

Foundry's \$19M Seed, Outcome-Priced Agents, and Open-Model Pressure

VC Tech Radar

2026-04-10

Foundry's \$19M Seed, Outcome-Priced Agents, and Open-Model Pressure

By VC Tech Radar • April 10, 2026

This brief covers Foundry Robotics' new seed round, several early teams worth tracking across devtools, cyber, and agent-native software, and the key market signals around faster startup growth, outcome-based pricing, and accelerating open-model adoption.

1) Funding & Deals

- **Foundry Robotics — \$19M seed for AI-first manufacturing.** Foundry says it is tackling American manufacturing with an AI-first, software-defined approach. The seed is backed by Khosla Ventures, Hana Bi Capital, Redglass VC, ZeroShot Fund, and others; Mike Volpi publicly backed the company and the team is hiring. [1, 2]

2) Emerging Teams

- **Contral.ai — AI IDE with a built-in teaching layer.** The product combines a VS Code fork, a repo-aware agent that reads, writes, and runs full codebases, real-time explanations, quizzes, and a Defense Mode that makes users explain their own code. Self-reported traction is strong for a bootstrapped product: #1 Product of the Week on Product Hunt, 400+ beta users, and \$0 marketing spend. [3]
- **numasec — open-source cyber agent with benchmarked recall and conservative controls.** numasec packages 21+ security tools, a security knowledge base, and PTES methodology into a terminal-native or Docker-isolated agent that chains findings instead of dumping a flat list. The founder reports 96% recall on Juice Shop and 100% on DVWA, with a permission model that defaults to asking before execution. [4, 5]

- **PixelGlass — strong founder-market fit in agentic web tooling.** The founder spent three years on Ghost core engineering and is now building a cloud Ghost dev environment where a Claude Opus-powered agent edits themes in a live preview, with one-click deployment or zip export. The stack uses detailed MCP/system instructions plus Ghost’s Gscan validator to keep generated themes clean. [6, 7]
- **Angles — private, local visual search with an unusually good early demo.** Angles focuses on finding photos and videos by visual similarity using local models for text-to-image, image-to-image, “find similar,” and live camera search. The team showed real-time search across an 80,000-photo library and is in early beta. [8]

3) AI & Tech Breakthroughs

- **Interpretability work is pushing from output checks toward internal-state monitoring.** Liberation Labs published *The Lyra Technique*, which aims to interpret structured internal states in transformer KV-caches rather than relying only on outputs. The authors argue this could matter for alignment verification, and point to independent convergence with Anthropic’s recent work on emotion concepts in LLMs. [9]
- **Some builders are trying to remove hallucinations at the architecture layer.** An open-source alternative to Harvey’s tabular review app uses only encoder-based models trained by the builder’s organization—no generative models—and turns contracts into an interactive knowledge graph of entities, spans, and relations. The builder says that design makes hallucinations architecturally impossible; the project was motivated by a reported hallucinated citation from Harvey. [10]
- **Persistent-agent architecture is hardening around memory, evals, and monitoring.** Harrison Chase argues that for 24/7 agents, memory is the core value layer and should live in open harnesses with portable memory. In production, Hex says its Notebook agent can work autonomously for 20 minutes on complex analysis, and its eval stack favors 30-50 handcrafted traps, long-horizon simulations, and LLM-as-a-judge clustering to surface failures without reading raw outputs. [11, 12, 13]
- **Document agents are converging on source-verifiable context.** Jerry Liu argues agents need more than naive PDF extraction: clean multi-modal markdown, bounding boxes for traceability, segmented images, and custom schemas. His `/research-docs` skill shows the direction—complex PDFs, Word files, and slide decks parsed into an auditable HTML report with word-level citations and bounding boxes back to source. [14, 15, 16, 17]

4) Market Signals

- **Early-stage growth expectations keep resetting higher.** A recent YC group-office-hour note put the lowest Demo Day target at \$800k of annualized revenue, versus \$150k two years ago, with most companies aiming for \$1-2M. Paul Graham separately argued that higher valuations have some basis in reality because companies are growing faster now. [18, 19]

“Later stage investors always grumble about increasing valuations. But there is some basis in reality for it: companies do grow faster now.” [19]

- **Outcome-based pricing is becoming a real business model for AI agents.** Sierra built pay-per-resolution in from day one and reached \$100M ARR in 21 months, then \$150M+ ARR by February 2026, with customers automating 50-90% of service interactions. Intercom’s Fin grew from \$1M to \$100M+ ARR, now resolves 2 million issues per week across roughly 8,000 customers, and improved from about 27% to 66-67% resolution rates; as rates rise toward 80-90%, the gap between per-resolution and per-conversation pricing shrinks. [20]
- **Open models are gaining share faster than expected, especially from China, even as frontier access may centralize.** Nathan Benaich highlighted data showing Chinese models accelerating in adoption, with China leading in derivative models and OpenRouter inference share, and Qwen 3.5, Nemotron 3, and Kimi K2.5 standing out on RAM. Bindu Reddy says open-source usage on OpenRouter is already higher than any closed model and that GLM 5.1 and Kimi are close on performance, while Martin Casado predicts only model creators will keep direct access to the strongest systems and everyone else will use distilled variants or first-party apps. Garry Tan’s counterpoint is that distillation should spread capability down the ability-to-pay curve. [21, 22, 23, 24, 25]
- **The sharpest capability gains are in technical work, which helps explain the spread of agentic coding.** Karpathy’s thread—endorsed by Marc Andreessen and Jerry Liu—argues that frontier paid models like OpenAI Codex and Claude Code can now handle programming work that used to take days or weeks, while writing and general advice remain weaker because verifiable-reward domains improve faster and get more B2B focus. Andrew Chen’s prediction that coding becomes a default white-collar skill within 18 months fits the builder behavior in this batch: Replit is being praised for multi-agent collaboration, and founders of MiraBridge and LeanAI openly describe codebases largely written by AI under human orchestration. [26, 27, 28, 29, 30, 31, 32, 33]
- **Cybersecurity is getting fresh early-stage attention.** TechCrunch said it is seeing more cybersecurity startups from the earliest stages. In parallel, Clem Delangue warned that widely used open-source projects are too lightly maintained for how critical they’ve become, and suggested more

funding plus better-resourced umbrellas such as the Linux Foundation or Hugging Face for the most important projects. [34, 35]

5) Worth Your Time

- **Anjuna / TechCrunch clip — hiring after contracts, not ahead of PMF.** Ayal Yogev explains why Anjuna now hires only after signed enterprise deals, after concluding the company had overhired before true



product-market fit. [34]

How to handle layoffs with compassion with Ayal Yogev, Anjuna (5:28)

- **Luminai founder fireside — enterprise sales from an unusually young founder.** Useful for the wedge itself (turning hospital faxes into AI workflows) and for the go-to-market lesson: founder Kesava Kirupa says personal narrative beat cold outreach in closing large customers such as Cleveland Clinic. Watch here [36, 37]
- **Hex production-agent thread — one of the better short reads on agent evals.** Practical notes on small eval sets, long-horizon simulation, and LLM-as-a-judge clustering from a team already running data agents in production. Thread [13]
- **Hallucination-free legal review write-up — useful diligence material for legal AI.** A concrete example of an encoder-only workflow for extracting structured legal knowledge without depending on generative output. Write-up [10]
- **Open-model adoption thread — good starting point on Chinese**

open-model momentum. Useful for tracking derivative models, Open-Router inference share, and which recent models are showing strong relative adoption. Thread [21]

Sources

1. X post by @FoundryRobotics
2. X post by @mavolpi
3. r/SideProject post by u/contralai
4. r/SideProject post by u/Away_Replacement8719
5. r/SideProject comment by u/Away_Replacement8719
6. r/SideProject post by u/ronaldl911
7. r/SideProject comment by u/ronaldl911
8. X post by @tylerangert
9. r/artificial post by u/Terrible-Echidna-249
10. r/SideProject post by u/Neon0asis
11. X post by @hwchase17
12. X post by @RoyZalta
13. X post by @hwchase17
14. X post by @jerryjliu0
15. X post by @llama_index
16. X post by @jerryjliu0
17. X post by @jerryjliu0
18. X post by @t_blom
19. X post by @paulg
20. HubSpot Switching AI Pricing From Per Use to Per Resolution. But Does It Really Matter?
21. X post by @natolambert
22. X post by @nathanbenaich
23. X post by @bindureddy
24. X post by @martin_casado
25. X post by @garrytan
26. X post by @karpathy
27. X post by @pmarca
28. X post by @jerryjliu0
29. X post by @andrewchen
30. X post by @ramkrishnapk
31. X post by @amasad
32. r/SideProject post by u/Either-Ad9196
33. r/SideProject post by u/Pattinathar
34. How to handle layoffs with compassion with Ayal Yogev, Anjuna
35. X post by @ClementDelangue
36. X post by @ycombinator
37. X post by @ycombinator