Model Routing, Deep-Tech Patience, and Linear's Fundraising Discipline

VC Tech Radar

2026-06-08

Model Routing, Deep-Tech Patience, and Linear's Fundraising Discipline

By VC Tech Radar • June 8, 2026

This brief highlights a growing thesis around model routing and cheaper inference, a practical fundraising template from Linear, and a small set of AI teams tackling concrete enterprise and industrial problems.

Funding & Deals

- **Linear.** With three co-founders, the company raised a small angel round so it could build the first version without much money or an early VC commitment. Sequoia later led the seed once the team decided to raise, and about a year later Linear was breakeven/profitable; later rounds were run as short, relationship-driven processes, with Accel's Series B discussion rooted in prior research rather than a standard pitch. [1]
- **Build.** Tony Fadell says Build backs hardware, software, and chemical technologies that can unseat incumbents across environmental, societal, and health markets. The firm also works with founders on product, operations, financing, org design, and storytelling so teams can get closer on version one or two rather than waiting until version four; Fadell adds that earlier positions in Grok and Cerebras were made before the hype cycle. [2]

Emerging Teams

- **Soryn.** The product connects a team's Google Drive and answers internal questions about company documents. The founder emphasizes tenant isolation via Postgres RLS and AES-256-GCM, automatic per-company OpenAI budget caps, query rewriting plus a larger context window, and Redis semantic caching for repeated queries; the system is stable with

automated onboarding and the founder is now asking for real user feedback. [3]

- **ApplyBoostAI.** The founder says the product direction came after speaking with dozens of job seekers and finding that many did not understand why they were failing to get interviews, often blaming the market instead of the resume itself. That insight became ApplyBoostAI.com. [4]
- **Simbi Robotics, Great Parrot, and Oriannis.** Fadell cites Simbi Robotics, which uses AI + robotics for retail inventory and is "just taking off" after 7-8 years; Great Parrot, which uses AI + cameras for recycling sorting and textile defect detection; and Oriannis, which has been applying AI to drug design for 10 years and is "taking off." [2]

AI & Tech Breakthroughs

- **Tracer.** After observing repeated production traffic spent on tasks like classification, tagging, routing, moderation, extraction, intent detection, and tool selection, the founder built Tracer to train lightweight local models from prior LLM traces. Those surrogates are activated only when they match the original model well enough, while uncertain cases go back to the frontier model; on one repeated classification workflow, the founder says this reduced LLM cost by about 95%. [5]
- **Model routing.** Jerry Liu argues that significant value will accrue to startups building "model routing as a service," extending from document infrastructure such as parsing, extraction, and search to web-search analogs and vertical applications. He also says that finding the right point on the accuracy/cost/latency curve requires substantial evaluation, benchmarking, and reliability work. [6]
- **Operational AI in the physical world.** The Build examples share a common structure: AI is paired with robotics, cameras, or scientific workflows to address inventory labor, recycling and textile waste, and drug design rather than only software-side productivity. [2]

Market Signals

- **Model economics.** One X post argued that demand for intelligence is near-infinite, but 80% of workloads will run on models that are 99% cheaper within 12-18 months, while 20% stay on latest-generation models; under that view, energy and compute become the bottleneck rather than incremental model quality. [7]
- **Value capture.** Liu's framing implies that more of the value may sit in the layer that chooses the right model and operating point for each task, because accuracy, cost, latency, and long-tail reliability have to be optimized together. [6]

- **Investor filter.**

  > "I always start from pain." [2]

  Fadell says new technology should solve real pain in a way that creates a fundamentally different product and can unseat incumbents, not just feature-compete. He also argues that deep-tech founders often need help translating strong research into product and storytelling. [2]

Worth Your Time

- **Tony Fadell on Lenny's Podcast.** Useful for his pain-first deep-tech framework and for concrete examples across retail robotics, recycling/textiles, and AI drug design. [2]

Father of the iPod and iPhone on building taste, judgment, and creativity in the AI era (22:25)

- **Jerry Liu on model routing as a service.** Useful if you are screening routing, document-infra, or vertical-agent companies; the thread explains why the hard part is benchmarking and reliability, not just switching APIs. [6]
- **Kari Saarinen on Linear's fundraising history.** Useful founder-side reading on delaying VC commitment, then running tight rounds once product and business quality are stronger. [1]
- **Tracer: I reduced LLM costs by 95% and open-sourced the tool.** Useful as a short operator case study on replacing stable frontier-model calls with local surrogates. [5]

Sources

1. X post by @karrisaarinen
2. Father of the iPod and iPhone on building taste, judgment, and creativity in the AI era
3. r/SaaS post by u/StrengthFinancial693
4. r/SaaS post by u/Radiant_Freedom9451
5. r/SaaS post by u/Adr-740
6. X post by @jerryjliu0
7. X post by @brian_armstrong