44 Trillion Tokens on Chinese AI Chips: The Inference Race Is On

2026-08-27T10:42:08.14997Z

44 Trillion Tokens on Chinese AI Chips: The Inference Race Is On

body{font-family:-apple-system,BlinkMacSystemFont,'Segoe
UI',sans-serif;max-width:680px;margin:0 auto;padding:24px;color:#111;line-height:1.7}
h1{font-size:24px;font-weight:700;margin-bottom:8px;line-height:1.3}
.hero{width:100%;border-radius:8px;margin:16px 0 28px;display:block}
p{margin:0 0 18px}
a{color:#0057ff}
.source{font-size:13px;color:#888;margin-top:32px;border-top:1px solid
#eee;padding-top:16px}

44 Trillion Tokens on Chinese AI Chips: The Inference Race Is On

We were all stress testing the new Chinese AI chips!

44 Trillion Tokens in a few days only using Chinese AI Chips!

This is just opencode, add OpenRouter and other providers to the list!

This seems to be the largest inference experiment by a non-US lab!

I'm waiting to see how the hedge fund managers pick this up and see NVIDIA's impact on it.

All the US Sanctions might have forced these Chinese model labs to be the least dependent on Nvidia.

Couple this with OpenAI's latest move with Jalapen?o's OpenAI chip for faster inference.

NVIDIA might have ruled the training market, but the companies are hedging for inference!

https://myspacexai.site/issues/44-trillion-tokens-on-chinese-ai-chips-the-inference-race-is-on