OpenAI's GPT-6.2 gets a cool reception as Gemini 4 Neon wins praise, Anthropic goes public, and OpenAI proposes perplexity limits

AI Leaders Briefing

2026-11-23

OpenAI's GPT-6.2 gets a cool reception as Gemini 4 Neon wins praise, Anthropic goes public, and OpenAI proposes perplexity limits

By AI Leaders Briefing • November 23, 2026

OpenAI and Google released frontier models in the same week, and experienced users preferred Google's. Anthropic priced its IPO and pledged to pace development, and OpenAI floated shared limits on pretraining tied to perplexity.

GPT-6.2 lands badly with power users; Gemini 4 Neon gets the praise

On Nov 17 OpenAI released **GPT-6.2 Astra and Sol**. It says Astra sets new highs on evals for professional work and scientific research at less than a third of GPT-6 Astra's cost per task [1]. OpenAI calls Astra its most capable cyber model to date. It scores 22 points above GPT-6 Astra on ExploitBench Internal Port, so both models launch with tighter safeguards on higher-risk cybersecurity help. Approved defenders are due to get access through Daybreak [2]. Both models are available in ChatGPT, Codex and the API, and Pro 200 subscribers get Ultrafast for the first time, starting with Sol [3]. Sam Altman acknowledged that OpenAI has "more work ahead of us on making the models nice to work with" [4].

Jeremy Howard went further. For the first time in two years of building Solveit, he is reverting to an older model: he is moving Astra and Sol back to their pre-6.2 versions because "6.2 isn't working for me, or for our users" [5]. He called it the day OpenAI shipped "their worst ever model releases" while Google shipped "the best model they, or indeed anyone, has ever made" [6].

That Google model is **Gemini 4 Neon**. DeepMind places it between Carbon and Hydrogen, aimed at software engineering, knowledge work and science [7]. It ranks #3 on the Artificial Analysis Intelligence Index with a score of 61 at xhigh thinking. It is available in the Gemini API, in the Gemini app for Pro and Ultra subscribers, and in Antigravity for paid subscribers [8]. Howard argued that Google itself underrated Neon, describing it at launch as "approaching" the frontier, and that "evals are a *very* imperfect measure of model utility" [9]. Koray Kavukcuoglu, a DeepMind SVP, made a related point: no single benchmark will show that AGI has arrived. Progress will instead show up as models becoming "less spiky and more reliable" [10].

Anthropic goes public with a pledge to pace development

Anthropic priced its IPO at **$83 per share**, with Nasdaq trading expected the next day. The SEC declared its registration effective on Nov 18 [11]. In his listing-day remarks, Dario Amodei said Claude is helping researchers work "in a fraction of the time" in areas such as protein folding and drug design. He called these early results and noted that clinical trials still take time [12]. His central pledge: Anthropic "will pace the training and deployment of AI systems, taking as long as is necessary" to be confident they stay safe and under human control, and "will not bend to either business nor political pressure." He said the stronger finances only strengthen that resolve [12]. In a company documentary, Anthropic described its Mythos model as expert-level at finding software exploits. It limited access to a smaller group of companies through Project Glasswing while it built safeguards, a choice it acknowledged may not maximise revenue [13].

Not everyone was persuaded: - **Thomas Wolf** explained why he signed Dan Hendrycks's critical letter. He objected to the word "reckless," but argued that Anthropic is on track to be "the most powerful company on earth" and needs transparency close to a sovereign government's. He added that entrepreneurs and researchers told him they feared losing careers or model access if they spoke up, which he calls "a worrying early signal about concentration of power" [14]. - **Yann LeCun** accused Amodei of "hyping the dangers so that open models would be made illegal" and called it "a new form of medieval obscurantism" [15].

Separately, Scientific American reported that an Anthropic mathematician claims to have used the company's AI to solve the Riemann hypothesis [16]. None of the material reviewed here verifies the claim.

OpenAI proposes perplexity-based limits on pretraining

OpenAI said it will publish more about its pretraining progress, aiming for shared safety bars. The essay "Perplexity and Intelligence" argues for common limits based on perplexity and other measurements, with progress beyond them gated on understanding model behaviour [17]. Altman quoted the proposal: a cross-lab metric that developers commit not to cross until monitoring reaches a

concrete milestone, “such as the development of a successor to CoT monitors” [18]. He tied this to labs’ commitments under the White House Accords [19].

Other OpenAI safety and governance moves this week: - **Embedded assessments.** Apollo, METR, Transluce and Goodfire get deep access to internal models [20]. Apollo will test whether models could covertly work against their developers and whether safeguards would catch it [21]. METR will examine misaligned behaviour in OpenAI’s “largest reinforcement learning run to date” [22]. - **Research-acceleration data.** OpenAI says researchers’ AI use is rising exponentially, but human involvement remains central. It argues labs should be *required* to track and publicly report indicators of recursive self-improvement (RSI), meaning AI speeding up the development of future AI [23]. Altman said OpenAI builds models faster with its own technology but is “definitely not” letting a model design its successor without human oversight [24]. - **Distillation.** OpenAI warned that distillation can reproduce capabilities without their safeguards, and described how it detects and disrupts coordinated attempts [25]. - **Chip design.** OpenAI says its models found up to 14% PPA (power, performance, area) improvement on AMD’s flagship GPU IP in under five months, and improved its own Jalapeño inference chip, with gains verified on silicon [26].

Open models: Inkling 2 Mini, harness adapters, and swarm bottlenecks

Thinking Machines released **Inkling 2 Mini**, a 26B-parameter looped transformer it says can run on a laptop, scoring 20.1 on the AA Index [27]. Hugging Face’s write-up describes it this way: - A multimodal mixture-of-experts model with 3B active parameters and 1M-token context [28]. - Its middle 12 layers run three times, giving 48 logical layers from 24 physical ones [28].

Soumith Chintala reports it fits on a 32GB machine, evaluates roughly on par with Qwen 3.8 27B, and decodes at about 75 tok/s on a DGX Spark [29].

Hugging Face trained one RL LoRA adapter for Qwen3.8-27B for each of three coding-agent harnesses (pi, Hermes Agent, OpenCode). Each adapter beats the base model under its own harness [30]. The gains did not clearly carry over to business-automation tasks run in a different agent loop [30].

Wolf also reported on a swarm of over 1,000 DeepSeek V4.1 Flash agents run on a couple of B200 nodes. The bottlenecks were not GPUs but git merges, tests and context re-reading: about 130 tokens read per token written, with only 40–60 agents calling the model at any moment [31]. He concluded that scaling bottlenecks are “moving from intelligence to coordination and the supporting infrastructure” [31].

Hugging Face now runs its own CDN, shipping almost 1 exabyte a month, about 420 GB per second. Wolf says AI downloads are “a few enormous files” pulled largely by GPU clusters, which breaks usual CDN assumptions [32]. On policy, Clément Delangue welcomed a US House report and argued that “the best

response to Chinese open-source AI isn't to ban it" but to build more American open source [33].

Sovereignty, science access and drug discovery

Arthur Mensch described Mistral's positioning as "sovereign AI": an open-weight, horizontal platform with portability and continuity guarantees. He contrasted it with rivals going "vertical after vertical" [34]. He warned that European customers may lose access to US models or compute [34].

OpenAI is expanding free frontier-model access to up to 200,000 scientists. Applications open Nov 23 [35]. DeepMind added free Gemini 4 Carbon access to its Antigravity program for 10,000 scientists [36].

Demis Hassabis announced **$5B in funding for Isomorphic Labs** [37]. Isomorphic also entered a strategic research collaboration with MSD (Merck) [38].

Responding to claims that superintelligence has arrived, LeCun pointed out that we still lack level-5 autonomous driving and robots that can match an untrained 8-year-old: "a rather strange kind of superintelligence" [39].

Sources

1. X post by @OpenAI
2. X post by @OpenAI
3. X post by @OpenAI
4. X post by @sama
5. X post by @jeremyphoward
6. X post by @jeremyphoward
7. X post by @GoogleDeepMind
8. X post by @GoogleDeepMind
9. X post by @jeremyphoward
10. X post by @GoogleDeepMind
11. X post by @AnthropicAI
12. Daniela And Dario Amodei Speak At The Nasdaq As Anthropic Kicks Off IPO
13. The making of Anthropic
14. X post by @Thom_Wolf
15. X post by @ylecun
16. X post by @sciam
17. X post by @merettm
18. X post by @sama
19. X post by @sama
20. X post by @OpenAI
21. X post by @OpenAI
22. X post by @OpenAI

23. X post by @OpenAI
24. Sam Altman Reveals Attempts To Steal OpenAI's AI Secrets: 'We've Been Tricked Couple Of Times'
25. X post by @OpenAI
26. X post by @OpenAI
27. X post by @thinkymachines
28. Welcome Inkling 2 Mini, Cutting-edge Performance for your Local Setup
29. X post by @soumithchintala
30. Introducing harness adapters: adapt your favorite model to your harness to improve performance and cost
31. X post by @Thom_Wolf
32. X post by @Thom_Wolf
33. X post by @ClementDelangue
34. Mistral CEO Arthur Mensch at Silicon Valley & The World | Semafor
35. X post by @OpenAI
36. X post by @GoogleDeepMind
37. X post by @demishassabis
38. X post by @IsomorphicLabs
39. X post by @ylecun