US: powerful AI jet stuck in pilot approvals. China: rougher engine racing into factories on subsidies.

US vs China AI: Who Actually Deploys It

Syed Umair Shoaiby
Technical Program Manager at Cirrus Logic | Leading Semiconductor Silicon NPI from Tape-out to Production | AI...

June 8, 2026

The United States just put its most powerful AI models on every federal desk for 42 cents.

China is paying its small companies up to 80% of the bill to actually use theirs.

Same race. Two completely different bets. And after watching how programs actually move from a working demo to a deployed workflow, I think the second bet is the one quietly winning.

Both countries are now racing to spread AI, not just build it. They are simply enabling it from opposite ends.

The US bet is supply-side. Make the best capability cheap and available, and trust that adoption follows. The GSA struck a deal that hands any federal agency a frontier model for 42 cents for eighteen months. Federal AI contract obligations hit 7.2 billion dollars in 2026, up roughly 966 percent from 2024. The theory is clean. Lower the price of access to near zero and the rest takes care of itself.

The China bet is demand-side. Do not just supply the tool. Pay people to put it into production. Local governments in Beijing, Shanghai, Shenzhen and others are issuing computing-power vouchers that cover up to 80% of an AI bill for smaller firms. Shanghai alone put around $84 million behind it. On top of that, the 15th Five-Year Plan names AI more than fifty times

and explicitly backs open-weight models, so the cheap, deployable option is the one the state is pushing.

And the deployable option is genuinely cheap. The cluster of Chinese open-weight models released this spring runs several times cheaper to operate than the leading US frontier model, on the order of three to seven times cheaper per token, at benchmark scores that sit one tier below, not a generation below. For most real workflows, one tier below at a fraction of the cost is the better business decision, not the worse one.

So what shows up on the factory floor? Roughly twice the deployment. The cleanest single figure here is aggregator-sourced, so I will not lean my whole case on it, but the direction is corroborated everywhere. A large share of Chinese industrial firms are running AI in production while a large share of comparable US firms are still stuck in what people are openly calling pilot purgatory. You can see it in the field. JD Logistics is offering 12-hour delivery in core cities. Cainiao cut cross-border delivery times by about half. That is fielded throughput, not a slide.

Now the part I want to be honest about, because the easy version of this post is a China-wins take, and that version is wrong.

China is winning the rollout on a fragile base. Its chip self-sufficiency was only about a third in 2024, and the domestic processes underneath all this deployment still run at yields reported below a third of what TSMC gets. The 80% self-sufficiency goal for 2030 still depends on tooling China cannot make yet. Cheap, subsidized, low-yield compute pushed into fast open-weight deployment buys speed now and borrows against quality, durability, and governance later.

The US gap is the mirror image. A real frontier-model and compute lead that keeps stalling at the one step that actually creates value. Two countries. Two different gaps. One is deploying faster than its foundation can safely carry. The other is sitting on the best foundation and not deploying.

Read at program level, this is the research-versus-deployment-reality gap I keep writing about, just blown up to the size of a country. Capability is not the binding constraint for either side. Diffusion is. The winner is not whoever has the smartest model or the best fab roadmap. It is whoever moves capability into real, owned, maintained workflows the fastest and keeps it running.

That is the same thing I watch decide programs at much smaller scale. The team that wins is almost never the one with access to the better tool. Everyone has access now. It is the one that does the unglamorous work of closing the distance between a pilot that worked once and a workflow somebody actually owns on Monday.

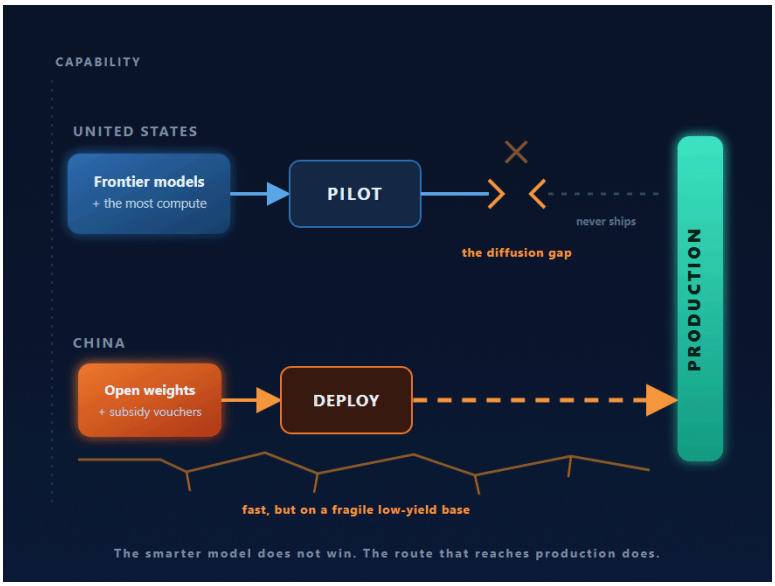
I have seen the failure version of this up close. A genuinely useful automation built for an NPI program, demoed clean, everyone nodded. Then it sat. No owner. No place in the standard build flow. No one accountable for it when an input changed. Six months later people were back to doing it by hand, not because the tool was wrong, but because nobody ran the diffusion. The teams that get it right treat that last mile as the actual project. They name an owner before the pilot, wire it into an existing ritual people already do, and budget for the boring maintenance, so the thing survives contact with a real quarter.

Access is solved. Diffusion is the job.

So here is my question for the people who run this for a living. In your organization, what is the real thing standing between a pilot that worked and a workflow in production? Is it the tool, or is it everything around the tool that nobody wants to own?

Disclaimer: The views shared here are my own and do not represent the positions of my employer or any organization I am affiliated with. Company and country examples are drawn from public reporting and are used for illustration, not as commentary on any specific business or government. All figures come from the sources linked with this article and reflect the data available at the time of writing.

#AI #PMO #ProgramManagement #Semiconductor



US: best models stall at pilot. China: open models reach production on a cracked base.

REFERENCE GUIDE: WHERE EACH CLAIM COMES FROM

- **"The GSA struck a deal that hands any federal agency a frontier model for 42 cents for eighteen months"**

GSA-xAI OneGov agreement: Grok 4 and Grok 4 Fast available to any federal agency for $0.42 for 18 months (through March 2027).

https://www.gsa.gov/about-gsa/newsroom/news-releases/gsa-xai-partner-to-accelerate-federal-ai-adoption-09252025

- **"Federal AI contract obligations hit 7.2 billion dollars in 2026, up roughly 966 percent from 2024"**

Brookings, Where does federal AI spending stand in 2026. Potential awards run up to $91.8 billion.

https://www.brookings.edu/articles/where-does-federal-ai-spending-stand-in-2026/

- **"Computing-power vouchers that cover up to 80 percent of an AI bill for smaller firms; Shanghai alone put around 84 million dollars behind it"**

Tom's Hardware, on China's computing-power vouchers across Beijing, Shanghai, Shenzhen, Chengdu and others. Shanghai allocated roughly CN 600 million (about $84 million).

https://www.tomshardware.com/tech-industry/artificial-intelligence/china-subsidizes-ai-computing-for-small-domestic-companies-computing-power-vouchers-spread-across-multiple-chinese-cities

- **"The 15th Five-Year Plan names AI more than fifty times and explicitly backs open-weight models"**

IISS analysis of China's 15th Five-Year Plan (2026-2030): AI mentioned 52 times with a standalone AI+ action plan. The State Council AI Plus Opinions (approved July 2025, via CSET) back an open-source AI ecosystem.

https://www.iiss.org/online-analysis/online-analysis/2026/03/chinas-15th-five-year-plan/

https://cset.georgetown.edu/publication/china-ai-plus-opinions-2025/

- **"Chinese open-weight models run several times cheaper to operate than the leading US frontier model, on the order of three to seven times cheaper per token, at benchmark scores one tier below"**

Artificial Analysis: the spring 2026 Chinese open-weight cluster (DeepSeek V4 Pro, Kimi K2.6, GLM-5.1, MiniMax M2.7). DeepSeek V4 Pro scores 52 on the Intelligence Index, the number-two open-weight reasoning model, and is priced at about $1.74/$3.48 per million input/output tokens. Claude Opus 4.7 runs about $5/$25 per million tokens (per MindStudio's pricing comparison), so V4 Pro is roughly three times cheaper on input and about seven times cheaper on output. Artificial Analysis puts V4 Pro at more than four times cheaper on total benchmark cost. The body claim is stated as three to seven times to stay inside what these sources support.

https://artificialanalysis.ai/articles/deepseek-is-back-among-the-leading-open-weights-models-with-v4-pro-and-v4-flash

https://www.mindstudio.ai/blog/deepseek-v4-vs-gpt-55-vs-claude-opus-47-pricing

- **"Roughly twice the deployment" / "pilot purgatory"**

AI Frontiers, China and the US Are Running Different AI Races, citing an industrial-adoption split of about 67 percent of Chinese firms versus about 34 percent of US firms. NOTE: this exact split traces to an aggregator (SecondTalent), not a first-tier survey, so it is stated as "roughly twice" and not leaned on alone. Direction corroborated by Capital Economics (China's

AI rollout could rival the US, Jan 14 2026) and PIIE's US Census BTOS data showing large US-firm adoption still in low double digits.

https://ai-frontiers.org/articles/china-and-the-us-are-running-different-ai-races

https://www.piie.com/blogs/realtime-economics/2026/adoption-ai-industrial-sectors

- **"JD Logistics is offering 12-hour delivery in core cities. Cainiao cut cross-border delivery times by about half"**

AI Frontiers, China and the US Are Running Different AI Races. Used as concrete examples of fielded operational AI.

https://ai-frontiers.org/articles/china-and-the-us-are-running-different-ai-races

- **"China's chip self-sufficiency was only about a third in 2024" and "the 80 percent self-sufficiency goal for 2030"**

TrendForce, on China's reported 80 percent chip self-sufficiency target by 2030. Self-sufficiency stood at about 33 percent in 2024, and the domestic-equipment line still depends on tooling China cannot yet make.

https://www.trendforce.com/news/2026/03/31/news-china-reportedly-targets-80-chip-self-sufficiency-by-2030-eyes-domestic-7nm-line-and-14nm-production-stability/

- **"Domestic yields reported below a third of TSMC's"**

Mark Lapedus (Semiconductor Ecosystem), Can China make 5nm chips. SMIC's 7nm yields are reported around 20 to 40 percent and 5nm lower still, against TSMC's 80 percent-plus, putting domestic advanced-node yields at roughly a third of TSMC's or below.

https://marklapedus.substack.com/p/can-china-make-5nm-chips

- **The US-supply-side-versus-China-demand-side framing**

Brookings, Competing AI strategies for the US and China, background framing of the two strategies.

https://www.brookings.edu/articles/competing-ai-strategies-for-the-us-and-china/

- **The pilot-to-production framing and the "diffusion is the job" practitioner take** are the author's own lived program-management experience, illustrative and not externally cited.

Comments

Add a comment...

No comments, yet.

Be the first to comment.

Start the conversation

Syed Umair Shoaiby

Technical Program Manager at Cirrus Logic | Leading Semiconductor Silicon NPI from Tape-out to Production | AI Enabled Program Execution, Workflow Automation & Operational Strategy | Scaling Cross Functional Programs