World Models, Agent Sandboxes, and New Vertical AI Wedges

VC Tech Radar

2026-05-22

World Models, Agent Sandboxes, and New Vertical AI Wedges

By VC Tech Radar • May 22, 2026

The clearest signals this cycle are a concrete shift toward world models, strong traction in agent infrastructure, and a fresh set of vertical AI startups in fraud, pathology, and human-agent coordination. The brief also covers new financing structures, frontier-model economics, and a short list of source material worth reading or watching.

Funding & Deals

- **OpenAI's YC token-for-equity program is still the clearest financing experiment in this batch.** Sam Altman offered $2M in OpenAI tokens to every YC startup in the current batch in exchange for equity; YC separately said the offer covers the spring and summer batches and extended the summer deadline to May 25. External commentary framed the tokens as compute credits that can de-risk early product work and may lift valuations at the margin. [1, 2]
- **Round mechanics are lagging company growth.** Harry Stebbings said founders are agreeing terms around $3M ARR and reaching $15M-$20M ARR by the time legals finish, with company progression outpacing legal completion. [3]
- **A small angel round came with a clear product lesson.** An AI video editor founder said they raised $30K two weeks earlier, then learned from 10 beta users that the real pain was workflow speed, not output fidelity. They responded by cutting validation from 18 gates to 5, limiting retries, and moving to a preview-first flow. [4]

Emerging Teams

- **Daytona: stateful compute for agents.** CEO Ivan Burazin previously co-founded CodeAnywhere, used by about 3 million people, and later ran developer experience at InfoBip. After a January 2025 pivot from human dev environments to agent sandboxes, Daytona reported 74% month-over-month growth; one customer runs about 850K sandboxes a day; RL/eval workloads moved from 0% to roughly 50% of usage. The system runs on bare metal with its own scheduler, using local NVMe snapshots to start one sandbox in about 60 ms or 50,000 in about 75 seconds. [5]
- **Incandor: behavioral intelligence for bank fraud.** YC says the product links behavior across accounts, making fraud rings, mule handoffs, and banned operators visible. Founders are Matthew Yekell and Luc Rosenzweig. [6]
- **Limrun: mobile development infrastructure for cloud agents.** The product provides remote Xcode plus iOS and Android simulators so cloud agents can build mobile software; YC says customers already include Replit, Rork, and Momentic AI. Founder: @muvaff. [7]
- **Voquill: voice AI for pathologists.** Voquill listens while pathologists work and drafts sign-out-ready reports in real time, targeting a workflow where many pathologists spend more time writing reports than diagnosing. Founders are @HenryHabibAI, @josiahsrc, and Michael. [8]
- **Human-agent coordination is becoming a software layer.** Pentagon, launched by @edgarpavlovsky, argues that agents are already doing coding, research, ops, and customer work but still operate in isolation, turning humans into middleware. Lightsprint is attacking the adjacent problem with a platform for visual planning, parallel cloud agents, live previews, and more reliable shipping. [9, 10]

AI & Tech Breakthroughs

- **World models are moving from research rhetoric into startup formation.** Yann LeCun said he founded Advanced Machine Intelligence to pursue world models and physical AI beyond LLMs, predicted 2026 will be "the year of the world model," and argued that LLM-style next-token architectures do not work for video, sensor, or biological data because there are infinitely many plausible next states. Fei-Fei Li said World Labs is building foundation models for spatial intelligence, with world models and world action models that learn from pixels to generate states, policies, and actions for robots and physical systems. Bioptimists is applying similar beyond-language ideas to biology with multimodal, multiscale models aimed at drug discovery and rational medicine design. [11, 12]

  "I think 2026 is going to be the year of the world model" [11]

- **OpenAI's unit-distance result is a real symbolic milestone.** An OpenAI model discovered a new family of constructions for the planar unit distance problem, outperforming square-grid-based approaches and disproving a belief held since Erdős posed the problem in 1946. Multiple sources framed it as the first time AI autonomously solved a prominent open problem central to a field of mathematics; one account said the model connected geometry to deep number theory, and experts including Noga Alon, Melanie Wood, and Tim Gowers called it "a milestone in AI mathematics." [13, 14]
- **Runway is productizing a stronger video-editing primitive.** Aleph 2.0 lets users edit a single frame, preview the change, and propagate that edit through the rest of the video inside the web-based Edit Studio. Cristóbal Valenzuela said Aleph 1.0 had already changed editing workflows and positioned 2.0 as a new standard for the category. [15, 16]
- **Fast inference remains one of the few infrastructure advantages users immediately feel.** Cerebras said its wafer-scale AI systems are 15-20x faster than GPUs at inference and are built around a 46,000 square millimeter chip. CEO Andrew Feldman said demand accelerated in 2025 once models became useful enough for everyday work, and argued that speed opens new business models rather than just marginal efficiency gains. [17]

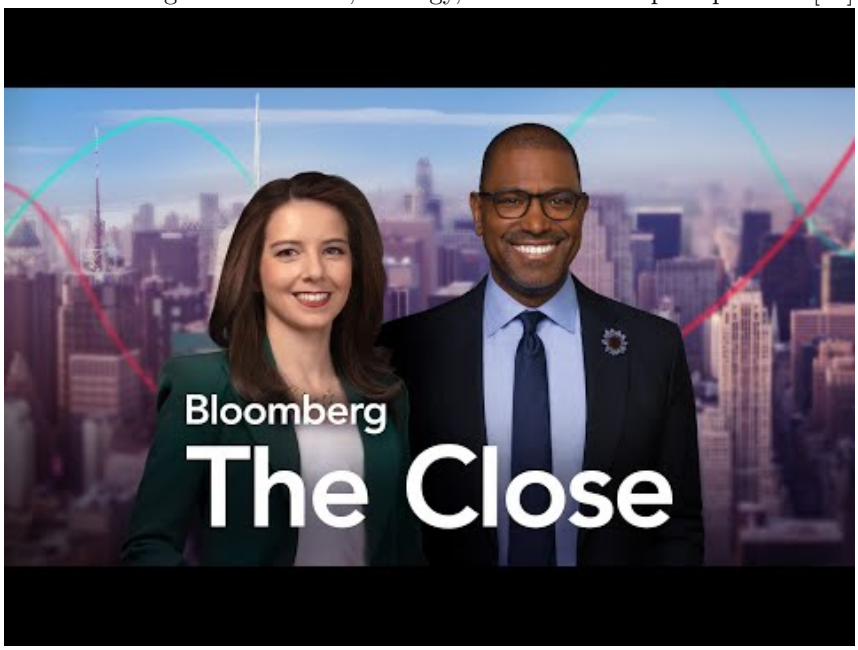
Market Signals

- **AI adoption is now showing measurable GTM leverage.** ICONIQ/SaaStr data says companies with AI fully embedded in GTM generate roughly 2x the net new revenue per FTE of medium and low adopters. AI-heavy pipelines also show better top-of-funnel conversion: new lead to MQL is 38% versus 27%, and MQL to SQL is 37% versus 29%. Daily AI use passed 50% in marketing, SDR/BDR, and RevOps. [18]
- **Returns still appear concentrated at the frontier, and the supporting stack is expensive.** In recent conversations cited by Patrick O'Shaughnessy, Anthropic's Krishna, Dylan Patel, and Gavin Baker all argued that frontier models capture most economic returns at the model layer; Krishna said customers spend heavily on newer models because frontier intelligence drives meaningful ROI. Sarah Guo added that this is a capex-intensive cycle, that Nvidia is 2-5 years ahead in areas like neoclouds and inference cloud, and that startups still want frontier chip performance because it enables products such as current coding agents. [19, 20, 21]
- **The geopolitics of open versus closed models are shifting.** Fei-Fei Li said the 2026 AI Index shows the US-China capability gap has closed for the first time; she added that China now leads in open LLMs, video models, and even world models, while the US is closing models. [12]

- **Efficiency gains are real, but the energy map will get more complicated.** Fei-Fei Li said inference costs for language models fell about 280x in the last two to three years through distillation, quantization, and newer chips. At the same time, she said AI's current power buildout is being driven by training and inference on language models, while embodied AI will eventually add a much more distributed pattern of on-machine compute and energy demand. [12]
- **Early-stage distribution is being subsidized with tokens, and that may invite backlash.** Harry Stebbings said token spend is becoming a core marketing line item, with founders willing to give away $20K-$50K per month in tokens to drive usage and temporarily out-hustle incumbents. He also warned that layoffs and capital shifts into machines are creating a political and social backlash the tech industry is underestimating. [22, 23]

Worth Your Time

- **Yann LeCun on world models and physical AI** — Best single watch in this batch if you want the technical case for why LLMs are not enough for robotics, biology, or real-world perception. [11]



Spacex's IPO & the Future of the AI Build-Out | The Close 5/21/2026 (68:49)

- **Giving Agents Computers — Ivan Burazin, Daytona** — Strong diligence material on agent sandboxes, why Kubernetes breaks for this

workload, and what real product pull looks like in infra. [5]

- **Fei-Fei Li at the Stanford Sustainability Forum** — Useful for the 280x inference-cost decline, the world-model roadmap, and the energy consequences of embodied AI. [12]

Stanford Sustainability Forum | Powering the AI Revolution (6:55)

- **OpenAI's planar unit distance thread** — Primary-source summary of the math result and what changed relative to long-standing square-grid intuition. [13]
- **Runway Aleph 2.0 demo** — Quick product demo of a potentially important editing primitive: change one frame, then propagate the edit across the clip. [15]

Sources

1. Andrej Karpathy Joins Anthropic | SpaceX Files S1: How Does it Trade | Cerebras Smashes Day 1
2. X post by @ycombinator
3. X post by @HarryStebbings
4. r/SaaS post by u/kp160200
5. Giving Agents Computers — Ivan Burazin, Daytona
6. X post by @ycombinator
7. X post by @ycombinator
8. X post by @ycombinator
9. X post by @ycombinator

10. X post by @ycombinator
11. Spacex's IPO & the Future of the AI Build-Out | The Close 5/21/2026
12. Stanford Sustainability Forum | Powering the AI Revolution
13. X post by @OpenAI
14. X post by @bindureddy
15. X post by @runwayml
16. X post by @c_valenzuelab
17. The Story Behind Cerebras' $63 Billion IPO with Founder and CEO Andrew Feldman
18. The Modern GTM Org in 2026: 20-30% Leaner, 9x Flatter, ~2x More Net New Revenue Per Rep. The Latest from ICONIQ Growth.
19. X post by @patrick_oshag
20. X post by @saranormous
21. X post by @saranormous
22. X post by @HarryStebbings
23. X post by @HarryStebbings