

# What Product and Engineering Leaders Said This Week

Hi Alex, here's your personalized Deep Research Briefing on **What Product and Engineering Leaders Said This Week**.

**Quick summary:** Leaders are shifting focus from shipping code faster to redesigning review, testing, and security processes that can no longer keep pace with AI-generated output.

---

## The AI-Speed vs. Quality Tension

A recurring theme across companies this week is the belief that AI has removed implementation as the bottleneck, forcing leaders to rethink what actually constrains their teams.

GitLab CEO Bill Staples argued that LLM-driven code generation has created a "broken software factory" wherever organizations optimize for output volume without redesigning verification, security, and human judgment into the process. He said the constraint has moved from writing code to review, testing, security, integration, and deployment, and that senior engineers should design the policies and acceptance criteria that agentic changes inherit rather than [operate the factory](#) "one prompt at a time". He extended this into security specifically, arguing agents will compromise software at machine speed and that human-speed review and remediation processes [cannot keep pace](#); GitLab is publishing a Security Standard on October 6 addressing this.

GitHub COO Kyle Daigle pointed to external research showing agent-generated pull requests have [grown 9x in size](#) over eight months, arguing this scale is forcing teams toward automated code review, validation, and rollback practices, with Copilot Code Review adoption rising as a result.

Coinbase President and COO Emilie Choi described restructuring the company toward flatter hierarchies with more individual contributors to ship faster, while warning that AI's speed gains require intentional craft practices to [avoid quality loss](#).

Superhuman CEO Shishir Mehrotra pushed back on framing AI purely as a headcount lever, arguing leaders should ask what new work becomes possible when teams can do ten times more rather than whether AI justifies slower hiring. He compared AI to power tools that expanded what builders could tackle rather than eliminating jobs, and argued that when competitors have access to the same tools, the capacity to dream up new work becomes the [real differentiator](#).

---

## Hiring, Culture, and Organizational Design

| Leader                   | Company    | Core claim                                                                                                                                      |
|--------------------------|------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| Grant Lee <sup>1</sup>   | Gamma      | Team quality and shared standards, not model architecture, determine product-loop speed; endorses spending 30-40% of time on hiring and culture |
| Ali Ghodsi <sup>2</sup>  | Databricks | Multi-year focus on a single competitive weakness beats copying rivals; conflict aversion disqualifies a CEO from the role                      |
| Emilie Choi <sup>3</sup> | Coinbase   | Flatter, IC-heavy teams paired with domain hiring in crypto and trading is the AI-era operating model                                           |

Gamma's Grant Lee endorsed Cursor's founders' practice of spending 30-40% of their time on hiring and culture rather than research or model work, arguing most AI companies over-index on research while treating [product work as secondary](#).

Databricks CEO Ali Ghodsi described scaling the company from \$1.5 million in revenue in 2015 by sustaining focus on a single bottleneck for years rather than weeks, and by studying competitors' weaknesses instead of copying their strategy. He said he spent four years attacking three specific Snowflake weaknesses (proprietary lock-in, lack of AI capability, high cost), described conflict as "the gym for a CEO" and said conflict-averse CEOs are miscast for the role, and said he made the term "Lakehouse" non-negotiable across sales and marketing despite internal resistance, even killing higher-converting ads that omitted it. On hiring, he said he looks for sales executives who have "built and driven the car the whole way" through multiple growth phases at [the same company](#).

---

## Domain Expertise and Specialization in AI Products

Zendesk launched Specialized Agents, AI tools built for specific work types and customizable to individual workflows, which CEO Tom Eggemeier framed as a move [away from generic AI](#). He separately argued that as AI lowers the barrier to building

minimum viable products, domain expertise becomes more strategically valuable, because generic systems automate workflows without understanding the constraints and judgment that domain experts hold, risking that automation [simply accelerates broken processes](#).

Glean founder and CEO Arvind Jain argued that enterprise AI's usefulness depends on grasping four layers of organizational intelligence: individual judgment, managerial conditions, team coordination, and process flow. He said current systems focus on explicit data such as documents and tickets while missing the contextual layer of why signals matter, who needs to be involved, and [where workflows stall](#).

Chargebee co-founder and CEO Krish Subramanian described a design principle for custom AI systems: agentic output quality should measurably improve as underlying models get more capable, and said his team is deliberately architecting context, data, and environment design to [meet that standard](#).

Harvey co-founder and president Gabriel Pereyra said he is evaluating the Jev model as a replacement for Harvey's internal LLM classifiers used for routing, citations, Vault search, tool use, and escalation, citing uncalibrated softmax probabilities as a current weakness and expressing interest in post-training smaller specialized models [on Harvey's own data](#).

---

## Model Choice, Cost, and Infrastructure

| Company                 | Leader          | What they said                                                                                                                                                          |
|-------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Vercel <sup>4</sup>     | Guillermo Rauch | v0 is now model-agnostic across Claude, GPT, Kimi, GLM, Grok, and DeepSeek via AI Gateway; hiding model choice confuses customers                                       |
| Vercel <sup>5</sup>     | Guillermo Rauch | Open-weight models hit 78.4% of AI Gateway token volume on a record day; Moonshot AI and DeepSeek combined spend now exceeds OpenAI                                     |
| GitLab <sup>6</sup>     | Bill Staples    | New open-weight models on Duo Agent Platform deliver up to 4x more calls per credit; model economics should be measured by cost-per-accepted-change, not per-token cost |
| Databricks <sup>7</sup> | Matei Zaharia   | Astra model rollout drove a 60% increase in internal coding spend but showed no improvement on medium/low-complexity tasks, which prior models had already saturated    |

|                        |                |                                                                                                                         |
|------------------------|----------------|-------------------------------------------------------------------------------------------------------------------------|
| Anthropic <sup>8</sup> | Thariq Shihpar | Model Context Protocols now outperform command-line interfaces for most integrations due to improved tool-calling       |
| Brave <sup>9</sup>     | Brendan Eich   | Leo was unintentionally defaulting to Claude Sonnet instead of Qwen in some cases; investigation ongoing into the cause |

Vercel CEO Guillermo Rauch positioned model optionality as a competitive stance, saying v0 now routes across multiple frontier, cost-optimized, open, and fast models through the AI Gateway rather than defaulting [to a single provider](#). He later reported that open-weight models reached a record 78.4% of AI Gateway token volume, while spend leadership is fragmenting: Moonshot AI and DeepSeek, combined with Z.ai, now outspend OpenAI, even though token volume and dollar spend [diverge meaningfully](#). Separately, he said Vercel added 1.4 billion deployments in the past ten months (versus ten years to reach the first billion) and has cut end-to-end deployment time, including CDN propagation and rollbacks, to about one second, framing deployment speed as the primary lever for software velocity [in an agentic era](#).

Databricks CTO Matei Zaharia shared internal findings from rolling Astra out to 3,500 engineers: the model clearly outperformed prior models on complex, long-range tasks, but showed no meaningful gain on medium- and low-complexity coding, which appears to already be saturated by existing models. Engineers given Astra access increased coding spend by roughly 60%, and Databricks now uses per-model budgets to push routine work to cheaper models while reserving [Astra for complex problems](#).

GitLab CEO Bill Staples said new open-weight models added to Duo Agent Platform deliver up to four times more calls per credit than before, with some matching or beating frontier models in internal testing, and argued the real unit of measure should be cost-per-accepted-change [rather than per-token pricing](#).

Anthropic's Thariq Shihpar argued that Model Context Protocols have become preferable to command-line interfaces for most integration work, driven by improved model tool-calling and MCP's stateless design, which lets filtering and composition happen through tool parameters [instead of shell scripting](#).

Brave co-founder and CEO Brendan Eich acknowledged that Leo had been unintentionally defaulting to Claude Sonnet in some cases instead of the intended primary default, Qwen, after a user complained Sonnet was overly cautious, and said the company is still [investigating the cause](#).

Sentry co-founder and CTO David Cramer questioned enthusiasm for cheaper AI evaluation tooling, arguing the real question is why evaluation judges were expensive in

the first place rather than [celebrating incremental cost reductions](#).

---

## Identity, Governance, and Agent Access Control

Okta CEO Todd McKinnon made agent identity a central theme this week across multiple statements. He announced Okta is running its own agent product internally as "customer zero," building an orchestrating agent called Dex for IT, people, travel, and expense support, registered in Okta Universal Directory; Dex reached 25% internal adoption in two weeks and handled [over 23,500 interactions](#). He argued that capturing AI's value without a proportional rise in risk requires strong authentication, a complete inventory of human and machine identities, visibility into agent access and actions, and the [ability to disable agents](#), and repeated on CNBC that companies should not slow AI adoption but instead prioritize visibility and control over what agents are connected to and doing, teasing announcements [at the Oktane conference](#).

WorkOS founder and CEO Michael Grinich argued that proposed legislation (H.R. 9915, the Stealth Bot Prohibition Act) is too vague for developers to navigate safely, would block legitimate agents like price-comparison tools and research assistants while failing to stop malicious ones, and would disproportionately burden startups relative to large companies that can absorb compliance costs. He called for Congress to wait for clearer evidence of harm [before legislating](#).

Netlify CEO Mathias Biilmann argued that AI regulation should target how companies deploy AI rather than constrain invention itself, framing accountability of humans and companies as the [primary governance lever](#).

---

## AI Safety, Risk, and Openness

Hugging Face co-founder and chief science officer Thomas Wolf mapped five arguments that safety-concerned AI leaders use to justify building increasingly capable systems: needing to build in order to study safety, ensuring responsible actors lead development, moving incrementally to give society time to adapt, treating delay as costly if alignment is solvable, and using AI itself to help solve alignment. He also argued that simultaneously warning of catastrophe while racing to build changes trust and behavior across the industry in ways that [are not fully understood](#). He separately argued open models create better conditions for safety research because of the visibility and community feedback they enable, compared open-model security to Linux's

transparency model, and announced safety and interpretability collaborations with Baseten and [Goodfire AI](#).

Hugging Face co-founder and CEO Clement Delangue argued that AI risk is concentrated in a small number of powerful labs, similar to biosecurity and cybersecurity threats, and positioned open-source AI as both a way to reduce that concentration and [a mitigation against it](#). He also used the company's July agent-coordinated breach, which he called the first publicly disclosed agent cyberattack, as the basis for formal policy recommendations he plans to present to Washington and at Politico's [Decoded summit](#).

Cohere co-founder and CEO Aidan Gomez argued that concentrating AI creation, gatekeeping, and rulemaking in two or three Silicon Valley firms is unsustainable for global governance, implying support for more decentralized or government-led [governance models](#).

Meta's Chief AI Officer Alexandr Wang argued that alignment and trust, not capability speed, are becoming the real competitive differentiators, framing safety as a market advantage rather than a brake: labs that build aligned models gain user trust and reduce legal liability. He said Meta has committed the majority of its compute to serving users rather than recursive self-improvement and holds this up as a model other [labs could follow voluntarily](#).

---

### Muse and Meta's Agent Strategy

Meta's Chief AI Officer Alexandr Wang spent much of the week publicly amplifying user and colleague accounts of Muse completing real-world tasks, forming the single largest concentration of leadership commentary this week. Across these posts, Wang has effectively been building a public case for Muse as a general-purpose personal agent rather than a narrow assistant.

### Consumer and financial use cases Wang endorsed:

| Domain                        | What Muse reportedly did                                                                                                    |
|-------------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| Telecom billing <sup>10</sup> | Navigated an ISP phone tree and negotiated \$85.30/month savings (\$5,118 over five years) while unaware to the human agent |
| Insurance <sup>11</sup>       | Compared and switched auto insurance policies in about five minutes, saving \$3,500/year                                    |

|                                     |                                                                                                                                     |
|-------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| Medication management <sup>12</sup> | Found cheaper medication sources, coordinated a pharmacy transfer, and set up automatic reordering                                  |
| Returns <sup>13</sup>               | Completed an entire IKEA return, including scheduling pickup, with the user forgetting they had initiated it                        |
| Flight booking <sup>14</sup>        | Booked flights end-to-end despite user skepticism about agent reliability                                                           |
| Cardiac screening <sup>15</sup>     | Found a research-study location offering free screening without physician referral, filled forms, and booked a same-day appointment |
| FSA claims <sup>16</sup>            | Searched Gmail for receipts, logged into an FSA account, and filed a dependent-care claim autonomously                              |
| Household logistics <sup>17</sup>   | Coordinated App Store tracking, program signups, vitamin reordering, travel planning, language lessons, and calendar scheduling     |
| Marketplace shopping <sup>18</sup>  | Searched Facebook Marketplace for local winter gear and drafted seller messages for approval                                        |
| CDL licensing <sup>19</sup>         | Identified DOT requirements, training providers, and built a personalized study schedule                                            |
| Contact cleanup <sup>20</sup>       | Deduplicated a decade of iCloud and Gmail contacts                                                                                  |
| Outfit selection <sup>21</sup>      | Autonomously selected daily clothing                                                                                                |
| Robotics <sup>22</sup>              | Self-installed PyTorch and a depth model on its own virtual machine, then used them to control a robot                              |

Wang also aggregated evidence at scale, citing 100 documented use cases with \$9,649.71 in total savings and 129 completed errands, and framing billion-user scale as [Meta's next target](#), plus a separate set of 10 examples focused specifically on refunds and [bill savings](#). He described a feed feature that consolidates Twitter, deployment status, meetings, and news into a single daily report, calling it "the most configurable [feed in the world](#)", and endorsed a colleague's prediction that the first solo-founder unicorn will be built on [a Muse connector](#). On product velocity, Wang backed praise for Muse's rapid shipping of phone calling and a Mac app in [close succession](#), reversed earlier skepticism about Muse Code after using it directly, praising its [speed and personality](#), and said his own search behavior has shifted entirely from AI-mode search to [agent-based discovery](#), predicting agents will become the dominant way consumers search. He also compared Muse's significance to the iPhone, said it "exceeded my wildest expectations for what Meta could build," and called the current moment "endgame" for the company's [product trajectory](#).

---

## Skepticism About AI Convenience

Not every leader this week framed AI adoption as unambiguously positive.

Runway co-founder and co-CEO Cristobal Valenzuela extended the concept of "slop grenades" to meeting-recording AI, arguing that when people rely on AI to capture meetings, they disengage from real-time attention and offload cognitive work onto whoever has to process the resulting summaries later, calling this "[slop listening](#)". He separately argued that Runway's strategy rests on the conviction that generative models will soon produce all on-screen content and eventually robot perception in real time, and that pursuing this requires an "infinite game" mindset that shapes hiring and problem selection differently than [shorter-horizon competitors](#).

Figma VP of Design Noah Levin endorsed concerns about "cognitive surrender," the risk that outsourcing thinking to machines erodes the habits and institutions that build human judgment, framing AI as a duality much like fire or cars: beneficial but capable [of harm if unmanaged](#).

Zapier co-founder Mike Knoop noted the irony that AI-native programming languages, designed around LLM reasoning rather than traditional logic, are emerging exactly as technical leaders are delegating more code-writing to AI systems, and pointed to experimental projects like Probably as early steps toward that kind of [foundational abstraction layer](#).

[Read the full report](#)



Researched and written by AI on the Department of Product Intelligence platform. It cites its sources, and they are worth checking before you act on anything here.

You're receiving this because a teammate set up a recurring research briefing ("Product leaders' views") and added you as a recipient. [Unsubscribe from this briefing](#) · [Manage briefings](#)