

The Agent Stack Is Moving Beyond Models—Into Context, Governance, and Power

VC Tech Radar

2026-08-27

The Agent Stack Is Moving Beyond Models—Into Context, Governance, and Power

By VC Tech Radar • August 27, 2026

This brief tracks the shift from model-centric AI toward agent context, governance, and interfaces, alongside the physical infrastructure constraints shaping defense, nuclear, and AI deployment.

1. Funding & Deals

Arga Labs announced a \$10M seed led by General Catalyst. The captured announcement gives only a short thesis—AI agents are taking consequential actions—and no product-level detail, making this a capital-allocation signal into agentic software rather than evidence of traction or differentiation. [1]

Radiant Nuclear received a \$750M U.S. Army award to deliver 15 Kaleidos microreactors through the Janus program. This is not a venture round, but it is a meaningful deployment and validation signal: the Army says it evaluated reactor design, technical maturity, manufacturing readiness, deployment strategy, and commercial viability, while the resulting operational data is intended to accelerate future commercial deployments. [2]

2. Emerging Teams

David Heinemeier Hansson is turning Omarchy, a little-over-one-year-old Linux project, into an agent-shaped product and community. DHH—creator of Ruby on Rails and CTO of 37signals—says the latest Quatro release had reached 100% agent acceleration: none of its shipped code was hand-written, although he reviewed critical model-layer code. He also reports tens of thousands of downloads. [3]

The stronger signal is community participation, not just code-generation speed. Omarchy merged more than 1,000 pull requests in three months, many from non-classical programmers, while its plugin marketplace reached 330 plugins in three days. DHH’s operating thesis is that agents move the bottleneck from implementation toward human bandwidth, product vision, and taste—useful context for evaluating teams whose technical capacity is expanding faster than their product judgment. [3]

Parag Agrawal’s p0/Parallel is treating agents as a distinct search customer. A profile describes Agrawal as the former Twitter builder and CEO who has spent three years building an agent-first search engine. The system shipped a search agent before a full search engine so it could grow its index incrementally; the stated technical targets are selecting the right 1,000 tokens from a trillion web pages, using fewer than half the tokens, and reducing the compute budget from three seconds to 200 milliseconds. [4]

The thesis is also economic: human click data is the wrong optimization signal for agents, and an internet funded by scarce human attention may need a new publisher-payment model. The discussion proposes Shapley-value attribution and positions Parallel as a complement to models; a Google Cloud deal would make Google Search or Parallel Search the grounding options on GCP. [4]

3. AI & Tech Breakthroughs

Physics AI is following a different scaling path from language models. Latent Space reports that Anima Anandkumar’s team built FourCastNet within a year, reaching performance competitive with leading physics-based simulations and enabling short-term weather forecasting on consumer-grade GPUs. Her Neural Operators combine data with physical laws while representing functions across multiple scales; spherical-harmonic variants help stabilize global weather rollouts. [5]

The investment angle is structural rather than merely larger-model scaling: relevant datasets may contain only tens or hundreds of thousands of examples, while high-resolution physical systems can demand hundreds of billions to a trillion tokens of context. In fusion, the source reports that a few thousand samples enabled plasma-disruption prediction one million times faster than traditional simulation. Anandkumar’s TorchLean work extends the same direction toward formally verifying neural networks for safety-critical control loops. [5]

OpenWiki and WikiBench turn codebase documentation into a testable agent-navigation layer. LangChain’s OpenWiki generates and maintains repository documentation; WikiBench pins a repository, generates coverage and retrieval questions, and uses a reader agent plus LLM judges to score both factual completeness and grounding. [6] The benchmark found that giving an agent both the wiki and raw source produced the highest mean score at lower cost than raw source alone, while the wiki alone performed much worse.

The practical conclusion is precise: generated documentation is an index and guide, not a replacement for source code. [6]

Persistent memory is becoming an explicit agent product layer. Perplexity describes Brain as a self-improving memory system that compiles sessions, files, and sources into a structured knowledge wiki; its new evaluations report gains of 9.3 points in correctness, 8.0 in currentness, and 8.9 in recall while using 15% fewer tokens. A background Dream agent continually ingests context from files and connected apps to build multi-hop context graphs. These are vendor-reported results, but they reinforce the same architectural direction as WikiBench: structured context can improve an agent’s navigation and efficiency around the underlying source. [7, 8, 6]

4. Market Signals

Box offers a concrete public test of AI monetization in incumbent SaaS: sell governance and context, not inference. Its Enterprise Advanced tier is priced 20–40% above Enterprise Plus; after five quarters of data, billings grew 17% year over year versus 9% revenue growth, while net revenue retention moved from 103% to 105% to 106%. [9] Box held non-GAAP gross margin at 81.2%, down only 20 basis points, because competing agents pay the token bill while Box supplies governed content. It placed agent guardrails, prompt-injection detection, MCP-scoped permissions, classification policies, and audit trails behind the premium tier. [9] The relevant startup question is whether permissions, governance, and auditability can become a separately priced control plane rather than an unpriced cost of model consumption. [9]

Brand impersonation is becoming a multi-surface infrastructure-mapping problem. Doppel’s analyst-validated Threat Graph says more than 80% of confirmed impersonated brands appeared on at least two monitored surfaces, while near-simultaneous multi-surface campaigns grew roughly sevenfold in two years; domain-only tools capture only about a 6% slice of the problem. [10] Doppel reports that roughly 97% of takedown requests resulted in content removal or hosting pull-down, but almost 60% of taken-down domains served a new confirmed fake within 24 hours and close to 70% did so within 90 days. [10] Its proposed wedge is continuous monitoring and mapping the operator infrastructure behind recurring fakes, rather than removing one artifact at a time. The figures are customer-anonymized, raw volumes are withheld, and surface coverage is limited by the monitoring modules each brand purchased. [10]

European defense startups face a test-to-contract bottleneck that creates an enabling-infrastructure opportunity. The article contrasts a U.S. Army effort that opened five ranges to companies without procurement contracts, with a European pattern in which governments demand test data before awarding a contract while the needed ranges and specialist support become available afterward. [11] One founder was offered a slot at Germany’s WTD 91

in the fourth quarter of 2029; a GNSS-free navigation startup found three suitable sites worldwide, faced a Canadian bill above \$120,000 for a few flights, and found UK range limits too restrictive to produce useful results. [11] Simulation cannot reproduce every component failure or adversarial adaptation, and one anti-drone company keeps a quarter of its staff in Ukraine for weekly testing against deception and jamming. [11] The proposed investment theme is a small network of mission-ready ranges with transparent capabilities and availability, supported by reusable cross-border “campaign passports” and buyer-agreed outcomes such as paid integration, an operational trial, or an order. [11]

At the AI hardware boundary, power is the contrarian watchpoint. Flex CEO Revathi Advaiti argues that AI is increasingly an infrastructure story—and specifically a power story—rather than only a compute story. The thesis challenges a market still priced primarily around chip scarcity. [12]

5. Worth Your Time

- **Watch — DHH: Future of Programming, AI, Agentic Engineering, Vibe Coding & Linux.** The useful segment is the operating evidence: agents filter a 1,000-PR stream for human decisions while the project’s extension system lets non-classical programmers contribute



quickly. [3]

DHH: Future of Programming, AI, Agentic Engineering, Vibe Coding & Linux | Lex Fridman Podcast #501 (34:02)

- **Read — Anima Anandkumar on physics foundation models.** A compact explanation of why physical-world AI needs neural operators,

physical priors, and formal verification rather than simply more token scale. [5]

- **Read — Europe’s defense startups struggle to test their way to a contract.** The piece turns a procurement failure mode into a concrete range-network and “campaign passport” thesis. [11]
- **Read — Faking a brand is easy. Making it stop is hard.** The useful diligence detail is the gap between high takedown success and rapid domain recurrence, plus the methodological limits on the underlying data. [10]

Sources

1. X post by @ycombinator
2. X post by @RadiantNuclear
3. DHH: Future of Programming, AI, Agentic Engineering, Vibe Coding & Linux | Lex Fridman Podcast #501
4. X post by @sonyatweetybird
5. “We have foundation models for language, not for physics” — Anima Anandkumar, Bren Professor of Computing
6. X article by @LangChain
7. X post by @perplexity_ai
8. X post by @AravSrinivas
9. 5 Interesting Learnings from Box at \$1.29 Billion in Revenue: 17% Billings Growth, 106% NRR, and 20 Basis Points of AI Margin Cost
10. X article by @KevinTian00
11. X article by @nathanbenaich
12. Flex Ltd: Will Power Be The New Silicon?