

# OpenAI's Home Device Plans, DeepSeek's \$7.4B Raise, and Agent Demand Surge

AI High Signal Digest

2026-07-15

## OpenAI's Home Device Plans, DeepSeek's \$7.4B Raise, and Agent Demand Surge

*By AI High Signal Digest • July 15, 2026*

OpenAI's reported home-device plans, DeepSeek's large financing round, and surging Codex and ChatGPT Work use lead the day's AI developments. The brief also covers efficient local models, active video reasoning, research-agent tooling, and emerging AI governance proposals.

### Top Stories

*Why it matters: AI's competitive landscape is widening beyond model releases into consumer hardware, capital formation, and rapidly scaling agent use.*

- **OpenAI's first hardware device is reportedly a mobile, screen-free smart speaker designed as a home AI companion.** Bloomberg reporting describes a battery-powered device that can be carried around the home, control appliances, play media, answer messages, and use cameras and sensors to understand its surroundings. It is also reported to learn from habits, context, and personal information including emails. [1, 2]
- **DeepSeek has reportedly raised \$7.4 billion as its ARR approaches \$500 million.** The company is said to maintain a gross margin above 50% on access to its V4 flagship model, while structuring a second round aimed at overseas, particularly Middle Eastern, dollar investors and preparing for a Shanghai STAR Market IPO next year. [3]
- **OpenAI says it has reached 8 million active users across Codex and ChatGPT Work.** The company reset usage limits without a five-hour cap for GPT-5.6 Sol; Sam Altman said demand growth was "insane" and warned that further scaling could bring service hiccups. [4, 5]

## Research & Innovation

*Why it matters: advances in local inference and active information retrieval could make capable systems more efficient to deploy and operate.*

- **PrismML introduced Bonsai 27B, which it calls the first 27B-class multimodal model to run on a phone.** Based on Qwen3.6 27B, it supports multi-step reasoning, tool use, long-context workflows, and agentic loops. Its phone-oriented 1-bit variant is 3.9 GB, while a 5.9 GB ternary version targets laptops; both are open-sourced under Apache 2.0. [6]
- **OmniAgent-7B frames long-video understanding as an active evidence-search task.** On LVBench, the 7B model scored 50.5 versus 47.3 for Qwen2.5-VL-72B while using about 203 frames rather than 768. It iteratively requests only needed frames or audio, summarizes findings into memory, and can stop once it considers the evidence sufficient. [7]

## Products & Launches

*Why it matters: research-agent, voice, and search capabilities are moving into directly accessible products.*

- **Perplexity open-sourced WANDR**, a 500-task benchmark built from real knowledge work for agents that must research both broadly and deeply. Wide Research is now available in the Perplexity Agent API, using a “Search as Code” architecture intended to execute a designed research plan deterministically at scale. [8, 9, 10]
- **OpenAI shipped a full-duplex conversation engine for GPT.** A live test described the system as enabling real conversation, while noting remaining quirks. [11]
- **Google is adding image generation to AI Overviews in Search.** The feature uses the Nano Banana model to create custom visuals from text prompts directly in results pages, with an English rollout planned over coming weeks in supported regions. [12]

## Industry Moves

*Why it matters: AI companies are pairing agent products with capital and distribution to expand into established enterprise workflows.*

- **Cognition says its first year after acquiring Windsurf lifted revenue run rate from \$73 million to more than \$500 million.** The combined company launched models including SWE1.7, Devin Review and CLI, and Devin Desktop; it is also developing event-driven agents and agent swarms. [13, 14]

- **ChaiDiscovery raised a \$400 million Series C at a \$3.8 billion valuation.** The company says it is pursuing medicine design at engineering-like precision and scale, and has deployments with Eli Lilly, Pfizer, and Novartis. [15, 16]

## Policy & Regulation

*Why it matters: government contracting and corporate-law proposals are beginning to define how advanced AI can be deployed and governed.*

- **A DeepMind researcher, citing The Information, said Google’s classified Pentagon agreement permits “any lawful government purpose.”** The reported terms require Google to help adjust safety settings at government request and provide no veto over lawful operational decisions. More than 600 DeepMind employees signed an open letter opposing deployment of models on classified networks. [17, 18]
- **Delaware is discussing an “Artificial Intelligence Company” legal entity for businesses run autonomously by AI agents.** Norm Ai is leading a public-private partnership to further develop the proposed framework. [19]

## Quick Takes

*Why it matters: capabilities continue to spread across science, speech, hardware, and local deployment.*

- Meta reported that its model scored **30/30** on the Asian Physics Olympiad theoretical exam, tying the top three student contestants. [20]
- Tencent released 1-bit and 4-bit versions of its 295B-parameter Hy3 model, saying it can be served on one GPU. [21]
- NVIDIA introduced its Vera Arm CPU, claiming up to a **1.9×** speedup for the CPU-intensive portion of agent loops. [22]
- Anthropic launched Claude for Teachers, offering verified U.S. K–12 educators free premium access, teaching skills, and standards-mapped curricula. [23]

---

## Sources

1. X post by @TheRunDownAI
2. X post by @kimmonismus
3. X post by @jukan05
4. X post by @thsottiaux
5. X post by @sama
6. X post by @PrismML
7. X post by @ZhihuFrontier

8. X post by @perplexitydevs
9. X post by @perplexity\_ai
10. X post by @AravSrinivas
11. X post by @thursdai\_pod
12. X post by @Google
13. X post by @ScottWu46
14. X post by @jeffwang
15. X post by @joshim5
16. X post by @saranormous
17. X post by @BlackHC
18. X post by @BlackHC
19. X post by @johnjnay
20. X post by @AIatMeta
21. X post by @TencentHunyuan
22. X post by @Kseniase\_
23. X post by @claudeai