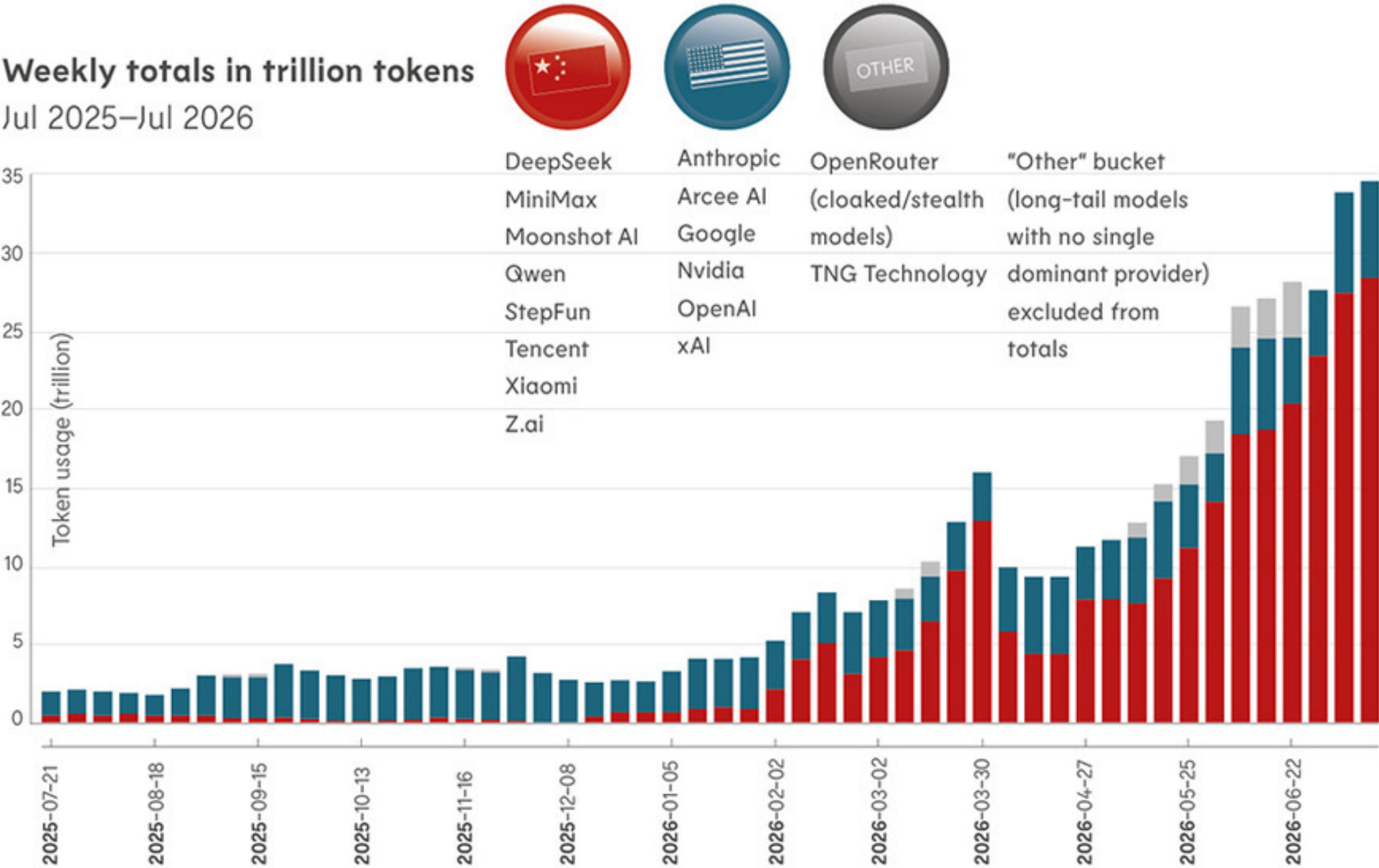


China's cheap AI flood:

DeepSeek and rivals overtook U.S. models



Weekly totals in trillion tokens
Jul 2025–Jul 2026



Source: OpenRouter model-rankings data

Weekly Facts & Figures.

- **Model adoption:** Chinese AI models surpassed U.S. models in weekly token usage on OpenRouter in Feb 2026 and now account for >4 times the token volume.
- **Pricing gap:** DeepSeek V4 Flash is priced at 0.28 USD per 1M output tokens, versus 25 USD for Claude Opus 4.8 (about 89 times higher).
- **Price-war escalation + scrutiny:** Free tiers (Tencent Hy3, Nvidia Nemotron 3 Ultra) drove a ~150% weekly token-volume spike; in parallel, U.S. lawmakers opened a review and China's MIIT alleged a Claude Code "backdoor."

What This Means.

"Token economics is something companies have to reckon with. If the model is good enough, more for less wins. For Europe market, the deciding factor will be whether Chinese providers can solve compliance and deploy on European servers."



Yuqi Cheng
Data & AI Engineer

Start the conversation.



yuqi.cheng@sinolytics.de



sinolytics.de