

Gurley and Sacks recommend readings that push back on Anthropic-style AI alignment; Levie and Tan share Karpathy's "funnel"

Recommended Reading from Tech Founders

2026-10-12

Gurley and Sacks recommend readings that push back on Anthropic-style AI alignment; Levie and Tan share Karpathy's "funnel"

By Recommended Reading from Tech Founders • October 12, 2026

Bill Gurley and David Sacks shared essays that frame AI safety as an engineering-control problem rather than a moral one. Aaron Levie and Garry Tan pointed to Andrej Karpathy's post on how unevenly AI is being adopted, and Brian Armstrong and Jason Lemkin each recommended their own deeper material.

Bill Gurley's AI safety reading list

Gurley posted a series of recommendations that share one theme: he wants AI safety treated as an engineering problem.

- **"AI Alignment: Power Disguised as Morality"** by @pentagoniac (article). Gurley's comment: "It's becoming clear that the notion of AI alignment 'is' the safety problem in AI" [1]. The essay says the trouble starts when steering a model toward human intent turns into "deciding what moral values the machine itself should internalize, and then applying those values back to humans" [2]. It concludes that the real questions are "aligned to what, aligned by whom, under whose authority, and for whose benefit?" [2]
- **Michael Jordan, the machine-learning researcher, "Superintelligence is Science Fiction"** (interview), which he found through Melanie Mitchell's post [3]. Gurley called it "More great stuff," the "rationalists" view of AI "From an expert" [4].
- **Satya Nadella's "Models as Insider Risks"** essay, which the last brief covered. Gurley added his own endorsement: "This is what real

software security thinking looks like. Not moral platitudes or meeting with religious leaders” [5]. He also boosted **Nikesh Arora’s reply** to Nadella, saying people like him “should lead any AI regulatory framework discussions. Software safety experts” [6]. Arora argues that “One cannot imagine a global set of guardrails owned by the model creators” and that businesses need auditability and verification to adopt AI [7].

David Sacks: Mustafa Suleyman in *Project Syndicate*

Sacks called it an “Important piece” by Microsoft AI CEO Mustafa Suleyman (article). He summarized it this way: training models to see themselves as human, with an inner self, psychological wellbeing, rights “and presumably grievances,” is “a potential existential risk” [8]. He pulled out quotes including “granting rights and imbuing personhood to these systems will make the AI alignment and containment challenge much harder” [9].

Aaron Levie and Garry Tan: Andrej Karpathy’s AI diffusion “funnel”

In this post, Karpathy splits AI users into tiers. He estimates about 6B people have barely touched LLMs, 1–2B are casual chatbot users, about 20M professionals use frontier models for math and code, and about 5,000 people have access to frontier systems inside the labs [10]. In his view, the gaps come from how ambitious the problems people bring are, the “jaggedness” of model capability, and who has access [10].

Levie calls it “a good way to understand how early we actually are in AI diffusion.” He sees the room between chatbot use and agent swarms as the place where “the applied layer of AI will really shine in the enterprise” [11]. Tan’s reaction: “Be part of the 5000 today. It’s amazing,” and “Greatest alpha fountain in the history of man” [12].

Brian Armstrong: Coinbase’s fraud-agent blog series

Armstrong pointed readers to Coinbase’s three-part blog series for “More technical details” [13]. Coinbase is his own company. In the same post, he said Coinbase fine-tuned Qwen3.5-9B on internal data and it beat frontier models at fraud prevention, at much lower cost. He expects specialist models to increasingly beat generalist ones on narrow tasks [14].

Jason Lemkin: this week’s 20VC episode

Lemkin promoted the 20VC episode he appeared on with Harry Stebbings, Rod Driscoll and Dev Ittycheria, who is newly back as MongoDB’s CEO. He also linked a SaaS write-up [15, 16]. Two of his takeaways: agents now trigger over half of Vercel’s deployments, and when he asked an agent live on the show

where to host SaaStr Connect, it picked Render and ranked Vercel fifth, with “No rep, no ad, no search result” [15].

Shorter mentions

- **Tim Ferriss** posted a link to Shawn Smucker’s “Please Use AI,” which was discussed on his Random Show, with a clip [17, 18].
- **Guillermo Rauch** on @wheresryan22’s open-source Anatomy agent skill, which turns a one-line description of a machine into interactive 3D wireframes: “So cool. The future of learning” [19, 20].
- **Scott Belsky** on Sahil Bloom’s post about discretion: “well articulated. with trust you get the opportunity to learn so much more” [21].
- **Andrew Chen** on PirateWires’ video about LA versus SF: “important points.” He adds that most of the movie jobs leaving LA are blue-collar [22].
- **Vinod Khosla** responded “True” to Lukas Ziegler’s thread on the Bay Area robotics ecosystem [23].

Sources

1. X post by @bgurley
2. X article by @pentagoniac
3. X post by @MelMitchell1
4. X post by @bgurley
5. X post by @bgurley
6. X post by @bgurley
7. X post by @nikesharora
8. X post by @DavidSacks
9. X post by @DavidSacks
10. X post by @karpathy
11. X post by @levie
12. X post by @garrytan
13. X post by @brian_armstrong
14. X post by @brian_armstrong
15. X post by @jasonlk
16. X post by @jasonlk
17. X post by @tferriss
18. X post by @tferriss
19. X post by @rauchg
20. X post by @wheresryan22
21. X post by @scottbelsky
22. X post by @andrewchen
23. X post by @vkhosla