

Consumer AI spending is concentrated in a small set of power users, and Claude is gaining share

VC Tech Radar

2026-10-06

Consumer AI spending is concentrated in a small set of power users, and Claude is gaining share

By VC Tech Radar • October 6, 2026

This period's data from a16z and Big Technology shows consumer AI spending concentrated in a small, heavy-paying minority, with Claude gaining share. Also covered: Reflection's open-weights Beam, Menlo's framework for VC in the AI cycle, startup work on agent harnesses and memory, and new raises in bio and hardware.

Consumer AI: wide use, narrow spending, and a shift in share

a16z's seventh Top 100 Consumer AI list ranks apps by revenue for the first time. The data is consumer card spend, so enterprise and SMB spending is not included. Traffic rankings are settling: only 11 new products entered the web and mobile lists, the fewest of any edition. Of the 50 products ranked by spend, 29 did not appear on either traffic list [1].

Spending is concentrated among a small group of heavy users: - About half of Americans say they use AI, but a16z's Yipit panel puts paying subscribers at about 4.5% of US consumers. Some other sources put it at 2-2.5%. a16z says the share has doubled in a year [2]. - Among people who pay, the top 10% generate more than half of revenue and the top 1% generate 20%. The top 1% spend \$903 a month, against a median of \$25. Most of the spending goes to developer, productivity and creative tools [2]. - 85% of web-list apps make money from subscriptions, 62% from credits or usage charges, and only 13% from ads. Olivia Moore calls this an "unnatural inversion" of the usual consumer-internet model and expects ads and transaction fees to return [3]. - Nine of 15 consumer internet categories have no AI product in the Top 100: streaming, social, dating, gaming, travel, retail, finance, real estate and jobs [4].

The diligence point: in consumer AI, revenue rankings and traffic rankings now pick out different winners. And the open categories are the ones that have historically been funded by ads and transactions, not subscriptions.

Claude is taking share. a16z says ChatGPT still leads, with about 6x Claude’s web usage and about 3x as many paid US consumer subscribers as Gemini or Claude. But a16z’s US spender panel shows Claude has passed Gemini in paid subscribers. About 7.5% of Claude subscribers are on plans costing \$100 or more a month, against about 1% for ChatGPT and Gemini [1]. Big Technology’s numbers ahead of Anthropic’s IPO point the same way: - **App users:** Apptopia puts Claude’s share of daily active users across AI apps at 14.9% in September, up from 1.4% in January. ChatGPT’s fell from 41.8% to 30.7% [5]. - **Downloads:** Claude’s US downloads peaked at 7.42M in March and fell to 3.26M in September [5]. - **Web traffic:** In August Claude’s site had 950M visits, up 540% year over year [5]. - **Enterprise:** Claude’s SSO login traffic, which Similarweb uses as a proxy for enterprise use, has roughly quadrupled since March [5].

The competitive question for agents. Andrew Chen says new horizontal consumer agents (Muse, Town, Instinct, Grokbot) are making thousands of vertical-agent startups rethink their position [6]. He argues agents don’t have network effects by default. Better models, UX and memory can build large companies, but those advantages are not network effects [6]. The question that matters is where network effects end up. If identity, reputation and context sit in open protocols, vertical agents can thrive. If horizontal agents keep them proprietary, switching agents means leaving your network behind [6].

Open weights: Reflection’s Beam

Reflection AI announced Beam, an “agentic open model” with 501B total parameters and 23B active, trained from scratch. Reflection says Beam advances Western open models on coding and agentic tasks and that full weights will be released this month [7]. These are the company’s own claims, and no outside benchmarks appear in this period’s sources.

How VCs are playing the AI cycle: Menlo’s approach and a dissent

Harry Stebbings credits Menlo’s Anthropic round at a \$4B valuation with making the firm “\$50BN+,” followed by Lovable, Higgsfield and Legora [8]. Venky Ganesan of Menlo made these arguments: - **Seed checks as options.** A seed check buys an option; you size up only when revenue data shows a company is an outlier [8]. Menlo has put 20% of a fund into one company, Anthropic, by “laddering up” as new data came in [9]. - **The return bar.** Startups pay compute and model costs to hyperscalers, while LPs can buy those same companies without fees or carry. So private VC has to beat public markets by at least 1,000 basis points [8]. - **What breaks a cycle.** Defaults on leveraged debt,

not equity write-downs [8]. - **Tranched rounds as froth.** These rounds raise lower-priced “build with me” money first and then more at a higher valuation. Ganesan says everyone now uses them, regardless of company quality [10].

Sarah Guo disagreed. A lead seed investor can treat the check as a full bet and back it for years: “structure determines strategy” [11]. Stebbings separately predicts incumbents will start acquiring AI customer-support companies [12].

Agent infrastructure: harnesses, test environments, memory

- **Harnesses as training environments.** Hugging Face turned unmodified coding harnesses into RL environments by routing their API calls through a capture proxy. The same model scored 62% in Mini-SWE-Agent and 33% in Claude Code. Training a 2.6B model across four harnesses improved all four (42%→54%), while fine-tuning on rollouts from a larger model plateaued at 47.5%. All of it is open source [13].
- **Grep AI (YC F26)** builds agents that turn repeatable steps into code and reserve frontier models for exceptions [14]. It says customers including IG, Airwallex and Worldpay have saved more than 250,000 hours [14]. Garry Tan argues that lab harnesses have an incentive to burn tokens, which leaves real room for startup-built harnesses [15].
- **Era** launched free. It generates a simulated enterprise across Salesforce, Slack, Jira, Zendesk and other tools with known ground truth, so agents can be benchmarked and post-trained on their failures [16].
- **Cognition’s “Dreaming”** builds Devin a memory graph across sessions and cleans it up overnight. Cognition is also proposing an open Agent Memory Repo standard [17]. Harrison Chase’s open question is how inferred memories get validated before an agent uses them [18].

Deals and early-stage signals

- **Sensible Bio (YC S21)** raised a \$47M Series A to make mRNA inside engineered living cells. Current cell-free methods, it says, can’t supply the doses and purity that therapies like in vivo CAR-T and gene editing need. Its platform already makes about 1,000 sequences a week [19].
- **Jeevy Fabrication** makes rocket tanks, piping and structures for customers including SpaceX and Lockheed Martin. Standard Capital led its Series A. The founder previously worked on Starship ground-support equipment, and the company is using AI to get parts delivered to spec and on time [20].
- **SF Tech Week** lists more than 1,700 events, over double the count two years ago. Hardware-themed events tripled. Cybersecurity events more than tripled, driven by AI security as companies deploy agents. Deep tech doubled [21].
- **Yandex Music’s Sona** replaced more than 15 candidate generators and

its ranking stack with a single transformer in an A/B test. It raised active users 4.53% and listening time 6.30%, but catalog coverage fell and it has not shipped to full traffic [22].

Policy

Sam Altman argued for lighter-touch regulation that accepts “some bad things” while guarding against catastrophic loss of control. He said he would not support licensing for models a generation or two behind the frontier [23]. He said tight silicon wafer supply, not financing, is driving up costs, and that he expects OpenAI to be ready for an IPO once it has “a few safety cases under our belt” [23]. He believes OpenAI’s best internal model is “well ahead” of the best Chinese ones but can’t say by how much [23]. Scott Kupor says the White House Super Intelligence Force will advise both on staying ahead and on doing so responsibly [24]. Brad Feld and others launched Founded Colorado, an advocacy group that came out of the fight over Colorado’s SB 24-205 AI law [25].

Sources

1. The Current State of Consumer AI
2. X post by @a16z
3. X post by @a16z
4. X post by @a16z
5. The Key Anthropic Numbers You Need To Know Ahead Of Its IPO
6. X post by @andrewchen
7. X post by @reflection_ai
8. X post by @HarryStebbing
9. Can You Still Win in Venture Without a \$1BN Fund? Menlo’s Venky Ganesan on the New Rules of VC
10. X post by @HarryStebbing
11. X post by @saranormous
12. X post by @HarryStebbing
13. X post by @ClementDelangue
14. X post by @_aj
15. X post by @garrytan
16. X post by @OfirEhrlich
17. X post by @cognition
18. X post by @hwchase17
19. X post by @ycombinator
20. X post by @daltonc
21. X post by @andrewchen
22. r/MachineLearning post by u/SettingAccording8986
23. Sam Altman: The benefits of AI are worth ‘some bad things’
24. X post by @skupor
25. X post by @bfeld