Anthropic Pulls Ahead, China's Efficiency Moat Deepens, and Infrastructure Bets Multiply

VC Tech Radar

2026-05-14

Anthropic Pulls Ahead, China's Efficiency Moat Deepens, and Infrastructure Bets Multiply

By VC Tech Radar • May 14, 2026

A light funding day still surfaced a $7M air-traffic systems round, new defense procurement, and seed-fund allocation signals. The larger story was the mix of standout infrastructure and vertical AI teams, agent-runtime and multimodal breakthroughs, and market shifts around China's efficiency edge, Anthropic's enterprise adoption lead, and application-layer moats.

1) Funding & Deals

- **Air traffic systems:** Eric Jacob Button announced a **$7M** round led by Initialized to build next-generation air traffic systems, with United Airlines, Y Combinator, and other investors participating. Garry Tan said YC was the first investor. [1, 2]
- **Defense procurement is emerging as a financing substitute.** The DoD signed framework agreements for low-cost containerized missiles with Anduril, CoAspire, Leidos, and Zone 5, and for low-cost hypersonic missiles with Castelion. The department said it intends to procure **10,000+** LCCMs over three years starting in 2027 and, after testing and validation, award Castelion a multi-year contract for at least **500 Blackbeard missiles annually** while seeking authorization for **12,000+** missiles over five years. [3]
- **Allocator signal:** Scribble Ventures says it is now on its **third fund** with nearly **$300M AUM**, still writing roughly **$750k-$1.5M** initial checks at pre-seed and seed. Founder Elizabeth Weil positioned the firm around AI-native companies and top 1% founders, sourced through an operator-heavy network spanning OpenAI, Meta, Twitter, Instagram, and a16z. [4]

2) Emerging Teams

- **Heron Power:** Heron is building solid-state transformers that use silicon and software to replace steel, oil, and copper in power conversion for data centers, solar, and battery projects. The company is targeting a grid that Drew Baglino described as largely unchanged for a century, and its first large factory is expected to create around **500 jobs**. [5]
- **Mariana Minerals:** Mariana is a software-first mining and refining company building Capital Project OS for agentic project delivery, PlantOS for refinery autonomy, and MineOS for autonomous mining control. It already operates a copper mine in southeast Utah, is building a lithium refinery in Texas, and is targeting **10 projects in 10 years**. [5]
- **Foresight:** YC says Foresight builds AI-powered consumer simulations for CPG, retail, and tech teams, and reports **95% accuracy** versus traditional research in tests with Fortune 500 clients. [6]
- **Adialante:** The YC team says its mobile MRI model can reduce scans to **hundreds per scan** and wait times to **hours**, with the explicit aim of making annual cancer screening routine. [7]
- **Rudus:** Rudus is going after a painful construction workflow: concrete takeoffs that can require **100+ hours** of manual tracing per bid. YC says its AI platform lets teams bid on **3-5x** more work without adding headcount. [8]
- **Surtr Defense:** Surtr's ParallaxOS is pitched as an open operating system for drone defense, unifying any sensor into one threat picture with AI fire control, while keeping integrations with partners and data with customers. [9]

3) AI & Tech Breakthroughs

- **Multimodal efficiency is getting materially better.** A new open **30B-parameter** multimodal model processes images, video, and audio at almost **10x real-time video**, about **3x faster** than Q3 Omni on video and up to **7x faster** on documents. The gains come from Mamba layers that scale linearly with context length, direct audio tokenization that preserves emotion and tone, 3D convolutions over frame blocks, a distilled vision encoder, and efficient video sampling that removes duplicate frames. [10]
- **Agent infrastructure is hardening into a real stack.** LangChain launched **SmithDB**, a database built for agent trace data, and **LangSmith Engine**, which sits on top of traces to identify issues and suggest fixes like code changes or additional evaluators. Perplexity, meanwhile, described an agent runtime with hardware-isolated sandboxes per task, proxy tokens instead of raw API keys, safety detection on accessed content, encrypted connector data, and separated storage and compute. [11, 12, 13, 14]
- **Voice generation is separating identity from performance.**

Scenema Audio released open-source weights and inference code for zero-shot expressive voice cloning based on the idea that emotional performance and voice identity are independent. The team argues diffusion-based speech sounds more natural and less robotic than autoregressive TTS, especially for emotional delivery, and it is already being used in audio-first video workflows. [15]

- **Web agents are getting resettable training environments.** WebHarbor packages **15** popular websites into local Flask + SQLite apps inside one Docker image, resets them to byte-identical state in **<1 second**, and supports all **643 WebVoyager tasks** out of the box—explicitly solving live-web issues like reCAPTCHA, geo-blocks, content drift, and non-resettable environments for RL training. [16]
- **Physical AI demos are getting longer and less scripted.** Figure Robotics showed humanoid robots running a full **8-hour shift** at human performance levels, fully autonomously on **Helix-02**. [17]

4) Market Signals

- **China's competitive threat is now efficiency, not just catch-up.** Exponential View estimates Chinese labs are extracting **4-7x** more intelligence per unit of compute than naive scaling would predict, despite being **2-3 years** behind the US in compute and facing an **8x** US lead in deployed capacity. Even so, their models are described as only **3-8 months** behind the US frontier on benchmarks, with much cheaper inference and reported **50-70%** gross margins at some providers. [18]
- **Enterprise model share just flipped.** Ramp AI Index data cited on X shows **34.4%** of businesses using Anthropic versus **32.3%** using OpenAI; Anthropic adoption quadrupled over the last year while OpenAI rose only **0.3%**. Nathan Benaich added that revenue comparisons may also be affected by model **4.7** using up to **1.3x** more tokens for the same query. [19, 20]
- **Application-layer defensibility is moving down-stack.** a16z argues that as incumbents such as Salesforce open APIs and ship headless products, they are implicitly betting the data layer—not the UI—retains value. In that world, startups compete on proprietary data, control of the action layer, real-world execution, and selling to technical buyers, while next-generation systems of record capture context, initiate work, and record data exhaust. [21, 22]
- **Enterprise knowledge work is already landing significant task volume.** PayPal runs **74,000 weekly tasks** in Perplexity Enterprise across model validation, channel performance, market trend research, competitive intelligence, and product analysis. [23]
- **Investor sentiment remains split between platform risk and productivity upside.** Jason Calacanis is explicitly asking how many startups survive Anthropic and OpenAI, and Dalton Caldwell says the answer is not a simple "no," pointing to moats and the danger of building on the

assumption that models will not improve. Marc Andreessen amplified the counterview that firms have incentives to blame AI for layoffs, while GitHub commit activity has risen sharply and the industry is not seeing layoff "armageddon." [24, 25, 26]

> "It's hard to believe you are prepared for a problem you haven't spent any time considering could be a problem" [27]

5) Worth Your Time

- Exponential View's China AI lab analysis is the most detailed item in the batch on how export controls may have created an efficiency moat, with concrete numbers on compute lag, benchmark distance, pricing, and margins. [18]
- Seema Amble's a16z essay on headless software is useful for diligencing application moats as systems of record become agentic and value shifts toward data and execution. [22]
- The a16z conversation with Heron Power and Mariana Minerals is the clearest video in the batch on why the AI bottleneck is increasingly physical infrastructure—materials, energy, and grid capacity—not just models and chips. [5]

The Founders Who Left Tesla to Rebuild America | a16z (0:50)

- Sequoia's Suno interview is worth watching for a founder-level explanation of why music generation required modeling raw audio, why the team chose full songs over short clips, and why usage

looks more like creative participation than passive listening. [28]

Suno's Mikey Shulman: Everyone Can Make Music Now (28:26)

- LangSmith Engine and SmithDB are the most relevant product posts in the batch if you track agent reliability tooling and the growing importance of traces as a control plane. [12, 11]

Sources

1. X post by @ericjacobbutton
2. X post by @garrytan
3. X post by @DoWCTO
4. Elizabeth Weil (Scribble Ventures) - A People-Centered Journey [Entire Talk]
5. The Founders Who Left Tesla to Rebuild America | a16z
6. X post by @ycombinator
7. X post by @ycombinator
8. X post by @ycombinator
9. X post by @ycombinator
10. NVIDIA New AI Is An Efficiency Monster
11. X post by @hwchase17
12. X post by @hwchase17
13. X post by @perplexity_ai
14. X post by @AravSrinivas
15. r/MachineLearning post by u/a___side_of_fries

16. r/MachineLearning post by u/ArtichokeHelpful7462
17. X post by @Figure_robot
18. Inside the Chinese AI labs where America's AI controls created its toughest competiton
19. X post by @arakharazian
20. X post by @nathanbenaich
21. X post by @a16z
22. X post by @a16z
23. X post by @perplexity_ai
24. X post by @Jason
25. X post by @Standard_Cap
26. X post by @perrymetzger
27. X post by @daltonc
28. Suno's Mikey Shulman: Everyone Can Make Music Now