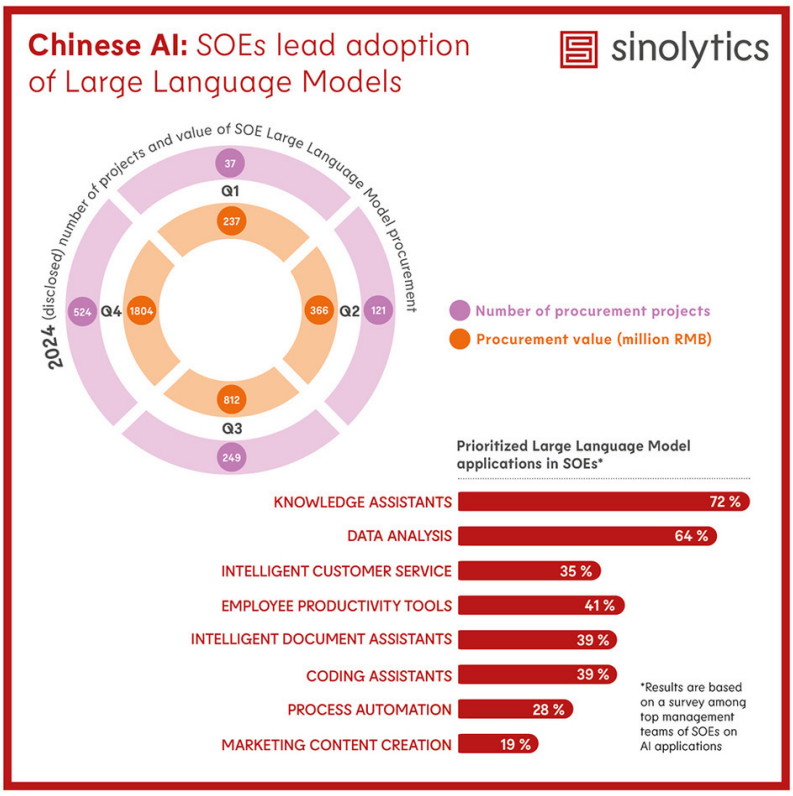
In the race to develop commercially successful AI products, China's tech giants and startups have adopted different strategies. Tech giants continue to research basic models while leveraging the advantages of their existing commercial ecosystems to drive applications, focusing on both business- and consumer-facing products.

Meanwhile, leading AI startups have pivoted to smaller, domain-specific models (Baichuan and 01.AI) or put more effort into application development (Moonshot and Minimax).

The Chinese government serves as both a major customer and a facilitator of AI commercialization, but high adoption has yet to translate into financial gains.

In addition to advancing the underlying technologies, the Chinese government is also committed to promoting industrial applications. Government support is focused on stimulating demand, with numerous government agencies and state-owned enterprises as key business customers for large language models.

Since 2023, central government-supervised SOEs have met several times to roll out industry-specific application scenarios. In 2024, a total of 931 central and state-owned enterprises launched large-scale model procurement tenders, accounting for 61.25% of the total market share. Leading SOEs such as State Grid, CNPC, and China Merchants Group are actively investing in AI deployment. As of February 21, 2025, it is estimated that 45% of China's 98 centrally supervised SOEs have completed the deployment of DeepSeek models.



Sources: 2024 central State-owned Enterprise Large Language Model application tracking report, China Large Language Model bidding project monitoring report (2024), Sinolytics research

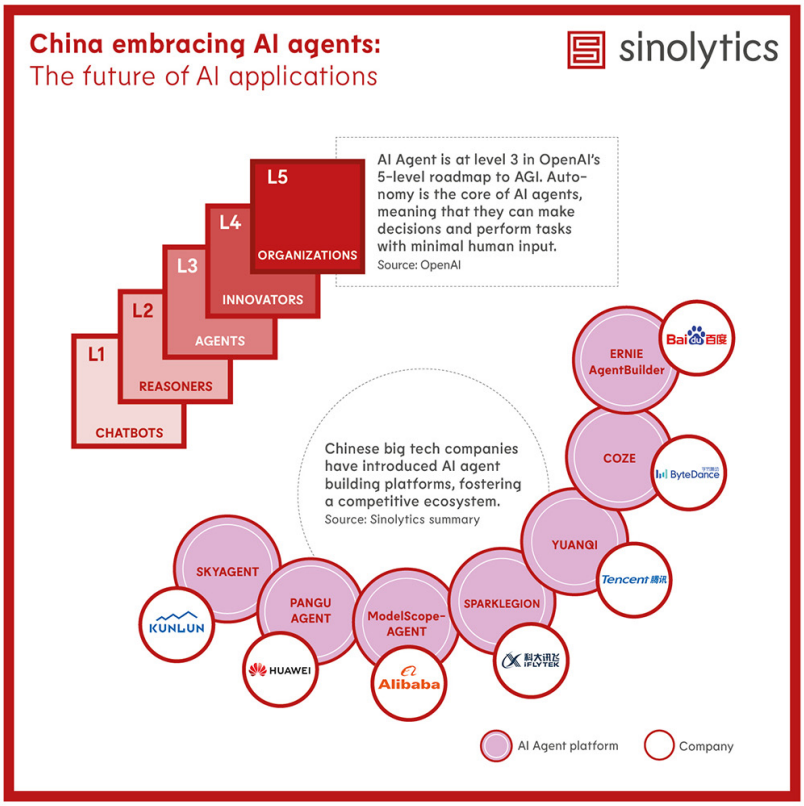
Political pressure is driving government agencies and state-owned enterprises to lead the deployment of large models, although the impact on operational efficiency and corporate profitability remains unclear. Intense competition is forcing large model companies to lower their bidding prices, resulting in squeezed profits and the adage "the more contracts won, the more blood lost."

The government is also acting as a bridge for AI commercialization by connecting startups with key resources such as data, computing, capital, and application scenarios. In Shanghai, the Mosu Space incubator aims to nurture thousands of AI applications by connecting startups with resource providers such as SenseTime, which has donated 1000P of compute, and venture capital funds such as IDG Capital.

Similarly, Shenzhen's Model Power Camp supports AI companies by providing free rent, computing subsidies, and access to data sets and application scenarios.

China aims to expand from digital AI agents to humanoid robotics, although large-scale deployment remains a long-term goal.

AI agents are widely considered to be the most promising application trend in 2025. An agent can use tools to access external information and perform autonomous reasoning and action planning. With the advent of RL-based reasoning models such as DeepSeek R1, the performance and deployment of AI agents is expected to continue to improve.



China's AI agent market has already attracted a wide range of players. Tech giants such as Baidu, Alibaba, Tencent, and ByteDance have all launched agent products and comprehensive AI agent development platforms to enable enterprises and individual developers to create their own agents for specific application scenarios, while smaller AI startups are focusing on AI agent applications in vertical scenarios.

Although most AI agents exist as virtual assistants or specialized systems integrated into industrial applications, embodied AI as a subset of AI agents is gaining significant traction in China, especially in the fields of humanoid robotics and autonomous driving. Embodied intelligence is an AI system that incorporates a physical body to interact with the environment, make decisions, and take actions. Large models have provided embodied AI with enhanced data processing and learning capabilities, enabling more accurate analysis of complex environments and efficient responses to diverse task requirements.

There is a strong consensus in the technology sector that embodied intelligence is a critical pathway for transitioning AI from the digital realm into the physical world, with humanoid robots being the ultimate expression of this transition. Some argue that achieving AGI requires physical embodiment, suggesting that robotics companies are best positioned to deliver true AGI.

The humanoid robotics industry has taken off, driven by strong policy support and robust investment. China's tech giants, such as Huawei and Tencent, have established in-house R&D capabilities. Automotive companies such as Xpeng with its AI humanoid robot "Iron" are leveraging their advantages in autonomous driving and manufacturing scenarios, while home appliance companies such as Haier with its robot "Kuavo" are investing in home scenarios. Leading startups such as UBTECH, Fourier, Unitree and Zhiyuan are growing fast and have started small-scale production. Although some robots are already working in factories and even performing at the Spring Festival Gala, large-scale commercial adoption is still on the horizon.

What does it mean?

China is demonstrating unique strengths at the application layer that are driving progress towards AGI. The country's ability to leverage diverse scenarios and a large user base is generally seen as a competitive advantage, potentially providing a counterstrategy if the U.S. achieves AGI hegemony.

China's expansive government institutions and state-owned enterprises have opened up vast scenarios for AI companies, providing tremendous demand-side support for an industry in its early stages. And in frontier applications such as embodied intelligence, China's complete manufacturing industrial chain, built over decades of industrial development, is a distinct advantage.

China offers multinational companies a valuable testing ground for AI application products, particularly in scenarios such as consumer applications, smart cities, and smart manufacturing. The Chinese market allows MNCs to explore ideas, test products, and iterate quickly. It also serves as a source of inspiration and insights that can inform and enhance their global operations. However, China's market also presents unique challenges, making it essential for multinational companies to develop a go-to-market strategy in advance.

#7

Conclusion

China's path to AGI is fraught with challenges. AI chips remain China's most critical vulnerability in the AI value chain, making it difficult to efficiently develop advanced models without significant investment and technological breakthroughs in domestic alternatives. However, China has strong advantages in the downstream application layer – a large AI-enabled user base, diverse application scenarios, and a robust manufacturing industrial chain that supports the transition from digital AI agents to embodied intelligence. These strengths could serve as a counterbalance should the U.S. take the lead in realizing AGI.

The Chinese government plays a critical role in this landscape. It acts as a major investor channeling money into the sector, an strategist designing national roadmaps, a facilitator creating synergy across sectors, and a key client purchasing AI services. Chinese AI companies are key players in driving AGI progress, especially as they contribute to chip advancement and the big model race, where breakthroughs from companies like Huawei and DeepSeek are accelerating progress and reshaping global competition.

Growing competition between China and the U.S. is driving innovation but also accelerating the decoupling of AI between East and West. As multinational companies develop AI-powered products, they will face new challenges navigating divergent ecosystems, making a China-specific strategy essential to tap into the market's growing potential.

Get
in
touch.

Turn insights into action:

Contact us today and unlock the full potential of your business.

info@sinolytics.de
www.sinolytics.de

sinolytics

Imprint:

Sinolytics GmbH,
Leipziger Platz 16,
10117 Berlin

HRB 199715 B
VAT ID: DE317668739
CEO: Björn Conrad