

Stack battles: the US-China artificial-intelligence rivalry is moving beyond chips alone

China is challenging US leadership in both AI hardware and software, with Europe unlikely to catch up

Alicia García-Herrero, Bertin Martens

Despite Chinese progress, the United States remains for now ahead in the race for dominance over the so-called artificial intelligence hardware stack – the resources and equipment, especially semiconductors, needed to run AI models. US multinational Nvidia is the global market leader with the best-performing AI chips. It accounts for roughly half the world's installed AI chip stock but as much as two thirds of installed computing capacity¹.

China's Huawei is the most credible foreign challenger, but its chips remain technologically inferior². Nevertheless, demand for Huawei's chips within China is strong because of the combination of incentives and restrictions – not least US export controls on advanced chips – that create a captive market. In this context, Nvidia has largely conceded the Chinese AI chip market to Huawei³.

Beijing's demand-side support for Huawei and other domestic chipmakers – through subsidies, cheap energy and procurement preferences – is part of a strategy to close the gap with the US. Another part of this strategy is diplomatic: seeking to tackle US export controls (García-Herrero and Martens, 2026).

But the competition over hardware tells only part of the story. Nvidia's durable advantage also rests on CUDA (Compute Unified Device Architecture), its proprietary software platform that makes its best-in-class chips run efficiently and that has become the *de-facto* standard for AI development. CUDA generates powerful network

effects between chip/infrastructure suppliers, AI model developers, application deployers and end users. More of one type of user attracts more of other types of users. Nvidia makes CUDA freely available to developers – a classic zero-price strategy to maximise adoption and lock in demand for its hardware.

Most AI developers write in PyTorch, the leading open-source machine-learning framework⁴. Because Nvidia dominates the installed chip base, developers optimise for CUDA-compatible hardware. This creates a self-reinforcing cycle: the larger the developer community and the richer the library of pre-optimised code, the harder it becomes for alternatives to gain traction. The result is exceptional pricing power (Nvidia's margins on chips are 70 percent to 80 percent)⁵.

Competing ecosystems face a classic chicken-and-egg problem. Several US players (Amazon, Google, Meta) have developed their own chips but still largely rely on the PyTorch/CUDA ecosystem for broad adoption⁶. The network effects and switching costs (code rewrites, performance uncertainty, loss of optimised libraries) are simply too high for most developers and companies.

Huawei's strategy to bypass Nvidia's software moat

A rival AI stack to Nvidia will require an alternative software ecosystem. Huawei has CUDA equivalents in CANN (Compute Architecture for Neural Networks) and MindSpore, an open-source machine learning framework optimised for its Ascend AI processors⁷. CANN is still maturing and according to developers is less user-friendly than CUDA⁸. Nevertheless, Huawei has moved aggressively to close the usability gap. In August 2025, it announced the open-sourcing of the CANN toolkit – a direct strike at CUDA's proprietary model – and has been recruiting Chinese AI labs, universities and research institutions to contribute to an open Ascend developer community⁹.

Just as important is torch_npu, a backend plugin that lets standard PyTorch code run on Ascend processors. Because most AI developers write in PyTorch, rather than directly in CUDA, this lowers the single biggest switching cost: developers no longer need to abandon their familiar framework to leave Nvidia hardware. The Huawei strategy mirrors, layer by layer, the playbook China has used on the hardware side: open up what the incumbent keeps closed, subsidise adoption and nudge a captive domestic generation of developers towards the alternative stack (García-Herrero and Martens, 2026).

Notably, Huawei collaborates with DeepSeek, one of China's leading AI labs. DeepSeek's models (open-weight and highly competitive on price/performance) have been engineered for compatibility with both Nvidia and Huawei processors. This is strategically clever: it lowers the barriers for developers already familiar with Nvidia ecosystems to try Ascend hardware, while widening the addressable market for Huawei chips. DeepSeek is rapidly becoming one of the most popular AI models globally¹⁰. Unlike some earlier closed efforts by US labs, leveraging popular open-weight models as a bridge appears to be accelerating Huawei's software ecosystem momentum. Whether this can generate self-reinforcing network effects comparable to CUDA's two-decade head start remains the core open question.

The early evidence suggests that the software gap, like the hardware gap, is no longer fixed or unbridgeable. The signs of Chinese catch-up are real: an open-sourced toolkit with a state-backed contributor pipeline, falling switching costs through PyTorch compatibility, flagship open-weight models running on Ascend and a protected domestic market large enough to sustain the ecosystem through its immature phase. None of these existed in meaningful form two years ago.

Nevertheless, obstacles remain formidable. Chinese developers continue to report serious usability problems with CANN¹¹, and CUDA's accumulated stock of optimised libraries is clearly large. Whether China can close the gap is genuinely uncertain.

The strategic implication, however, is clear. If Chinese companies succeed in crossing the CUDA moat – even only within their domestic market – the durable US advantage will narrow. More generally, the dual US hardware and software strength could become less mutually reinforcing.

European absence

Though it has AI assets (such as chip lithography equipment and AI research; see García-Herrero and Martens, 2026), Europe has no AI chip designer of global significance and no software layer analogous to CUDA or CANN that translates high-level AI algorithms into efficient machine code across a broad developer community.

As CUDA and CANN show, the hardware and software layers are closely intertwined through co-design. Without domestic chip-design capability at scale, it is much harder to build and optimise a competitive software platform, because there is less tight

feedback between hardware architects and compiler/runtime developers. The economic value in the AI stack accrues overwhelmingly to those who control chip design and the software platform, not merely to equipment suppliers. With assets such as lithography systems, Europe is, in effect, an indispensable input provider to a race in which it is not itself a full contestant.

However, the absence of a full domestic stack does not mean Europe has no options. Targeted industrial policy can help close the compute gap (García-Herrero and Martens, 2026). Europe could also pursue a more deliberate open-source and vertical-specialisation strategy, with a focus on where it has comparative advantage (examples include automotive AI, climate modelling and trustworthy AI). Whether such a partial strategy can deliver genuine sovereignty or merely reduce dependence is the central strategic question for European policymakers. They should realise that the contest is shifting from who designs the best chips to who controls the full stack, including the software layer.

Reference

García-Herrero, A. and B. Martens (2026) 'Europe needs a strategy to close the artificial intelligence compute gap', *Analysis* 13/2026, Bruegel, available at <https://doi.org/10.64153/HFQM9739>

Endnotes

- 1 See Epoch.ai, global AI chip sales by H100 equivalents, available at https://epoch.ai/data/ai-chip-sales?view=graph&tab=h100_equivalents
- 2 Huawei's most advanced Ascend processors lag Nvidia's, including on software maturity, developer tooling, memory bandwidth optimisation and power efficiency for equivalent workloads. Huawei compensates by packing large numbers of processors closely together but this comes at a significant energy cost (estimates suggest two to four times more energy for equivalent output). See for example, Anton Shilov, 'Huawei's new AI CloudMatrix cluster beats Nvidia's GB200 by brute force, uses 4X the power', Tom's Hardware, 18 April 2025, <https://www.tomshardware.com/tech-industry/artificial-intelligence/huaweis-new-ai-cloudmatrix-cluster-beats-nvidias-gb200-by-brute-force-uses-4x-the-power>.
- 3 Lee Ying-Shan (2026) Nvidia says it has 'largely conceded' China's AI chip market to Huawei, CNBC Tech, 20 May 2026. <https://www.cnbc.com/2026/05/21/nvidia-jensen-huang-china-ai-chip-market-huawei.html>
- 4 PyTorch bridges between machine learning commands and CUDA commands that are understood by Nvidia AI chips. PyTorch thus remains deeply connected to CUDA. PyTorch is governed by members of the Linux Foundation, including Nvidia, Meta, Microsoft, Google, Amazon, AMD, Intel and Huawei. See <https://pytorch.org/>.
- 5 Elongated Musk, 'Who Actually Earns the GPU Dollar?' Medium, 16 December 2025, https://medium.com/@Elongated_musk/who-actually-earns-the-gpu-dollar-5a2756a06bb1. A legitimate competition concern arises from Nvidia's dominance – high concentration and switching costs – but a full analysis of antitrust implications lies outside the scope of this analysis.
- 6 Dylan Sloan, 'Meta and Google announce new in-house AI chips, creating a "trillion-dollar question" for Nvidia', Fortune, 11 April 2024, <https://fortune.com/2024/04/11/meta-google-ai-chips-semiconductor-in-house-nvidia-trillion-dollar-question/>.
- 7 Lily Ottinger and Mary Clare McMahon, 'Can Huawei Take On Nvidia's CUDA? The true US v China AI battle', China Talk, 5 May 2025, <https://www.chinatalk.media/p/can-huawei-compete-with-cuda>.

- 8 Michael Parekh, 'AI: Huawei's AI gains in China over Nvidia. AI-RTZ #1076', AI: Reset to Zero, 4 May 2026, <https://michaelparekh.substack.com/p/ai-huaweis-ai-gains-in-china-over>.
- 9 Xinmei Shen, 'Tech war: Huawei to open-source AI chip toolkit to take on Nvidia's proprietary platform', South China Morning Post, 5 August 2025, <https://www.scmp.com/tech/tech-war/article/3320852/tech-war-huawei-open-source-ai-chip-toolkit-take-nvidias-proprietary-platform>.
- 10 See Openrouter, 'Text Model Rankings', updated 8 June 2026, <https://openrouter.ai/rankings/text>.
- 11 Eleanor Elcott, 'Huawei's bug-ridden software hampers China's efforts to replace Nvidia in AI', Financial Times, 3 September 2024, <https://www.ft.com/content/3dab07d3-3d97-4f3b-941b-cc8a21a901d6>.

Bruegel

Rue de la Charité

33,

B-1210 Brussels

(+32) 2 227 4210

info@bruegel.org

Bruegel.org