

10x Science Seeds, Perplexity Ships a New Qwen Stack, and Agent Infrastructure Matures

VC Tech Radar

2026-04-23

10x Science Seeds, Perplexity Ships a New Qwen Stack, and Agent Infrastructure Matures

By VC Tech Radar • April 23, 2026

A new AI-biotech seed round leads this brief, alongside emerging teams in coding-agent QA, government workflow automation, and auditable research infrastructure. The broader pattern is clear: post-training, document AI, and agent tooling are improving quickly, while API quality, trace data, and workspace agents are becoming core investment questions.

1) Funding & Deals

- **10x Science — \$4.8M seed.** YC highlighted a \$4.8M seed for 10x Science. The company is building AI for molecular-level protein characterization, compressing a workflow that currently requires specialized scientists to spend weeks or months manually interpreting complex data into one that delivers insights in minutes [1].
- **Strategic deal watch: SpaceX x Cursor.** Jason Calacanis reported that SpaceX tapped Cursor to compete with Claude Code [2]. In commentary on the deal, Anand Nandkumar argued the more important question is whether Cursor's developer traces are the scarce input in frontier coding models, while Clement Delangue used the moment to call for open traces so open agent models can be trained too [3, 4].

2) Emerging Teams

- **Autosana.** YC launched Autosana as an end-to-end validation harness for coding agents across iOS, Android, and web apps. YC says engineering teams using it have cut QA time by more than 80%, caught major bugs, and increased shipping velocity; founders are Yuvan Sundrani and Jacob Steinberg [5].

- **YC’s vertical-agent cluster.** Trellis is building agents for short-term rental operators that learn workflows and replace 5–10 disconnected tools, while Gov_Guard is starting with FOIA workflows by helping clerk offices search records, flag redactions, and draft response letters for review [6, 7]. Founders include Lodo Benvenuti and Jan Sahagun at Trellis, and Adit Sabby and Gleb Hulting at Gov_Guard [6, 7].
- **Pulse.** After processing billions of pages for Fortune 50 enterprises, large investment firms, and leading AI startups, Pulse open-sourced PulseBench-Tab and T-LAG as the evaluation methodology it says it uses to train and measure production extraction models. Founders Sid Mank and Ritvik Pandey are worth tracking if you are looking at document-intelligence infrastructure [8].
- **StockFit API.** A solo developer spent about a year building StockFit after finding existing investing APIs unreliable. The product exposes SEC-direct fundamentals as structured JSON and adds AI economic models—business model, moats, flywheels, operating levers, strategic initiatives, and failure modes—with every claim tied to a filing URL, section, and verbatim quote. It now offers 83 endpoints plus a native MCP server for Claude, Cursor, and other AI tools [9].
- **logomesh.** Two engineers turned their AgentBeats-winning evaluation agent into a GitHub app for Python PR review. The system infers invariants for modified functions, generates adversarial inputs in an airgapped Docker sandbox, and only comments when it can prove a reachable crash after a second LLM validation pass [10]. Public repos can install it with no config [10].

3) AI & Tech Breakthroughs

- **Perplexity’s Qwen post-training stack is already in production.** Perplexity said its SFT + RL pipeline improves search, citation quality, instruction following, and efficiency, and that its post-trained Qwen-based model matches or beats GPT models on factuality at lower cost [11]. Aravind Srinivas added that the new model sits on the Pareto frontier for accuracy versus cost, has been trained to handle search and tool calls in one model, outperforms GPT and Sonnet on Perplexity’s production cost-efficiency curve, and is already serving a significant share of daily traffic [12].
- **Open and specialized models keep moving the cost/performance frontier.** Bindu Reddy said Kimi 2.6 scored above Opus 4.7 on LiveBench, beat Opus on reasoning and coding, came close on agentic coding, and looked strong in Abacus internal evals at roughly 10x lower cost [13]. In a self-reported OCR benchmark, Dharma-AI said its open 7B and 3B SLMs outscored GPT-5.4, Gemini 3.1 Pro, and Claude Opus 4.6, while DPO

on the model’s own degenerate outputs cut failure rate 87.6% and AWQ quantization lowered per-page inference cost about 22% [14].

- **Document AI is getting more evaluation-driven and more lightweight.** Pulse open-sourced PulseBench-Tab, a table-extraction benchmark across nine languages, along with T-LAG, a unified score for both structure and content [8]. LlamaIndex released LiteParse, a layout-aware PDF parser for agents that preserves structure with a grid-projection algorithm instead of VLMs or other ML models, making it entirely heuristics-based and fast [15, 16].
- **Security tooling for AI codegen is getting more hybrid.** Replit said a new whitepaper shows current-generation LLMs can reach much better performance—above 90% in some cases—when paired with static analysis tools rather than used alone [17].

4) Market Signals

- **Agent-era B2B will punish weak APIs and weak moats.** SaaStr’s operator notes argue that if Replit, Lovable, or v0 cannot build a working dashboard on top of your product in 15 minutes, you are losing the agentic era. The same piece argues many *60% solutions* will fail to monetize because customers can vibe-code better versions quickly, while native integrations and direct data access remain harder-to-copy moats [18].
- **Usage and artifact history are emerging as the new retention metrics.** SaaStr argues DAU/WAU/MAU are now top-five B2B AI metrics because usage can fall to zero well before a cancellation appears in NRR, and that persistent chat and artifact histories can create real switching costs beyond simple memory features [18].
- **Workspace agents are moving toward the enterprise mainstream.** OpenAI introduced workspace agents in ChatGPT as shared agents for complex tasks and long-running workflows across tools and teams [19].
These are cool! I think most companies will want to use them. [20]
- **Agent engineering norms are converging around harnesses, skills, and tests.** Garry Tan endorsed a thin-harness, fat-skills approach, arguing that better outputs and longer-running agents come from deterministic tools plus regression coverage through evals, unit tests, E2E tests, and smoke tests. In a separate thread, he pointed to self-improving agents as an emerging pattern [21, 22, 23].
- **The labor-market picture remains more nuanced than the public debate suggests.** New UK data cited by Marc Andreessen said there is still no significant evidence of an overall employment hit from AI, and that occupations with higher AI exposure have grown faster than least-exposed

ones across measures; wage compression in AI-exposed roles appears to predate generative AI [24, 25].

5) Worth Your Time

- **10x Science background.** The YC announcement linked a TechCrunch article on the company: TechCrunch [1].
- **Archil founder interview.** Dalton Caldwell linked a full interview covering what Archil is, why agent builders should evaluate it, Hunter Leath's 10-year AWS filesystem background, the Clay case study, and the Series A context: YouTube [26, 27].
- **Document AI reading pack.** Pulse published background on PulseBench-Tab and T-LAG, while LlamaIndex published the grid-projection deep dive for LiteParse: Pulse blog, LiteParse blog, repo [8, 15].
- **Replit's whitepaper.** Useful if you are diligencing AI code-security products or hybrid static-analysis + LLM workflows: whitepaper [17].
- **Robotics data infrastructure.** This My First Million segment is a concise look at the data-generation layer behind humanoid robotics: Object Ways, an Indian data-labeling company founded by Dev Mandal, uses 2,000 workers to generate real-world movement datasets for firms such as Tesla Optimus, Figure AI, and data intermediaries like Scale AI [28].



25% Of My Portfolio Is Tesla Stock, Here's Why (19:28)

Sources

1. X post by @ycombinator
2. X post by @Jason
3. X post by @anandnk24
4. X post by @ClementDelangue
5. X post by @ycombinator
6. X post by @ycombinator
7. X post by @ycombinator
8. X post by @ycombinator
9. r/SideProject post by u/Either_Door_5500
10. r/SideProject post by u/Not_Ok-Computer
11. X post by @perplexity_ai
12. X post by @AravSrinivas
13. X post by @bindureddy
14. r/deeplearning post by u/augusto_camargo3
15. X post by @jerryjliu0
16. X post by @llama_index
17. X post by @amasad
18. Our Own Agent Deleted Amelia, HubSpot Gave Us a Zero, and 100 Days Since I Opened Canva: The Latest In The Agents #002

19. X post by @OpenAI
20. X post by @sama
21. X post by @thanford7
22. X post by @garrytan
23. X post by @jc__gr
24. X post by @pdmsero
25. X post by @pmarca
26. X post by @daltonc
27. X post by @daltonc
28. 25% Of My Portfolio Is Tesla Stock, Here's Why