CHINA'S DRIVE TOWARD SELF-RELIANCE IN ARTIFICIAL INTELLIGENCE
From chips to large language models

WENDY CHANG, REBECCA ARCESATI, ANTONIA HMAIDI

China
Tech
Observatory

MERICS
Mercator Institute for China Studies

China
Tech
Observatory

ABOUT THE CHINA TECH OBSERVATORY

The China Tech Observatory (CTO) of Mercator Institute for China Studies (MERICS) is funded by the German Ministry for Research, Technology and Space (BMFTR). The three-year research project takes stock of China's progress in developing and using globally critical technology. It provides information and analysis to help decision-makers in government, business and other areas to better understand China's aims and efforts in future technologies.
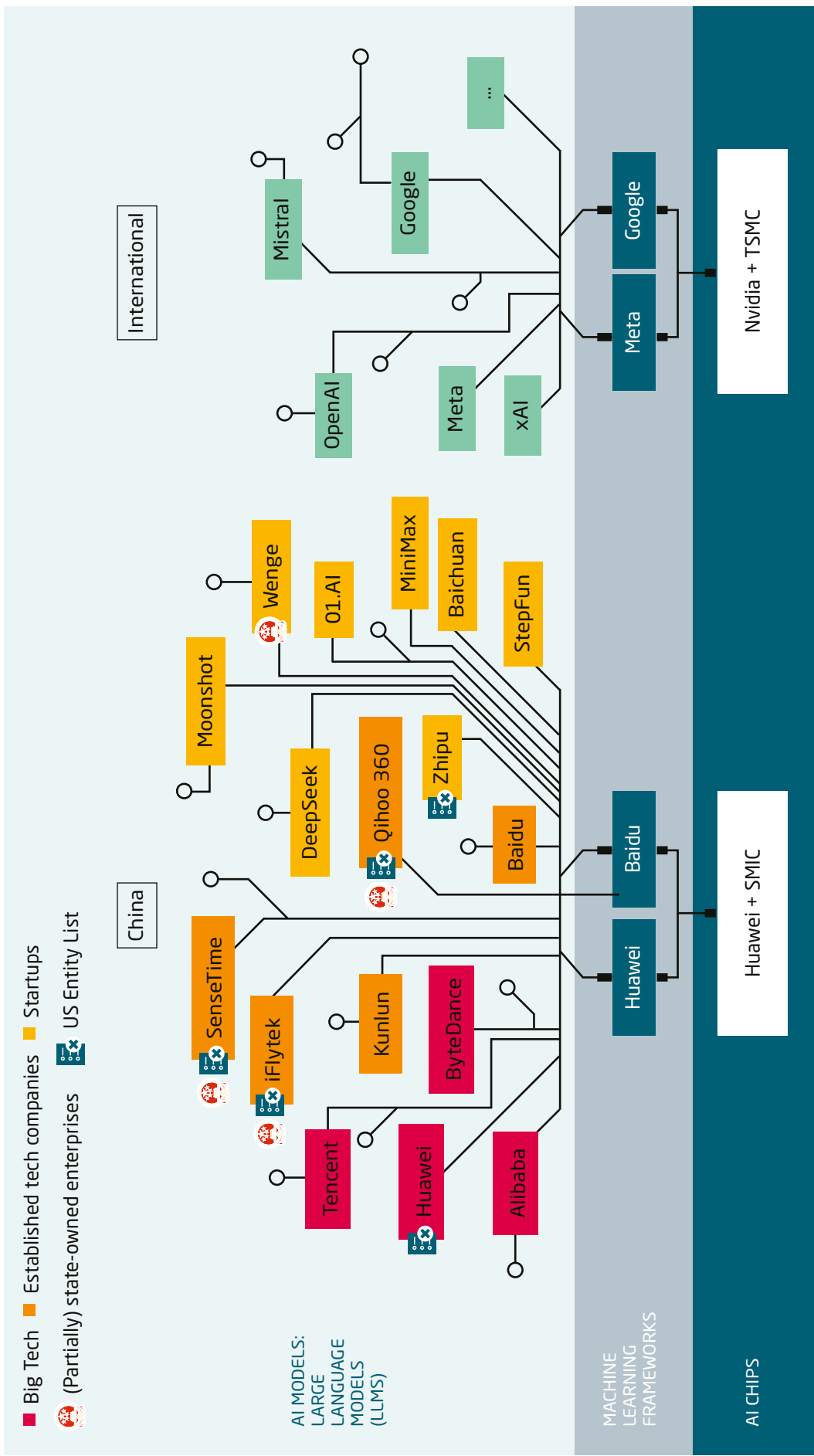
■ **China is pursuing self-reliance in AI at every level of technology.** It sees AI as strategic for national and economic security. Facing technology export controls from the US, Beijing has made "independent and controllable" AI a key objective. Examining China's efforts may provide useful learnings for Europe in its own pursuit of digital sovereignty.

■ **While China's government has long identified AI capabilities as a critical goal, it employs different strategies to aid each layer.** The heaviest state support is reserved for the capital-intensive semiconductor sector. Indigenization efforts for software frameworks are entrusted to Big Tech companies. Higher layers, i.e., AI models and applications, benefit from an enabling environment but receive less direct state support.

■ **China's semiconductor industry has managed to produce its own AI chips, but their performance does not yet match that of US semiconductor designer Nvidia.** Beijing has set indigenous capabilities as a top priority, especially faced with US export controls. Huawei leads this effort, working closely with domestic chipmakers.

■ **In models and applications, China is closing in on the US. China is heavily embedded in global open-source communities.** Coupled with a protected home market, this has spawned large language model (LLM) developers like DeepSeek. Hardware challenges still hinder wider deployment, but local adoption of LLMs is high, and China's AI industry is pivoting toward specialized applications.

■ **China's AI ecosystem can source critical inputs domestically, but its future will also hinge on external factors.** The country has nurtured a large talent pool, provided ample funding, promoted a maturing data environment and built computing infrastructure. Vulnerabilities include limited access to advanced chips and China's future participation in the global open-source community, which has long been key for its AI progress.

The AI Stack: Few players at the bottom, many on top

Legend:
- Big Tech
- Established tech companies
- Startups
- (Partially) state-owned enterprises
- US Entity List

International

- Mistral
- OpenAI
- Google
- Meta
- xAI
- ...

China

- Wenge
- 01.AI
- MiniMax
- Baichuan
- StepFun
- Moonshot
- DeepSeek
- Qihoo 360
- Zhipu
- Baidu
- SenseTime
- iFlytek
- Kunlun
- ByteDance
- Tencent
- Huawei
- Alibaba

AI MODELS: LARGE LANGUAGE MODELS (LLMS)

MACHINE LEARNING FRAMEWORKS
- Google
- Meta
- Baidu
- Huawei

AI CHIPS
- Nvidia + TSMC
- Huawei + SMIC

© MERICS

Source: MERICS

INTRODUCTION: CHINA IS MAKING RAPID PROGRESS ACROSS THE AI TECH STACK

The race for supremacy in Artificial Intelligence (AI) technology is now at the forefront of geopolitical competition between China and the United States. Not only is AI expected to reshape the way we live and work, but its potential use in military applications could also alter the global balance of power. And some believe that reaching artificial general intelligence, or AGI, where AI surpasses humans in capability, is akin to the race to build the atomic bomb: whoever unlocks AGI first will be the winner of the geopolitical competition.[1] The 2022 release of ChatGPT, the first publicly available generative AI tool based on a large language model (LLM) from US company OpenAI, kicked the AI race into high gear.

During the Politburo study session of April 2025, focused on AI, China's party and state leader Xi Jinping urged a nationwide mobilization to achieve "self-reliance and self-strengthening" （自立自强） in the technology by building an "independent and controllable" （自主可控） ecosystem across hardware and software.[2] The emphasis on AI sovereignty marks a relatively recent turn in Chinese AI strategy and policymaking. Previously, official documents still called for international cooperation. But since the US government's policy of constraining China's AI progress, many in China have called to sanction-proof its AI ecosystem. To be self-sufficient, China would need to produce the AI models, the software frameworks needed to create the models, and the chips that fuel their training and deployment – the whole AI technology "stack."

The US has identified AI and the semiconductor technologies that enable its advancements as among several general-purpose "force-multipliers", and US leadership as a "national security imperative."[3] Accordingly, Washington has rolled out policies aimed at slowing China's AI development, primarily through export controls on advanced semiconductors needed to train AI models, as well as on the software and equipment used to manufacture those chips.

China's leaders similarly view advancements in AI as key to national security and overall competitiveness, and they have characterized it as a strategic emerging field that offers the country a historic opportunity to leapfrog the US economically and militarily. Beijing published its first national AI strategy in 2017, outlining steps to become a "major AI innovation center in the world."[4] US export controls have reinforced Beijing's conviction that China's dependency on foreign countries – particularly the US – for its AI future is a national security risk.

China is arguably the first country trying to develop a national AI stack, but some in Europe have recently articulated this objective, too. So, where is China now in terms of AI chips, frameworks and applications? Who are the most important players? How is the government shaping the development and adoption of homegrown technology? How does this differ across the stack and why?

Understanding China's AI progress is crucial for European policymakers trying to navigate technology competition – for example to predict US export controls that might affect European companies. It can also help decision makers assess where Europe has leverage and how to engage, including companies that are increasingly embedding local solutions in their offerings for the Chinese market.

Moreover, examining China's self-sufficiency push in AI may also help Europeans predict possible problems in developing a "EuroStack."[5] Europe has some strategic presence, in-

> China is arguably the first country trying to develop a national AI stack

cluding key suppliers of semiconductor manufacturing equipment and a few notable LLM startups. However, the continent lags in developing and deploying the most advanced AI systems. Growing transatlantic tensions, especially under the new administration of Donald Trump, have renewed calls for "sovereign AI" solutions.[6]

CHINA'S AI STACK: DIFFERENT LAYERS, DIFFERENT APPROACHES

A simplified AI stack consists of three layers: chips that power computations, machine learning frameworks used to build AI models, and applications like large language models (LLMs). Beijing would ideally like self-sufficiency in all areas, but a clear hierarchy is emerging. State support is focused on the lower layers of the stack like chips and frameworks, with higher layers – AI models and applications – benefiting from an enabling environment but much less direct state support.

US and allied export controls now restrict sales to China of cutting-edge Graphic Processing Units (GPUs), like those designed by Nvidia, the high-bandwidth memory they are used in conjunction with, and the hardware and software required to manufacture these semiconductors. As a result, this is considered the most crucial layer for developing home-grown technology. Machine learning frameworks are open source and thus cannot be targeted by export controls at present. LLMs and their applications are moving rapidly, with companies constantly outpacing each other, bolstered by an active software ecosystem and access to global open-source developments.

In this report, we assess China's progress in these different layers of the AI technology stack. We also look at the ways in which the government shapes their growth, which differs considerably in each layer. We conclude with a brief discussion of key inputs on which the Chinese AI stack relies, namely capital, talent and data, and the outlook for Beijing to achieve its goals.

Direct state support is focused on lower layers of the AI stack

BOTTOM OF STACK: HUAWEI LEADS THE CHARGE

All AI is built on chips; they are necessary to run AI computational workloads. China's government views AI chipmaking as a fundamental capability in which it lags the US. Already in 2014 China launched the so-called "Big Fund" to bring together previously unconnected programs on chip development and manufacturing. The first phase ran until 2019 included 139 billion yuan, investing broadly across the supply chain, while the second phase with 204 billion yuan was designed to specifically focus on supply chain steps with Chinese weaknesses. In practice, however, both funds spent large chunks on front-end manufacturing.[7] A third phase of 340 billion yuan was announced in 2024. The Big Fund has had mixed results, for example with the collapse of high-profile chipmaker Tsinghua Unigroup.[8] But the big winners from the program currently dominate China's chipmaking industry: Semiconductor Manufacturing International Corporation (SMIC) as the most advanced logic foundry and Yangtze Memory Technologies Corp. (YMTC) as its top memory chip maker. SMIC is the only Chinese company that can make advanced 7 nanometer (nm) chips today.

Central state support is complemented by local state support, with many provinces or cities using their own investment vehicles to finance semiconductor companies. Tech giant Huawei has emerged as the main player in coordinating chipmaking, working closely with SMIC. For instance, Huawei's Hubble investment arm now invests across the supply chain, and importantly, it co-invests in many companies with the Shenzhen Major Industry Investment Group, a state-owned investment fund in the city of Shenzhen.[9]

Tech giant Huawei has emerged as the main player in coordinating chipmaking

Domestic manufacturing is holding back China's design capabilities

Huawei, in its quest to compete with US Big Tech like Apple, designed its first AI mobile chip in 2017, adding a server chip meant to rival Nvidia in 2018.[10] However, these chips were all manufactured at Taiwan's TSMC. Huawei's inclusion on the US Department of Commerce's Entity List in 2019, followed by the imposition of the "Foreign Direct Product Rule," barred the Chinese company from fabricating chips at TSMC.

Since 2016, several other "fabless" design companies have emerged in China. Design is a high value-added part of the semiconductor supply chain that is less capital-intensive than manufacturing. Thus, the barrier to entry is low, especially if companies can use electronic design automation (EDA) software and other tools for development.

China is now home to several AI chip design companies. A full version of the LLM Deep-Seek, for instance, can be run on Huawei, Cambricon, Moore Threads and Kunlun chips.[11] Importantly, these companies are still dependent on the ARM microarchitecture, designed by the Japanese-owned British company of the same name, as well as on EDA software from the West, although Huawei is also leading efforts to indigenize EDA software.

Exhibit 1

Producing advanced semiconductors

A highly complex process with many dependencies on global tech and equipment

1. Design
Chinese firms: Huawei, Cambricon, Moore Threads, Kunlun
Critical dependencies: Electronics design automation software (EDA)

2. Front-end manufacturing
Chinese firms: SMIC
Critical dependencies: Lithography machines, chemicals, other equipment

(Advanced chips only) Advanced packaging
Chinese firms: Tongfu, Xingsen

3. Assembly, testing and packaging (Back-end manufacturing)
Chinese firms: Tongfu, Xingsen

Source: MERICS

© MERICS

Front-end manufacturing is a bottleneck

All Chinese design companies are competing for the same limited 7nm manufacturing capacity at SMIC, as well as for high bandwidth memory (which cannot yet be sourced locally) and advanced packaging (to bundle multiple chips for enhanced performance). While SMIC and others work on expanding manufacturing, Huawei has a large advantage because it is the de-facto leader of the semiconductor "national team" with access to up to 70 percent of SMIC's capacity.[12]
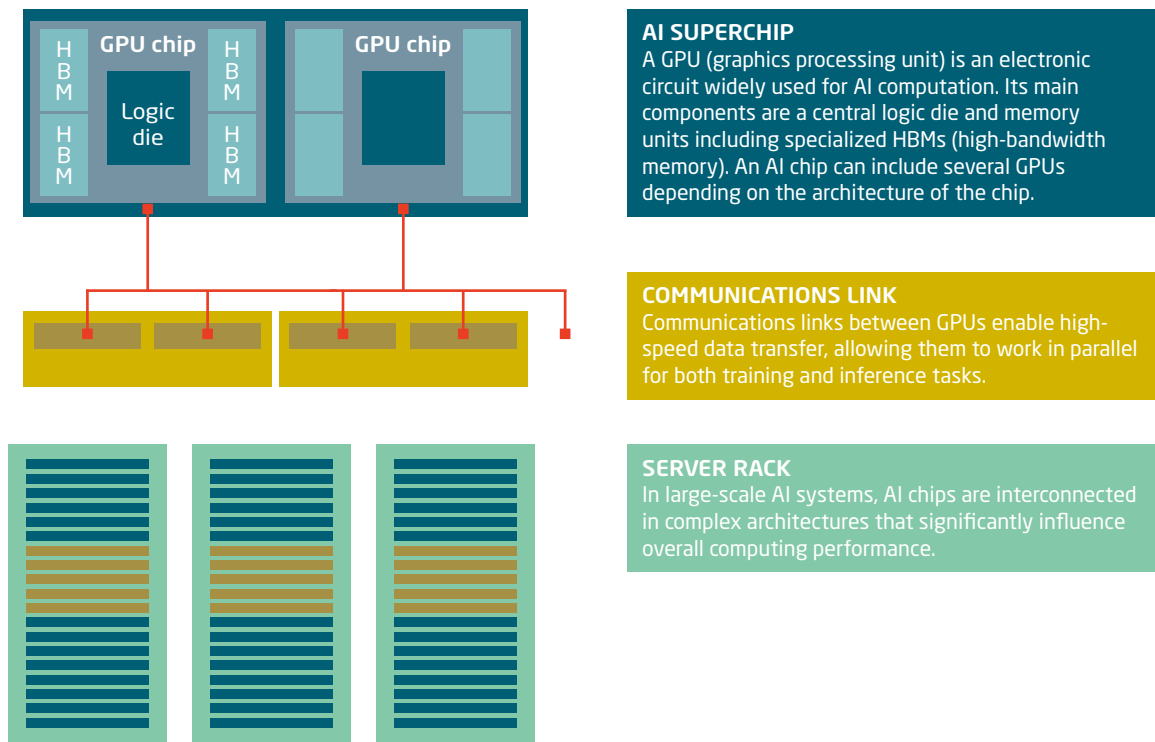
Huawei also made changes in 2022 from its 910 chip, manufactured at TSMC, to its newer 910B chip, designed for manufacturing at SMIC, that point to some production limitations.[13] The new chip outperforms its predecessor but contains fewer cores, despite a larger die size, which can limit performance.[14] At the end of 2024, moreover, credible rumors spread that Huawei had manufactured chips at TSMC in violation of US export controls, suggesting that SMIC's capabilities are more limited and lower in quality than believed.

Another key bottleneck is high bandwidth memory (HBM). Many AI tasks, notably those involved in training and running large language models, are highly dependent on memory. Huawei is working with memory chipmaker Tongfu Microelectronics to develop HBMs,[15] but it reportedly also stockpiled significant memory from South Korea before export controls came into effect.[16]

Exhibit 2

The anatomy of a modern AI chip

AI chips in a typical data center setup

GPU chip
HBM Logic die HBM
HBM HBM

GPU chip

AI SUPERCHIP
A GPU (graphics processing unit) is an electronic circuit widely used for AI computation. Its main components are a central logic die and memory units including specialized HBMs (high-bandwidth memory). An AI chip can include several GPUs depending on the architecture of the chip.

COMMUNICATIONS LINK
Communications links between GPUs enable high-speed data transfer, allowing them to work in parallel for both training and inference tasks.

SERVER RACK
In large-scale AI systems, AI chips are interconnected in complex architectures that significantly influence overall computing performance.

Source: MERICS

© MERICS

Due to manufacturing limitations, Huawei's newer 910C chip is just two 910B dies packaged together for better performance. This makes advanced packaging even more important. China has in recent years led the world in outsourced assembly and testing and thus has significant packaging capacity at home. But these fabs are focused on non-advanced packaging and must now be adapted to build up advanced packaging capacity. Xingsen Technologies is just one company doing so with Huawei's support.

Software issues hold back Chinese AI chips in applications

The 910B's performance is, on paper, roughly on par with Nvidia's A100, a chip that came out in 2019. In practice, however, Huawei's chips are beset by problems that make them less useful than their performance stats would suggest.

The strength of individual chips is not the most important factor. Most AI workloads run on multiple chips. Nvidia is the undisputed leader in AI chip design globally, not least because its chips are far easier to connect to each other, even between servers.

Part of Nvidia's competitive advantage lies in CUDA (Compute Unified Device Architecture), a programming interface and parallel computing platform. CUDA allows developers to port code easily between Nvidia chips and provides solutions to problems already in code. Most AI models so far have been written using CUDA, so switching to non-Nvidia chips is costly.

Chinese companies are trying to challenge Nvidia's dominance in two ways. Smaller companies like Moore Thread offer a compatibility layer for CUDA, so that developers switching to their chips do not need to change their CUDA code. But CUDA is optimized for Nvidia chips and is regularly updated, so compatibility layers need continuous support and will never be as efficient as a similarly high-performance framework developed natively.

By contrast, Huawei chose to develop its own alternative to CUDA, CANN. But CANN is not nearly as mature and usable: Huawei's chips still have random, frequent crashes without recovery, and developers are hesitant to port their models to CANN.[17] As changes are made to CANN, successive models may not be compatible, so a model running on the Ascend 910B will need significant changes to run on a future Ascend-based chip. This slows model innovation and deployment.

The bottom line: China still relies on foreign AI chips

Huawei chips are especially unsuitable for large LLM training runs. A few months ago, DeepSeek considered Huawei's most advanced AI chip at the time, the Ascend 910C, inadequate for training.[18]

Nvidia continues to lead the market for AI chips in China, having sold more than 1 million of its H20 chips in 2024. Huawei, on the other hand, only sold 200,000 AI chips, despite their lower price.[19] Until recently, the H20 was the most advanced AI chip Nvidia was still allowed to sell to its Chinese customers. In April 2025, the Trump administration slapped a licensing requirement (with a presumption of denial) on H20 exports to China,[20] which will likely make life even harder for China's AI industry.

In addition to manufacturing innovation, Chinese developers are also working on making better use of the limited chips available to them. For instance, Alibaba developed a technique for training the same model using a mix of foreign and domestic chips. [21] Moreover, Chinese hardware makers are offering server-scale machines that rival Nvidia's performance using significantly more chips and power.[22] China currently has abundant energy and subsidized energy prices, so it is better positioned than the West to meet AI's power demand.

MIDDLE OF STACK: CHINESE BIG TECH DOMINATES DOMESTIC MACHINE LEARNING OFFERINGS; USERS PREFER GLOBAL ONES

At the next layer up, machine learning frameworks are used to streamline development and make it easier. They are standard interfaces, libraries and toolkits for designing, training and verifying AI algorithms. While developers could write code from scratch to train and deploy models, using a framework is much faster and easier to port to new chips and iterate on. Frameworks can be viewed as the common language of AI development.

Leading AI frameworks globally are PyTorch and TensorFlow, both open source and originally developed by US companies. Due to their wide global adoption, Chinese developers like Huawei are contributing to these frameworks to include support for domestic GPUs, while also developing domestic alternatives.

The first major framework released by a Chinese company was Baidu's PaddlePaddle, published in 2016.[23] PaddlePaddle has been relatively widely adopted in China, especially in industry, and includes a lot of pre-trained models for typical industrial AI workloads.[24]

Nvidia continues to lead the market for AI chips in China

Huawei's machine learning framework, MindSpore, is available both for the company's own flagship AI chips, the Ascend GPUs, and for CPU-based machine learning. It was open-sourced in 2020 as a full-scenario AI computing framework. However, beyond its use for Huawei's own pre-trained models and some initial traction in industrial AI, adoption seems limited. On the Chinese source code repository Gitee, a search for MindSpore returns 684 results, with TensorFlow and PyTorch each returning more than 10,000.[25]

Considering Huawei has moved MindSpore development onto Gitee, this shows that adoption lags significantly. As a result, Huawei is still actively developing extensions for PyTorch and TensorFlow to enable support for Ascend GPUs. Huawei joined the PyTorch Foundation as a Premier member in 2023.[26] The company is also active in contributing to open source projects on Github.[27] Since PyTorch and TensorFlow are both open source with very permissive licenses, it is unlikely the US government will be able to restrict China from accessing them.

TOP OF STACK: FIERCE COMPETITION OVER LARGE LANGUAGE MODELS AND APPLICATIONS

At the top of the stack are AI applications. Due to the recent focus on LLMs, this report emphasizes LLMs and their many applications. They have become the focal point of global technology competition since the launch of ChatGPT in late 2022.

There is a vibrant scene in China rising to meet this challenge, with academia as well as private enterprises entering the race en masse to develop LLMs. Ample private funding and access to global open-source models have allowed for rapid Chinese progress. Although China lagged in the first years after ChatGPT's release, domestic models did not have competition from leading Western ones, since these were quickly banned after release. With the release of DeepSeek's models in late 2024, China became a contributor to the cutting edge of global LLM development.[28]

There is a vibrant LLM scene in China

Software development is a more open environment, so direct government support to this layer of the tech stack has been less relevant. One notable form has been provincial government vouchers for startups to purchase or rent computing power for model training. These typically range from 20 to 50 percent of compute costs, and up to a yearly total of 500 million yuan for all companies in the most generous locality (Shenzhen).[29] Aside from this mechanism, which is not insignificant, the government's role can be best described as incentivizing the ecosystem rather than direct subsidization.

China's LLM scene comprises large and small players. Established tech giants are all working on their own models. So are academic and state-backed research labs, which produce much of China's top AI talent and startup spin-offs. A crop of startups has emerged, alongside existing tech companies trying to capture the trend. Finally, there is the surprise disruptor DeepSeek, which hailed from Hangzhou, an emerging hub of AI innovation.[30]

While still trailing the most advanced Western competitors, Chinese LLMs now perform well across different evaluation metrics. UC Berkeley's Chatbot Arena is a popular ranking. Its scores are derived by comparing pairs of chatbots blind-tested and voted on by volunteers. In March 2025, the top 20 scores include five Chinese entries, led by DeepSeek-R1 in sixth place. Qwen2.5-Max is ranked tenth and Zhipu's GLM-4-Plus-0111 is fourteenth.[31] The rest of the leaderboard models are from leading Western companies including xAI, OpenAI and Google.

Table 1

Notable Chinese LLMs and their developers in 2025

CATEGORY	NAME	LEADING MODEL(S)	NOTES	SCORE*
Big Tech	Alibaba	Qwen	Leader in fostering China's AI open source community.	66.38 ②
	ByteDance	Doubao	Parent company of TikTok, its model powers one of the most popular chatbots in China.	64.68 ③
	Tencent	Hunyuan	Focused on integrating AI into its many tech products including WeChat.	62.49
	Huawei	PanGu	Stresses support for industrial applications.	
Established companies	Qihoo 360	360zhinao	Cybersecurity firm with ties to Chinese military, added to US Entity List in 2020.	59.08
	Kunlun	Sky-Chat	Originally a web gaming company.	49.17
	Baidu	ERNIE	Associated Ernie Bot was first major chatbot to debut in China, after ChatGPT's launch.	47.56
	iFlytek	Spark	Known for its leading voice recognition tech, added to US Entity List in 2019.	40.76
	SenseTime	InternLM	Added to US Entity List in 2019, developed InternLM with Shanghai AI Lab.	32.02
Startups	DeepSeek	Deep-Seek-R1	Little-known startups prior to breakout models, backed by a quant hedge fund.	70.33 ①
	Wenge	YAYI	Spun off from CAS.	55.81
	Moonshot	Kimi	Alibaba-backed startup planning to get listed.	51.47
	StepFun	Step	Exploring multi-modal models and integration with the auto industry.	50.81
	Zhipu AI	GLM	A spin-off of Tsinghua University, added to the US Entity List in January 2025.	48.61
	01.AI	Yi	Once among the top players, has pivoted away from building LLMs.	23.29
	Baichuan	Baichuan	Has pivoted towards applications, especially in medical AI.	
	MiniMax	MiniMax	Focuses on consumer apps.	
Research labs	BAAI	WuDao, Emu	Early developer of ultra-large-scale models, added to the US Entity List in March 2025.	
	Peng Cheng Lab	Peng Cheng Mind	Provides key infrastructure and platforms, added to US Entity List in January 2025.	

* Scores are from the March 2025 leaderboard of the SuperCLUE benchmark, which evaluates LLM performance using a suite of Chinese-language tasks.

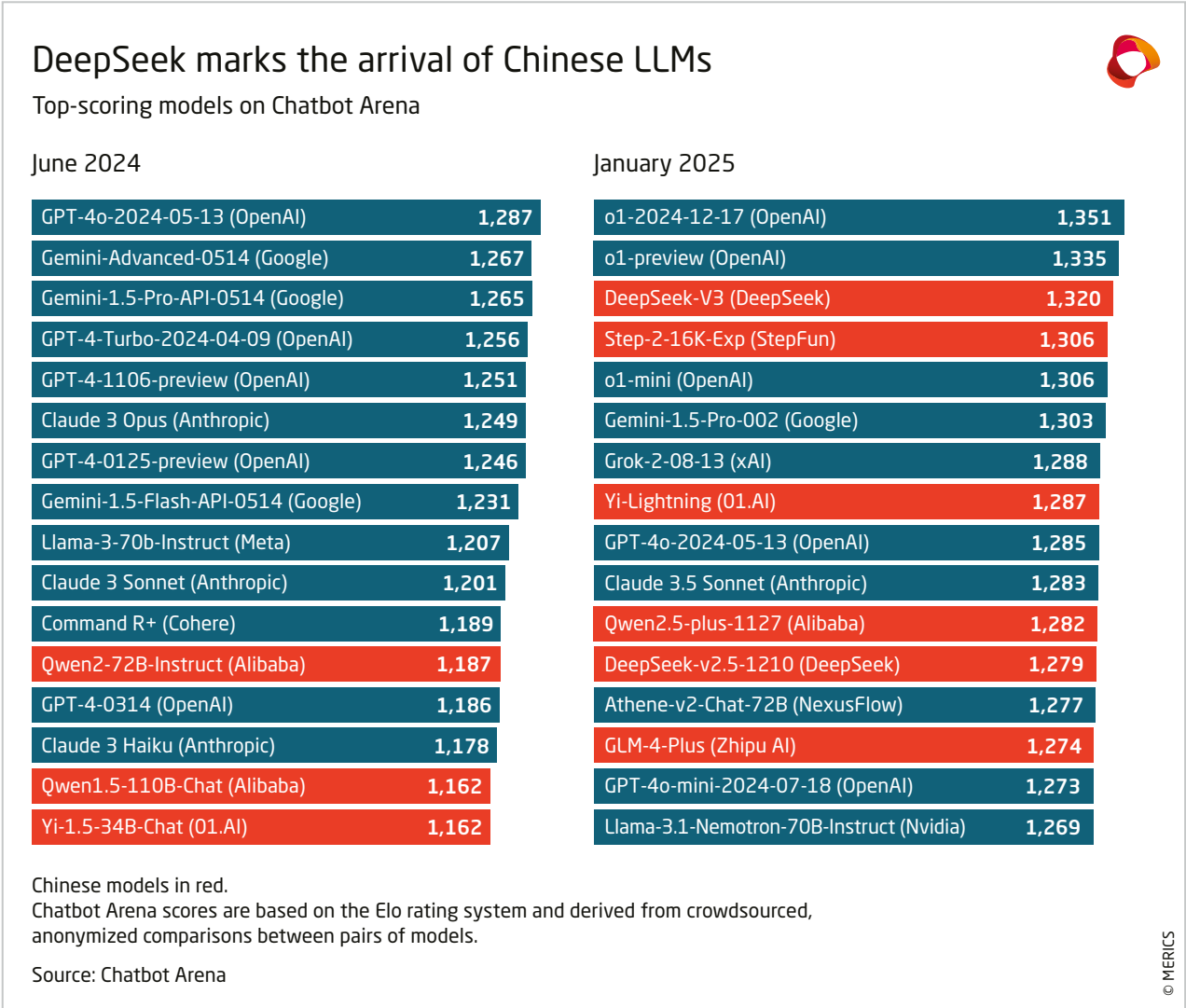Source: SuperCLUE leaderboard, March 2025

© MERICS

The benchmark SuperCLUE offers the most thorough look at the wide range of Chinese LLMs. It evaluates LLM performances in tasks using the Chinese language. In March 2025, the top scoring model was OpenAI's o3-mini, followed by DeepSeek's R1. Other Chinese LLMs, including the latest Qwen, Hunyuan and 360zhinao models, follow closely behind top models from Google and Anthropic.[32]

Government fosters the open-source ecosystem

LLMs have enjoyed a relatively easy development path in China compared to the hardware layers of the AI stack, in part thanks to the longstanding practice of open-source software development. Much of the foundational LLM research has been in the form of public papers and code repositories, free for developers around the world to finetune or build on. As a result, most top models are very similar in architecture, differentiating themselves instead in training data sets and fine tuning to yield competitive results against benchmarks. Meta has been particularly influential in open sourcing its state-of-the-art LLaMA series.[33] Many leading Chinese LLMs use the underlying architecture for their models, including Baichuan's Baichuan series and 01.AI's Yi series.[34]

Open sourcing for LLMs often refers to a slightly different concept than traditional software projects. In what is more accurately described as "open-weight," model weights, which are

Exhibit 3

DeepSeek marks the arrival of Chinese LLMs

Top-scoring models on Chatbot Arena

June 2024

Model	Score
GPT-4o-2024-05-13 (OpenAI)	1,287
Gemini-Advanced-0514 (Google)	1,267
Gemini-1.5-Pro-API-0514 (Google)	1,265
GPT-4-Turbo-2024-04-09 (OpenAI)	1,256
GPT-4-1106-preview (OpenAI)	1,251
Claude 3 Opus (Anthropic)	1,249
GPT-4-0125-preview (OpenAI)	1,246
Gemini-1.5-Flash-API-0514 (Google)	1,231
Llama-3-70b-Instruct (Meta)	1,207
Claude 3 Sonnet (Anthropic)	1,201
Command R+ (Cohere)	1,189
Qwen2-72B-Instruct (Alibaba)	1,187
GPT-4-0314 (OpenAI)	1,186
Claude 3 Haiku (Anthropic)	1,178
Qwen1.5-110B-Chat (Alibaba)	1,162
Yi-1.5-34B-Chat (01.AI)	1,162

January 2025

Model	Score
o1-2024-12-17 (OpenAI)	1,351
o1-preview (OpenAI)	1,335
DeepSeek-V3 (DeepSeek)	1,320
Step-2-16K-Exp (StepFun)	1,306
o1-mini (OpenAI)	1,306
Gemini-1.5-Pro-002 (Google)	1,303
Grok-2-08-13 (xAI)	1,288
Yi-Lightning (01.AI)	1,287
GPT-4o-2024-05-13 (OpenAI)	1,285
Claude 3.5 Sonnet (Anthropic)	1,283
Qwen2.5-plus-1127 (Alibaba)	1,282
DeepSeek-v2.5-1210 (DeepSeek)	1,279
Athene-v2-Chat-72B (NexusFlow)	1,277
GLM-4-Plus (Zhipu AI)	1,274
GPT-4o-mini-2024-07-18 (OpenAI)	1,273
Llama-3.1-Nemotron-70B-Instruct (Nvidia)	1,269

Chinese models in red.
Chatbot Arena scores are based on the Elo rating system and derived from crowdsourced, anonymized comparisons between pairs of models.

Source: Chatbot Arena

© MERICS

the parameters resulting from a trained model like LLaMA or DeepSeek, are shared, while the training code and datasets are not. Developers can build on these open models but cannot recreate them.[35] This has implications for entities who want to deploy open-weight models, as the accuracy, proper handling of sensitive data and IP, and potential bias in the training data cannot be verified.

China's government has long embraced the open-source movement, both to decrease dependency on foreign technology companies and to ramp up progress toward self-sufficiency.[36] It has supported using open-source standards in many technologies, such as RISC-V for chip design architecture and Kylin Linux for operating systems. Responding to policy direction, a consortium of Chinese tech companies launched the Open Atom Foundation in 2020 to promote open-source development in cutting-edge tech projects. The platform is seeded by projects from tech giants, such as Huawei's OpenHarmony operating system.[37]

Chinese companies and software developers are very active in open-source communities. In 2024, Chinese developers made up around 9 percent of all developers on GitHub, which is the largest source code repository in the world and one of the few remaining major Western internet platforms to still operate inside China.[38] To create a domestic alternative, both for ensuring continued access and for better control of activities on the platform, China's Ministry of Industry and Information Technology (MIIT) backed local hosting platform Gitee in 2020.[39] The platform sees some traction from Chinese users, gaining praise for faster access speed inside China and better localization, but many still prefer GitHub to engage with the international community.[40]

Alibaba Cloud, or Aliyun, is a leader in Chinese AI open-source efforts. Its Qwen series regularly tops open-source leaderboards.[41] Alibaba also operates ModelScope, a platform for open-source AI models fashioned after Hugging Face, a platform that has become the default destination for open source LLM projects globally. Hugging Face itself was banned in China, a reflection of the political sensitivity around generative AI.[42] Many Chinese com-

Gitee – China's Github alternative

Gitee is built a lot like GitHub, organized around individual projects, issues, and pull requests for collaborative development. It features built-in integrations with more Chinese services, such as Baidu and Tencent Cloud, and Gitee Go for continuous integration and deployment. In 2022, Gitee announced that all code would be manually reviewed before public availability, raising censorship concerns.
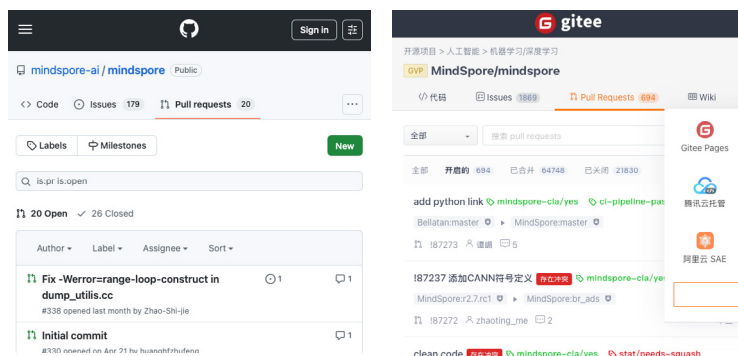
panies participate in the open-source community by open sourcing slightly older models, while keeping their most advanced commercial models close-sourced.

DeepSeek caught the world's attention with low costs and high performance, but its deployment remains a challenge

DeepSeek's release of its R1 model in January this year marked the arrival of the little-known startup in the LLM scene, and brought wider awareness of Chinese LLMs outside China.[43] Deep-Seek's remarkably efficient use of computing resources to achieve performance similar to far more expensive models has sparked discussions ranging from the competitiveness of Chinese AI to the effectiveness of chip export controls, even briefly causing a sharp drop in Nvidia share prices.[44]

DeepSeek's breakthrough lies in combining existing technical solutions to radically reduce the computing resources required to train and run a world-class model.[45] Custom communications schemes between GPUs also helped cut training time.[46] DeepSeek's models are open weight, and its research papers describing these innovations have contributed to the state of the art in LLM development.[47]

Outside the tech community, DeepSeek has garnered attention for its high performance offered at a dramatically lower price, although there are some disputes around the true training and access costs of its models.[48] Importantly, it sparked discussions of Chinese LLMs entering global market competition, raising questions around Chinese government censorship and data gathering.[49] However, as a startup, DeepSeek does not have the infrastructure to offer its products on a large scale. Many Chinese companies are rushing to fill this void, although these early efforts have yielded mixed results.

Results of a test on the Chinese platforms offering DeepSeek illustrate the challenges companies face in supporting top models. The test was conducted by SuperCLUE on 18 Chinese platforms that offered DeepSeek-R1 in February 2025, including both free and paid offerings. The results were poor, showing long response times for questions and high rates of cut-off responses or none at all. [50] The best performers were US firms Perplexity and to-gether.ai, a testament to the importance of access to cutting-edge chips in deploying LLMs effectively. In a follow-up test in March, US platforms disappeared from the chart. [51]

The best performing Chinese host of DeepSeek is VolcEngine, the cloud service from ByteDance. A relative newcomer after the big three – Alibaba, Huawei and Tencent – it is trying to catch up by doubling down on AI offerings.[52] Overall, China's cloud market is much smaller than that of the US, due in part to lower adoption of public cloud solutions by enterprises.[53] While cloud service providers want to expand by investing in AI infrastructure, enterprises are also embracing private solutions that allow them quick access.[54] The growing popularity of "all-in-one" machines offering DeepSeek models reflects this.[55] In the end, widespread adoption and easy access to LLMs are what will propel the dissemination of AI technology into innovations across different fields. It remains to be seen if China's divergent approach will yield better results.

External constraints and government priorities urge a pivot toward applications

DeepSeek's arrival in a way marks the end of the "battle of hundred models" era in which many firms pursued LLM development.[56] 01.AI has announced it will stop training new models, focusing instead on offering DeepSeek-based solutions[57]; Baichuan has pivoted into medical AI, and Kimi is also exploring opportunities in more specific applications.[58]

Chinese companies and software developers are very active in open-source communities

Some in the local AI community already advocate for a shift in focus from cutting-edge LLM capabilities toward building applications for existing models.[59] Beyond China's traditional strength in taking existing innovations and building on them rather than developing new paradigms, there are several other reasons for this "pragmatic" approach.

First is the shortage of cutting-edge AI chips in China due to US export controls. While model developers are attempting to use Chinese-made AI chips, dependence on Nvidia chips is widely acknowledged.[60] Many companies and institutions rely on stockpiles acquired prior to export controls, access them remotely through cloud service providers, or in some cases get their hands on smuggled chips. [61]

With a limited supply of AI chips for training and deploying large foundational models, China may see a better path to commercial success in AI applications. Qihoo 360 CEO Zhou Hongyi, for example, has suggested that a vertical approach involving smaller models with proprietary data may be best for resource-constrained Chinese enterprises.[62] The government, for its part, wants to focus on the application aspect of AI technology. In both 2024 and 2025, the government work reports presented at the National People's Congress featured an "AI+ Initiative" to promote the use of AI solutions to boost manufacturing and other sectors like electric vehicles, robotics, education and medicine.[63]

The Hangzhou government stands out as a successful example for fostering its startup environment with financial support and other incentives. Six AI startups there, including DeepSeek, represent an emerging wave of innovative technologies coming out of the Chinese tech ecosystem. They showcase the range of opportunities AI brings to different sectors – Unitree and Deep Robotics in robotics, Game Science in gaming, BrainCo in brain-computer interface, Manycore in spatial simulations applied to design.[64]

CHINA INCREASINGLY RELIES ON ITSELF FOR AI INPUTS, BUT FACES CHALLENGES

Achieving AI self-reliance will hinge on domestic and external factors. The AI stack relies on several inputs whose availability is crucial for any country or region to compete in advanced AI systems. These are capital, talent, data and infrastructure. Understanding the availability and role of these AI inputs in China is critical in assessing its capabilities as well as potential technological progress in the years ahead.

In terms of capital, the role of US investment, once a significant contributor to China's AI ecosystem, has sharply declined. Already before the Biden administration introduced a framework allowing the Treasury Department to scrutinize, and in some cases bar, US investment in Chinese technology industries including AI, US venture capital had started pulling back, reaching a ten-year low of USD1.3 billion in 2022, down from USD14.4 billion in 2018.[65] This is causing Chinese AI firms to shift to alternative sources of capital to fund their growth, namely CNY-denominated funds as well as dollars from deep-pocketed investors in places like the Gulf states. Although the country's venture capital market has been in decline since 2021, according to Pitchbook, China saw 715 deals in the AI sector in 2024 totaling USD7.3 billion, higher than any other country in Asia.[66]

Despite this enthusiasm, largely driven by the generative AI boom, Beijing's bet on a state-driven, techno-nationalist innovation model may hamper the future development of traditionally privately driven sectors.[67] In January, China launched a new AI investment fund with an initial capital of CNY 60 billion (USD 8.2 billion), followed by a distinct state
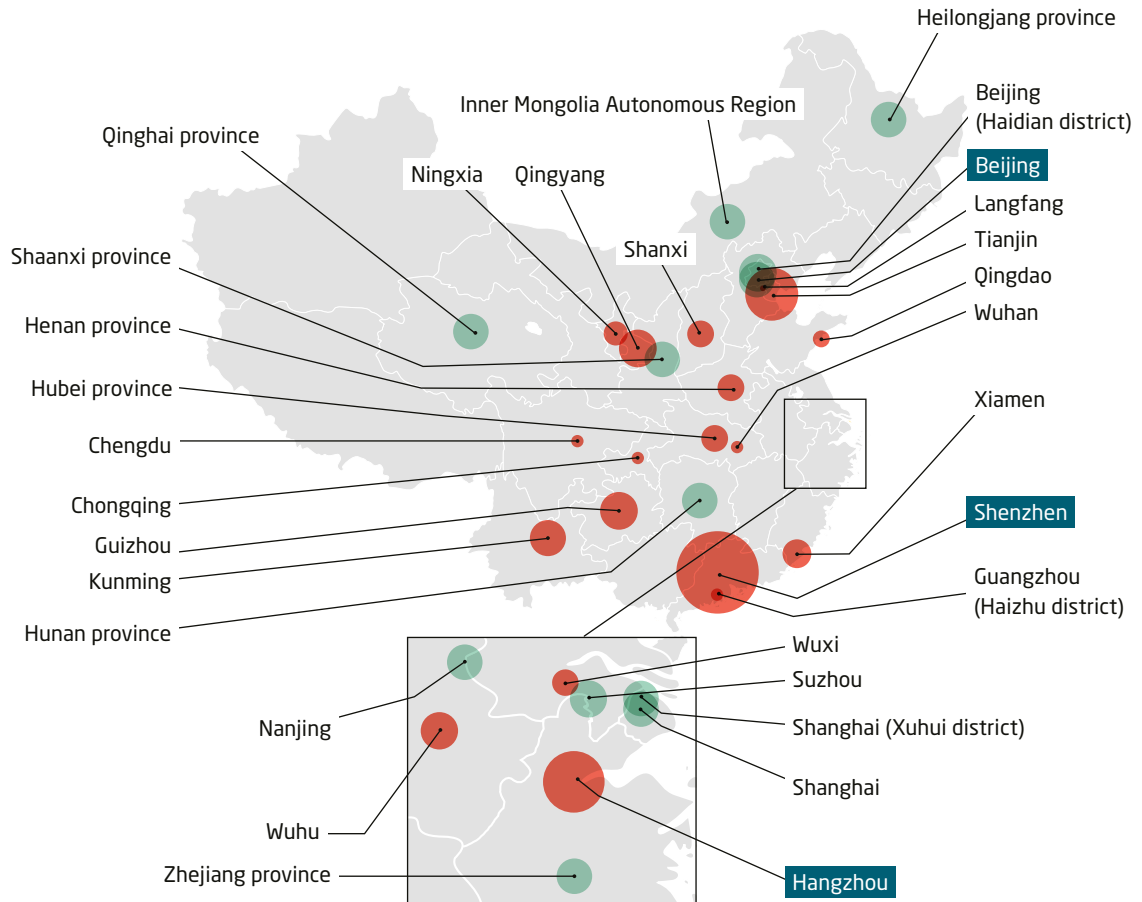
Exhibit 4

Chinese LLM developers are eligible for compute subsidies

Provincial and municipal governments distribute vouchers to support model training and adoption

SHENZHEN

Up to 10 million per enterprise per year for companies that undertake large model training, with a waiver of up to 50% of the contract value and up to 60% for startups.

HANGZHOU

For eligible LLM companies carrying out model training, inference or application, up to 8 million per enterprise per year, and up to 20-30% of the contract amount depending on circumstances.

BEIJING

Up to 1 million per enterprise per year, and up to 10 million and 20 million for eligible AI enterprises located within Haidian District and the Beijing Special Economic Zone, respectively.

Source: Local government policies

© MERICS

guidance fund for critical tech sectors that the National Development and Reform Commission (NDRC) unveiled at this year's Two Sessions.[68] Seasoned investors in China's tech ecosystem privately caution that state capital could introduce inefficiencies and slow innovation. This is not to mention the economic slowdown, which raises serious questions about the future viability of China's system of science, technology and innovation.[69]

In terms of talent, DeepSeek's experience demonstrates that a brilliant and motivated cohort of young, mostly domestically educated engineers and entrepreneurs is eager to develop and commercialize new innovations at home.[70] Indeed, research by MacroPolo showed that, in 2022, a quarter of the world's elite AI scientists received their undergraduate degrees from a Chinese university. More of these top talents are choosing to live and work in their home country: 28 percent compared to 11 percent only three years earlier.[71]

In other words, while the US remains the largest pool of elite AI talent, China is quickly closing that gap. While a general shortage of AI workers is still reported, and especially acute in specific fields and functions like speech recognition and algorithm engineering, China's talent outlook is clearly positive.[72] Research outputs reflect this trend, with China accounting for roughly 40 percent of highly cited AI papers in 2021, and already surpassing the US share in 2016, according to CSET's Emerging Technology Observatory.[73]

China's large ecosystem is not closed off, at least not so far. CSET found that a considerable proportion of China's highly cited AI research came from collaborations with US peers, although much of China's increase relative to the US could be attributed to non-collaborative papers.[74] AI research collaboration between China and the US remains resilient overall despite geopolitical tensions, although it is no longer growing at the same rate as prior to 2020.[75] Because AI and machine learning research thrives on cross-border collaboration, self-reliance would seem hardly feasible or desirable for China.

China's large ecosystem is not closed off, at least not so far

The extent to which Beijing regards such global AI communities as a strength rather than a vulnerability is increasingly in question, however. The Politburo Study Session on AI emphasized self-sufficiency and toned-down rhetoric on openness and global integration. The government is also trying to nurture domestic software development and code-hosting platforms. This seems to run counter to the interests of most Chinese developers, who continue to prefer GitHub due to its international exposure. Players like Huawei (through its MindSpore ecosystem) and China Software Development Network (CSDN, which counts about 10 million registered users and ties with American firms like Microsoft and Intel) are working to strengthen domestic AI developer communities.

Meanwhile, Beijing's continued political support for open models – an approach that Europe is also betting on[76] – is not a given. Some DeepSeek employees are reportedly under travel restrictions,[77] a sign that a control-obsessed party state may tighten its grip.

Data is another crucial ingredient in AI development. Although the availability of large-scale and rich datasets has provided an advantage to Chinese firms in sectors like computer vision and related applications like biometric recognition,[78] this is not true across the board. There are persistent issues with the quality of data corpuses used for training, and even with inadequate supply.[79] The government considers this a priority and has launched an ambitious effort, led by the National Data Administration, to better integrate China's data market.[80] This kind of proactive industrial policy for the data economy could help solve some of the challenges China's AI industry is facing in this area. Recent policies to stimulate the data labelling industry fit the same logic.[81]

The infrastructure through which data is processed and ingested by AI systems is just as important. This takes the form of data centers that typically contain hundreds of thousands of high-performance chips, network and storage architectures, as well as energy and cooling capabilities to handle AI workloads. China has poured billions into AI computing infrastructure, including through a megaproject to develop a nationwide grid of data cen-

ters and AI computing hubs, particularly by exploiting the abundant land and renewable energy sources of inland provinces.[82]

China has both strengths and weaknesses in the race to build AI infrastructure. On the one hand, the combination of its low electricity prices and forward-looking energy strategy – including plans to build the world's first hybrid fusion-fission reactor by 2031 – and the paradigm of hardware-efficient AI innovation pioneered by DeepSeek could position China relatively well to balance energy security and technological advances. On the other hand, the hype around LLMs among commercial players and local governments has led to wasteful investments, with up to 80 percent of newly built computing resources unused.[83]

Hardware export controls will impact China's trajectory

No discussion of the prospects for an indigenization of China's AI stack would be complete without considering external factors – specifically, foreign governments' export controls and their likely impact on the country's AI ecosystem. In October 2022, the US introduced unilateral export controls on certain semiconductors, chip design software, and chip-making equipment. Those controls were beefed up in 2023 and again in 2024.

So far, Chinese LLM developers have managed to access American-designed GPUs to train and deploy their models. To train the R1, DeepSeek claimed to have used merely 2,048 Nvidia H800 GPUs, which Washington only banned in October 2023. DeepSeek reportedly also had access to 10,000 of the less powerful A100 GPUs. Others estimate the company had access to many more Nvidia GPUs, and some in Congress have launched a probe into possible smuggling.[84] In a nutshell, export controls are not perfect firewalls, and Washington could tighten the screws further. US firms will keep adjusting to this: Nvidia is due to release a downgraded version of its H20 to keep servicing Chinese customers.[85]

The trajectory of AI export controls, however, is far from certain. In a big move, the Trump administration repealed the AI Diffusion framework, which was drawn up by the Biden administration to control the global spread of advanced AI capabilities.[86] This would have been executed mainly through a tiering system tied to a mix of export caps and security requirements for model training and deployment.[87] The previous administration's had eyes on countries like the UAE and Malaysia, believed to be transshipment hubs through which restricted GPUs end up in China. Essentially, the US would have sought to lock in the leadership of its companies in the global AI ecosystem and manage security risks – largely by isolating China through shaping the technology decisions of third countries.[88] Trump plans to replace this framework with his own, as yet unknown, version.

The trajectory of AI export controls is far from certain

For now, Trump's intent seems to be twofold. First, while subscribing to the same strategic objective of containing China's AI capabilities, his administration prefers soft guidance to help industry enforce existing rules, rather than new regulation. Second, Trump personally favors bilateral deals, as seen in agreements with Saudi Arabia and the UAE expected to open the door for future exports of large numbers of the most advanced GPUs.[89]

Corporate interests and third countries' choices, in addition to those made in Beijing and in Washington, will continue to influence the urgency with which China must indigenize its AI stack.

EXTERNAL FACTORS COULD CHANGE CHINA'S GROWTH TRAJECTORY ACROSS THE AI STACK

China's government regards AI as a crucial technology and is doing all it can to foster self-reliance, particularly in homegrown AI chips and the middle layer where Chinese firms are most vulnerable to being cut off.

While Chinese companies can design good, if not cutting-edge chips, they are still severely constrained by domestic chip-making capabilities. These are particularly hampered by US, Japanese and Dutch export controls on manufacturing equipment. Moreover, Nvidia's advantage exists not only in hardware, but also in software and cross-chip capabilities, which are difficult to replicate.

While a few chip designers in China are active in AI, Huawei has emerged as the leader of the "national team for AI chips." The company is building up an entire ecosystem spanning chips, computing frameworks and an AI framework. This, along with government support, makes it increasingly difficult for other Chinese AI chip companies to compete. While multiple Chinese chip designers can design competitive AI chips, they are all competing for SMIC's high-end manufacturing capacity. Huawei's outsized role gives it an important advantage. As a result, other designers are unlikely to emerge as real competitors. Some are likely to look for niches where they can be successful going forward.

For the machine learning frameworks used to build LLMs and other AI, tech giants Baidu and Huawei lead the long-term effort to develop local alternatives to global frameworks, the preferred choice of Chinese AI developers. Smaller companies try to add support for their chips to these global frameworks. Meanwhile, the government is supporting efforts to further the adoption of local frameworks, especially in application development.

China is quite self-sufficient in the development and adoption of competitive LLM models, with many firms in the space. This success is built on the ample access it has enjoyed thus far to the global open-source community. Chinese and US regulations keeping foreign models out have also helped local competitors thrive. The "DeepSeek moment" has turbocharged adoption of LLMs in China. DeepSeek has also encouraged the practice of open source (or open weight) in China's LLM ecosystem, possibly speeding up innovation.

Computing power restrictions and the government's own preferences are both pushing for a more application-focused approach to AI development, with fewer companies trying to build LLMs. If those companies open-source their models, this could lead to very fast iteration and widespread deployment. Instead of the "battle of a hundred models" seen at the beginning of the LLM boom, we could soon witness an explosion of applications built on the winners of the battle. This configuration could play more to China's strengths and alleviate the high hardware requirements of building LLMs from scratch.

It is also possible for China to fall behind. Long-term growth would suffer if access to the open-source community and AI computing power was hindered.[90] While China has built up a large talent pool, is shoring up investment and can provide the energy to power its data centers, government allocation of resources is often inefficient. China's AI ecosystem owes much to the ingenuity of its engineers and private entrepreneurs, but also to critical international input; the latter is looking increasingly difficult to maintain.

International input is critical but increasingly difficult to maintain

POLICY OPTIONS FOR EUROPE

- **Europe should decide whether to integrate China-origin AI technology into its AI stack.** DeepSeek has attracted European users, which has led to scrutiny of its data and content moderation practices. Lawmakers and regulators in charge of enforcing the GDPR and the AI Act should consider that Chinese LLM developers are bound by the CCP's national security and censorship demands.

- **China's bet to fully indigenize the AI stack is informed by its domestic conditions and political economy.** Its innovation model prices in inefficiency in pursuit of strategic objectives. Rather than throwing massive resources at building a sovereign stack, Europe may stand a better chance if it focuses on its comparative advantage and strategic interests, based on clear priorities.

- **Europe lags in data infrastructure to promote adoption of its own AI solutions.** This probably deserves its most urgent attention. China is pushing adoption through a mix of cloud-as-infrastructure projects, while US cloud service hyperscalers can count on sheer numbers of chips and users. Europe must find its own path. Reforms are also needed so the EU can attract and retain top talent.

- **Europe could leverage its industrial strengths to build AI applications.** Opportunities unlocked by DeepSeek's compute-efficient paradigm should be put into perspective. Hyperscalers with abundant chips will keep releasing the best general-purpose models, but the wider ecosystem is shifting toward test-time, or inference-time compute. Europe's abundant industrial data could be the foundation to deploy smaller models, which are often better for specialized tasks.

- **Competing in lower layers of the stack is a longer game.** Even China, with its clear government strategy and funding, has long struggled to develop competitive AI chips. Europe should leverage its niches in semiconductor supply chains rather than focus on areas where it is not competitive.

- **China's experience shows that state-led approaches have limitations.** Private companies that initially take off with limited state connections or support, like DeepSeek, can often be the most disruptive. In contrast, efforts to bend the choices of AI firms and developers to techno-nationalist agendas rarely bear fruit.

- **Geopolitical competition will continue to shape Europe's choices.** The US and China both aim for dominance in AI and their ecosystems are bifurcating. Reliance on US hardware could expose Europe to supply constraints or even coercion. But alternatives like Huawei's AI chips, besides technical difficulties, may present security risks or even be in breach of US export controls.

Research for this report was concluded in June 2025.

ENDNOTES

1 | Metz, Cade (2023). "The ChatGPT King Isn't Worried, but He Knows You Might Be." The New York Times. Retrieved from https://www.nytimes.com/2023/03/31/technology/sam-altman-open-ai-chatgpt.htmlTong, Anna, Martina, Michael, & Tong, Anna (2024). "US government commission pushes Manhattan Project-style AI initiative." Reuters. Retrieved from https://www.reuters.com/technology/artificial-intelligence/us-government-commission-pushes-manhattan-project-style-ai-initiative-2024-11-19/

2 | "习近平在中共中央政治局第二十次集体学习时强调：坚持自立自强 突出应用导向 推动人工智能健康有序发展＿＿中国政府网" (n.d.). Retrieved June 24, 2025, from https://web.archive.org/web/20250613142922/https://www.gov.cn/yaowen/liebiao/202504/content_7021072.htm

3 | House, The White (2022). "Remarks by National Security Advisor Jake Sullivan at the Special Competitive Studies Project Global Emerging Technologies Summit." Retrieved June 13, 2025, from The White House https://bidenwhitehouse.archives.gov/briefing-room/speeches-remarks/2022/09/16/remarks-by-national-security-advisor-jake-sullivan-at-the-special-competitive-studies-project-global-emerging-technologies-summit/

4 | "新一代人工智能发展规划 (New Generation Artificial Intelligence Development Plan)" (2017). Retrieved April 25, 2025, from State Council https://web.archive.org/web/20170721053549/https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm; Jochheim, Ulrich (2021). "China's ambitions in Artificial Intelligence." European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/ATAG/2021/696206/EPRS_ATA(2021)696206_EN.pdf

5 | "The White Paper" (2025). Retrieved June 13, 2025, from EuroStack https://euro-stack.eu/the-white-paper/

6 | "HPC Digital Autonomy with RISC-V in Europe" (2025). Retrieved June 13, 2025, from Jülich Supercomputing Centre https://www.fz-juelich.de/en/ias/jsc/projects/dare

7 | "China's Big Fund 3.0: Xi's Boldest Gamble Yet for Chip Supremacy – The Diplomat" (n.d.). Retrieved June 24, 2025, from https://thediplomat.com/2024/06/chinas-big-fund-3-0-xis-boldest-gamble-yet-for-chip-supremacy/

8 | Mozur, Paul (2021). "The Failure of China's Microchip Giant Tests Beijing's Tech Ambitions." The New York Times. Retrieved from https://www.nytimes.com/2021/07/19/technology/china-microchips-tsinghua-unigroup.html

9 | "bloomberg.com/graphics/2023-china-huawei-semiconductor/" (n.d.). Retrieved June 24, 2025, from https://www.bloomberg.com/graphics/2023-china-huawei-semiconductor/

10 | Ye, Josh (2023). "How Huawei plans to rival Nvidia in the AI chip business." Retrieved April 25, 2025, from Reuters https://www.reuters.com/technology/how-huawei-plans-rival-nvidia-ai-chip-business-2023-11-07/

11 | "部分国产芯片适配满血版 DeepSeek，仍「遥遥无期 (Some domestic chips are still 'far from being compatible with the full version of DeepSeek')" (2025). Retrieved April 25, 2025, from WeChat, AI Technology Review https://web.archive.org/web/20250310174412/https://mp.weixin.qq.com/s/BiieiP-ViGXzH_5j9x-LOyA

12 | "Huawei is quietly dominating China's semiconductor supply chain | Merics" (2024). Retrieved June 24, 2025, from https://merics.org/en/report/huawei-quietly-dominating-chinas-semiconductor-supply-chain

13 | Feldgoise, Jacob, & Dohmen, Hanna (2024). "Pushing the Limits: Huawei's AI Chip Tests U.S. Export Controls." Center for Security and Emerging Technology. Retrieved from https://cset.georgetown.edu/publication/pushing-the-limits-huaweis-ai-chip-tests-u-s-export-controls/

14 | Shilov, Anton (2024). "Huawei's homegrown AI chip examined — Chinese fab SMIC-produced Ascend 910B is massively different from the TSMC-produced Ascend 910." Retrieved April 25, 2025, from Tom's Hardware https://www.tomshardware.com/tech-industry/artificial-intelligence/huaweis-homegrown-ai-chip-examined-chinese-fab-smic-produced-ascend-910b-is-massively-different-from-the-tsmc-produced-ascend-910

15 | "Tongfu Microelectronics Co.," Retrieved June 18, 2025, from https://en.tfme.com/Zuhair, Muhammad (2025). "Chinese Firm Tongfu Announces Trial Production Of HBM2 Process; Expected To Be Featured In Huawei's AI Chips." Retrieved June 18, 2025, from Wccftech https://wccftech.com/chinese-firm-tongfu-announces-trial-production-of-hbm2-process/

16 | Yang, Heekyong, Potkin, Fanny, & Freifeld, Karen (2024). "Chinese firms stockpile high-end Samsung chips as they await new US curbs, say sources." Reuters. Retrieved from https://www.reuters.com/technology/chinese-firms-stockpile-high-end-samsung-chips-they-await-new-us-curbs-say-2024-08-06/

17 | McMorrow, Ryan, Olcott, Eleanor, & Hu, Tina (2024). "Huawei's bug-ridden software hampers China's efforts to replace Nvidia in AI." Retrieved April 25, 2025, from Financial Times https://www.ft.com/content/3dab07d3-3d97-4f3b-941b-cc8a21a901d6

18 | Allen, Gregory C. (2025). "DeepSeek, Huawei, Export Controls, and the Future of the U.S.-China AI Race." Center for Strategic & International Studies. Retrieved from https://www.csis.org/analysis/deepseek-huawei-export-controls-and-future-us-china-ai-race

19 | Potkin, Fanny (2024). "Exclusive: Nvidia cuts China prices in Huawei chip fight." Retrieved April 25, 2025, from Reuters https://www.reuters.com/technology/nvidia-cuts-china-prices-huawei-chip-fight-sources-say-2024-05-24/

20 | "Nvidia shares plunge amid $5.5bn hit over export rules to China" (n.d.). Retrieved June 24, 2025, from https://www.bbc.com/news/articles/cm2xzn6jmzpo

21 | Team, Ling, Zeng, Binwei, Huang, Chao, Zhang, Chao, Tian, Changxin, Chen, Cong, ... He, Zhengyu (2025). Every FLOP Counts: Scaling a 300B Mixture-of-Experts LING LLM without Premium GPUs. https://doi.org/10.48550/arXiv.2503.05139

22 | "Huawei AI CloudMatrix 384 – China's Answer to Nvidia GB200 NVL72" (2025). Retrieved June 24, 2025, from SemiAnalysis https://semianalysis.com/2025/04/16/huawei-ai-cloudmatrix-384-chinas-answer-to-nvidia-gb200-nvl72/

23 | PaddlePaddle/Paddle [C++] (2025). Retrieved from https://github.com/PaddlePaddle/Paddle (Trabalho original publicado em 2016)

24 | "PaddlePaddle: An Open-Source Deep Learning Framework - viso.ai" (n.d.). Retrieved June 24, 2025, from https://viso.ai/deep-learning/paddlepaddle/

25 | Conducted on April 1, 2025 on https://so.gitee.com/, archive of search here: "Gitee Search – 开源软件搜索" (2025). Retrieved April 25, 2025, from https://web.archive.org/web/20250401070846/https:/so.gitee.com/?q=mindspore

26 | "Huawei Joins the PyTorch Foundation as a Premier Member" (2023). Retrieved April 25, 2025, from PyTorch https://web.archive.org/web/20250425121040/https://pytorch.org/blog/huawei-joins-pytorch/

27 | "Add Ascend backend support · Issue #11477 · microsoft/onnxruntime" (2022). Retrieved April 25, 2025, from GitHub https://github.com/microsoft/onnxruntime/issues/11477

28 | DeepSeek-AI et al. (2025). DeepSeek-V3 Technical Report. https://doi.org/10.48550/arXiv.2412.19437

29 | "算力券全攻略｜一文搞懂全国补贴详情 (Compute Power Voucher Full Strategy Guide | All National Subsidy Details Explained in One Article)" (2025). Retrieved June 13, 2025, from Northern Computing Network on Zhihu https://archive.is/XNL4H

30 | "中文大模型基准测评2025年3月报告 (Chinese Large Model Benchmark Report, March 2025)" [2025.03] (2025). Retrieved June 13, 2025, from SuperCLUE https://web.archive.org/web/20250405011050/https://www.cluebenchmarks.com/superclue_2503

31 | "LLM Leaderboard: Community-driven Evaluation for Best LLM and AI chatbots" (2025). Retrieved April 25, 2025, from Chatbot Arena (formerly LMSYS) https://huggingface.co/spaces/lmarena-ai/chatbot-arena-leaderboard/tree/main

32 | "SuperCLUE中文大模型测评基准——评测榜单 (SuperCLUE Chinese Large Model Evaluation Benchmark List)" (2025). Retrieved April 25, 2025, from SuperCLUE https://web.archive.org/web/20250405011050/https://www.cluebenchmarks.com/superclue_2503

33 | Zuckerberg (2024). "Open Source AI is the Path Forward." Retrieved June 13, 2025, from Meta Newsroom https://about.fb.com/news/2024/07/open-source-ai-is-the-path-forward/

34 | baichuan-inc (2023). "Baichuan-7B." Retrieved June 13, 2025, from GitHub https://github.com/baichuan-inc/Baichuan-7B/blob/main/README_EN.md; Peng, Tony (2023). "Kai-Fu Lee's LLM: A LlaMA Lookalike, China's LLM Overload, and US-China AI Risk Talk Begins." Retrieved June 13, 2025, from https://recodechinaai.substack.com/p/kai-fu-lees-llm-a-llama-lookalike

35 | "Llama and ChatGPT Are Not Open-Source Few ostensibly open-source LLMs live up to the openness claim" (2023). Retrieved June 13, 2025, from IEEE Spectrum https://spectrum.ieee.org/open-source-llm-not-open

36 | Arcesati, Rebecca, & Meinhardt, Caroline (2021). China bets on open-source technologies to boost domestic innovation | Merics. Retrieved from MERICS https://merics.org/en/report/china-bets-open-source-technologies-boost-domestic-innovation

37 | "Open-Source Technology and PRC National Strategy: Part I" (2025). Retrieved June 13, 2025, from China Brief, The Jamestown Foundation https://jamestown.org/program/open-source-technology-and-prc-national-strategy-part-i/

38 | "Octoverse: AI leads Python to top language as the number of global developers surges" [GitHub] (2024). Retrieved April 25, 2025, from The GitHub Blog https://github.blog/news-insights/octoverse/octoverse-2024/

39 | "工信部携码云 Gitee 入场，国内开源生态建设进入快车道 (MIIT Enters the Scene with Gitee, Propelling Domestic Open-Source Ecosystem Development into the Fast Lane.)" (2020). Retrieved June 13, 2025, from Gitee Open Source Society https://archive.is/C7DuD

40 | "大佬们都是用啥托管代码的？gitee还是github？(What do you guys use to host code ? gitee or github?)" (2022). Retrieved June 13, 2025, from UCloud Community https://web.archive.org/web/20250518181300/https://www.ucloud.cn/yun/ask/78295.html

41 | Chinese AI models storm Hugging Face's LLM chatbot benchmark leaderboard — Alibaba runs the board as major US competitors have worsened." Retrieved June 13, 2025, from Tom's Hardware https://www.tomshardware.com/tech-industry/artificial-intelligence/chinese-llms-storm-hugging-faces-chatbot-benchmark-leaderboard-alibaba-runs-the-board-as-major-us-competitors-have-worsened

42 | Schneider, Jordan (2024). "Hugging Face Blocked! 'Self-Castrating' China's ML Development + Jordan at APEC." Retrieved June 13, 2025, from https://www.chinatalk.media/p/hugging-face-blocked-self-castrating

43 | "DeepSeek-R1 Release" (2025). Retrieved April 25, 2025, from DeepSeek API Docs https://web.archive.org/web/20250121060835/https://api-docs.deepseek.com/news/news250120

44 | Bradshaw, Tim, Olcott, Eleanor, Alim, Arjun Neil, Smith, Ian, & Lewis, Leo (2025). "Tech stocks slump as China's DeepSeek stokes fears over AI spending." Financial Times.

45 | Lambert, Nathan (2025). "DeepSeek V3 and the cost of frontier AI models." Retrieved June 13, 2025, from https://www.interconnects.ai/p/deepseek-v3-and-the-actual-cost-of

46 | The main innovations it used include a more memory-efficient implementation of attention operators (multi-head latent attention, or MLA), a key building block of LLMs; using smaller data field sizes during training (FPA-8); and using a Mixture-of-Experts approach to improve efficiency.

47 | "What DeepSeek Means for Open-Source AI Its new open reasoning model cuts costs drastically on AI reasoning" (2025). Retrieved June 13, 2025, from IEEE Spectrum https://spectrum.ieee.org/deepseek

48 | "DeepSeek Debates: Chinese Leadership On Cost, True Training Cost, Closed Model Margin Impacts" (2025). Retrieved June 13, 2025, from SemiAnalysis https://semianalysis.com/2025/01/31/deepseek-debates/

49 | Pearl, Matt, Brock, Julia, & Kumar, Anoosh (2025). Delving into the Dangers of DeepSeek. Retrieved from https://www.csis.org/analysis/delving-dangers-deepseekRollet, Charles (2024). "Hugging Face CEO has concerns about Chinese open source AI models." Retrieved June 13, 2025, from TechCrunch https://techcrunch.com/2024/12/03/huggingface-ceo-has-concerns-about-chinese-open-source-ai-models/

50 | "第三方平台DeepSeek-R1稳定性报告：18家网页端测评 (Third-party platform DeepSeek-R1 stability report: 18 web-based evaluations)" (2025). Retrieved April 25, 2025, from WeChat https://web.archive.org/web/20250228054143/https://mp.weixin.qq.com/s/_Wg4cDfYi8l9-Rb4dHACAg

51 | "DeepSeek-R1第三方平台稳定性测试（DeepSeek-R1 Third-Party Platform Stability Testing）" (2025). Retrieved June 13, 2025, from SuperCLUE https://archive.is/FJ8Pa

52 | "大模型，能改变火山引擎在云牌桌上的位置吗？（Can Large Models Change Volcengine's Standing in the Cloud Computing Game?）" (2024). Retrieved June 13, 2025, from Leiphone https://web.archive.org/web/20250218155026/https://www.leiphone.com/category/industrycloud/OfACPVmQbcieLUY6.html

53 | Schneider, Jordan (2025). "Why China's Cloud Lags." Retrieved June 13, 2025, from https://www.chinatalk.media/p/the-political-economy-of-chinas-cloud

54 | "2025年'云'展望：AI、出海、下沉市场或迎'黄金期'（2025 Cloud Outlook: AI, International Expansion, and Lower-Tier Markets May Usher in a 'Golden Era.'）" (n.d.). Retrieved June 13, 2025, from Shanghai Securities News https://web.archive.org/web/20250518172515/https://m.caijing.com.cn/article/360110?target=blank

55 | "爆火的大模型一体机，炒作还是真需求？(The exploding big model all-in-one, hype or real demand?)" (2025). Retrieved June 13, 2025, from 21Jingji https://web.archive.org/web/20250327080701/https://www.21jingji.com/article/20250327/herald/1792cf3df44d036cbc93f93dc1cfcad6.html

56 | "百模大战落幕，大模型'六小虎'开始分野（The Battle of a Hundred Models Concludes; The 'Six Little Tigers' of Large Models Begin to Differentiate.）" (2024). Retrieved June 13, 2025, from 36Kr https://archive.is/P5shw

57 | "零一万物发布大模型平台，李开复：放弃巨型模型研发，我们做不起（01.AI Launches Large Model Platform; Kai-Fu Lee: We're Forgoing Giant Model R&D, We Can't Afford It.）" (2025). Retrieved June 13, 2025, from Tencent Tech AI Official Account https://web.archive.org/web/20250517175547/https://news.qq.com/rain/a/20250317A0548M00

58 | "百川之后，Kimi悄然布局AI+医疗（After Baichuan, Kimi Quietly Ventures into AI + Healthcare）" (2025). Retrieved June 13, 2025, from 36Kr https://web.archive.org/web/20250516095834/https://36kr.com/p/3290880683078790

59 | "'Too many' AI models in China: Baidu CEO warns of wasted resources, lack of applications" (2024). Retrieved June 13, 2025, from South China Morning Post https://www.scmp.com/tech/tech-trends/article/3269338/too-many-ai-models-china-baidu-ceo-warns-wasted-resources-lack-applications

60 | "ChatGPT算力消耗惊人，能烧得起的中国公司不超过3家 (ChatGPT Demands Astronomical Computing Power; No More Than Three Chinese Companies Can Afford the Cost.)" (2023). Retrieved June 13, 2025, from Caijing Eleven https://web.archive.org/web/20230305153648/https://m.huxiu.com/article/811823.html

61 | Baptista, Eduardo (2024). "China's military and government acquire Nvidia chips despite US ban." Reuters. Retrieved from https://www.reuters.com/technology/chinas-military-government-acquire-nvidia-chips-despite-us-ban-2024-01-14/

62 | 《人工智能十问》(Ten Questions on Artificial Intelligence)" (2024). Retrieved April 25, 2025, from Baijiahao https://web.archive.org/web/20240423064140/https://baijiahao.baidu.com/s?id=1796125727205056736

63 | "政府工作报告 (Government Work Report)" (2025). Retrieved April 30, 2025, from At the Third Session of the 14th National People's Congress on March 5, 2025 Li Qiang, Premier of the State Council https://web.archive.org/web/20250317193732/https://www.gov.cn/yaowen/liebiao/202503/content_7013163.htm

64 | "'杭州六小龙'出圈背后：城市与创业者的相互成就 (Behind the popularity of the 'Hangzhou Six Little Dragons': the mutual achievement of cities and entrepreneurs)" (2025). Retrieved April 30, 2025, from Securities Times https://web.archive.org/web/20250224053509/https://www.stcn.com/article/detail/1529436.html

65 | "Big Strides in a Small Yard: The New US Outbound Investment Screening Regime – Rhodium Group" (2023). Retrieved June 24, 2025, from https://rhg.com/research/big-strides-in-a-small-yard-the-new-us-outbound-investment-screening-regime/

66 | Yu, Yifan (2025). "China venture capitalists hold back on AI deals despite DeepSeek buzz." Retrieved June 13, 2025, from Nikkei Asia https://asia.nikkei.com/Business/Technology/China-venture-capitalists-hold-back-on-AI-deals-despite-DeepSeek-buzz

67 | "Whole-of-Nation Innovation: Does China's Socialist System Give it an Edge in Science and Technology? - IGCC" (n.d.). Retrieved June 24, 2025, from https://ucigcc.org/ https://ucigcc.org/publication/whole-of-nation-innovation-does-chinas-socialist-system-give-it-an-edge-in-science-and-technology/

68 | "New AI fund in China to pour US$8 billion into early-stage projects" (2025). Retrieved June 24, 2025, from South China Morning Post https://www.scmp.com/tech/policy/article/3306047/new-ai-fund-china-pour-us8-billion-early-stage-projects; Liu, Nectar Gan, Juliana (2025). "China announces high-tech fund to grow AI, emerging industries | CNN Business." Retrieved June 24, 2025, from CNN https://www.cnn.com/2025/03/06/tech/china-state-venture-capital-guidance-fund-intl-hnk

69 | Rhodium Group (2023). Spread Thin: China's Science and Technology Spending in an Economic Slowdown –. Retrieved from https://rhg.com/research/spread-thin-chinas-science-and-technology-spending-in-an-economic-slowdown/

70 | Stanford HAI (2025). Policy Implications of DeepSeek AI's Talent Base. Retrieved from https://hai.stanford.edu/policy/policy-implications-of-deepseek-ai-talent-base

71 | "The Global AI Talent Tracker 2.0" (n.d.). Retrieved June 24, 2025, from MacroPolo https://archivemacropolo.org/interactive/digital-projects/the-global-ai-talent-tracker/

72 | "2025 AI技术人才供需洞察报告 (2025 AI Tech Talent Supply and Demand Insights Report)" (2025). Retrieved June 13, 2025, from East Money Information https://web.archive.org/web/20250613085006/https://data.eastmoney.com/report/zw_industry.jshtml?infocode=AP202503061644099941

73 | Feldgoise, Jacob, Aiken, Catherine, Weinstein, Emily S., & Arnold, Zachary (2023). "Studying Tech Competition through Research Output: Some CSET Best Practices." Center for Security and Emerging Technology. Retrieved from https://cset.georgetown.edu/article/studying-tech-competition-through-research-output-some-cset-best-practices/CSET,

74 | Ibid.

75 | "Country Activity Tracker (CAT): Artificial Intelligence, China and United States" (2024). Retrieved April 30, 2025, from Emerging Technology Observatory, Center for Security and Emerging Technology https://cat.eto.tech/?countries=China%20%28mainland%29%2CUnited%20States&countryGroups=&expanded=Summary-metrics%2CCountry-co-authorship CSET ETO

76 | "EU launches InvestAI initiative to mobilise €200 billion of investment in artificial intelligence" (2025). Retrieved June 13, 2025, from European Comission https://digital-strategy.ec.europa.eu/en/news/eu-launches-investai-initiative-mobilise-eu200-billion-investment-artificial-intelligence

77 | "DeepSeek, a National Treasure in China, is Now Being Closely Guarded" (2025). Retrieved June 13, 2025, from The Information https://www.theinformation.com/articles/deepseek-national-treasure-china-now-closely-guarded

78 | Beraja, Martin, Yang, David Y, & Yuchtman, Noam (2023). "Data-intensive Innovation and the State: Evidence from AI Firms in China." The Review of Economic Studies, 90(4), 1701–1723. https://doi.org/10.1093/restud/rdac056

79 | "专家解读 | 畅通数据汇聚、供给、利用堵点 凝力推进数据集高质量建设−国家数据局 (Expert Interpretation | Unblocking Data Aggregation, Supply, and Utilization Blockages and Concentrating on Promoting High-Quality Construction of Data Sets)" (2025). Retrieved June 13, 2025, from National Data Administration https://web.archive.org/web/20250613090140/https://www.nda.gov.cn/sjj/ywpd/szkjyjcss/0306/20250306143724097100325_pc.html '; "大力推动我国人工智能大模型发展 （Vigorously promote the development of large models of artificial intelligence in China）" (2025). Retrieved June 13, 2025, from People's Daily https://web.archive.org/web/20250509065808/https:/paper.people.cn/rmrb/pc/content/202505/08/content_30071810.html?utm_source=substack&utm_medium=email

80 | Arcesati, Rebecca, & He, Alex (2024). "Data Marketplaces and Governance: Lessons from China." Retrieved April 30, 2025, from Centre for International Governance Innovation https://www.cigionline.org/articles/data-marketplaces-and-governance-lessons-from-china/

81 | "专家解读之一 | 大力发展数据标注产业 推动我国人工智能创新发展 (Expert Interpretation Series 1 | Vigorously Developing the Data Labeling Industry to Drive China's AI Innovation and Development)" (2025). Retrieved June 13, 2025, from National Data Administration https://web.archive.org/web/20250613101022/https://www.nda.gov.cn/sjj/zwgk/zjjd/0114/20250114135427836801437_pc.html

82 | "China's AI Development Model in an Era of Technological Deglobalization - IGCC" (n.d.). Retrieved June 24, 2025, from https://ucigcc.org/ https://ucigcc.org/publication/chinas-ai-development-model-in-an-era-of-technological-deglobalization/

83 | Chen, Caiwei (2025). "China built hundreds of AI data centers to catch the AI boom. Now many stand unused." MIT Technology Review. Retrieved from https://www.technologyreview.com/2025/03/26/1113802/china-ai-data-centers-unused/

84 | "DeepSeek Debates: Chinese Leadership On Cost, True Training Cost, Closed Model Margin Impacts" (2025). Retrieved June 13, 2025, from SemiAnalysis https://semianalysis.com/2025/01/31/deepseek-debates/Robertson, Jordan, Hawkins, Mackenzie, & Leonard, Jenny (2025). "US Probing If DeepSeek Got Nvidia Chips From Firms in Singapore." Bloomberg.Com. Retrieved from https://www.bloomberg.com/news/articles/2025-01-31/us-probing-whether-deepseek-got-nvidia-chips-through-singapore

85 | "Exclusive: Nvidia to launch cheaper Blackwell AI chip for China after US export curbs, sources say | Reuters" (2025). Retrieved June 13, 2025, from https://www.reuters.com/world/china/nvidia-launch-cheaper-blackwell-ai-chip-china-after-us-export-curbs-sources-say-2025-05-24/

86 | "Department of Commerce Rescinds Biden-Era Artificial Intelligence Diffusion Rule, Strengthens Chip-Related Export Controls" (2025). Retrieved June 13, 2025, from https://www.bis.gov/press-release/department-commerce-rescinds-biden-era-artificial-intelligence-diffusion-rule-strengthens-chip-related

87 | "Framework for Artificial Intelligence Diffusion" (2025). Retrieved June 13, 2025, from Federal Register https://www.federalregister.gov/documents/2025/01/15/2025-00636/framework-for-artificial-intelligence-diffusionHeim, Lennart (2025). Understanding the Artificial Intelligence Diffusion Framework: Can Export Controls Create a U.S.-Led Global Artificial Intelligence Ecosystem? https://doi.org/10.7249/PEA3776-1

88 | Harithas, Barath (2025). The AI Diffusion Framework: Securing U.S. AI Leadership While Preempting Strategic Drift. Retrieved from Center for Strategic and International Studies https://www.csis.org/analysis/ai-diffusion-framework-securing-us-ai-leadership-while-preempting-strategic-drift

89 | Hawkins, Mackenzie, & Leonard, Jenny (2025). "Trump's Rush to Cut AI Deals in Saudi Arabia and UAE Opens Rift With China Hawks." Bloomberg.Com. Retrieved from https://www.bloomberg.com/news/articles/2025-05-15/trump-s-rush-to-cut-ai-deals-in-saudi-arabia-and-uae-opens-rift-with-china-hawks

90 | Heim, Lennart (2025). China's AI Models Are Closing the Gap—but America's Real Advantage Lies Elsewhere. Retrieved from RAND, https://www.rand.org/pubs/commentary/2025/05/chinas-ai-models-are-closing-the-gap-but-americas-real.html

ABOUT THE AUTHORS

Wendy Chang is an Analyst in the Science, Technology and Innovation at MERICS. Her research interest is on the intersection of technology, politics and the economy. In the age of Artificial Intelligence and increased geopolitical tensions worldwide, she researches how developments in AI, self-driving cars, and other technologies can transform China's digital economy and affect its relationship to the world.

Rebecca Arcesati is Lead Analyst in the Science, Technology and Innovation program, based in MERICS' Brussels office. Her research focuses on China's technology, innovation and digital policy, its artificial intelligence and data strategies, as well as Europe-China innovation relations and US-China technology competition

Antonia Hmaidi is a Senior Analyst in the Science, Technology and Innovation at MERICS. She works on the geopolitics of technology, China's pursuit of tech self-reliance (especially in areas like semiconductors, operating systems and internet infrastructure), China's cybersecurity and hacking campaigns. Hmaidi also develops modelling and big data analysis tools and leads MERICS' data task force.