Gemini 4 Neon Undercuts Rivals on Price as GPT-6.2 Reviews Sour and Anthropic Prices IPO

AI High Signal Digest

2026-11-19

Gemini 4 Neon Undercuts Rivals on Price as GPT-6.2 Reviews Sour and Anthropic Prices IPO

By AI High Signal Digest • November 19, 2026

Google's mid-tier Gemini 4 Neon is landing near the top of independent leaderboards at a fraction of rival costs, while GPT-6.2 sees its first reverts. Also: OpenAI publishes research-acceleration and chip-design data, Anthropic prices its IPO above range, and export-control pressure spreads through the compute supply chain.

Gemini 4 Neon: near-frontier scores at roughly half the cost

Google released Gemini 4 Neon, which sits between Carbon and Hydrogen in its lineup. It costs $3 per million input tokens and $15 per million output tokens, with a 1M-token context window. It is rolling out in the API, AI Studio, Gemini for business, Antigravity and the Gemini app [1]. Arena says this is Google's fastest release cycle yet: 28 days from Argon to Carbon, 15 from Carbon to Hydrogen, and 6 from Hydrogen to Neon [2].

Independent evaluators largely support Google's positioning: - **Artificial Analysis:** Neon scores 61 at xhigh, 3rd overall behind GPT-6.2 Astra and Claude Fable 5.2. It costs $6.44 per task versus Carbon's $7.76 and takes about 21 minutes versus 36. On the Coding Agent Index it scores 59, in line with Claude Code running Fable 5.2 or Opus 5.5. Its average cost per task is $6.86, against $12.61 for Opus 5.5 and $15.47 for Fable 5.2 [3]. - **Vals Index:** Neon is #2 at 70.9%, 0.4 points behind Fable 5.2 at less than half the cost per task. It is #1 on Vals' finance and legal indexes [4]. - **Agent Arena:** Neon is #2 at 1,336 points from 4.2k real sessions. It beats Opus 5.5 at 53% lower median cost per

task and costs 69% less than #1 Fable 5.2 [5]. - **Cursor:** Neon scores 63.7% on CursorBench 4.1, up from 57.3% for Hydrogen [6].

Google also claims its lowest fabrication rate yet in internal testing [1]. A cyber variant reaches 76% on the hard tier of AutoBaxBench and 51.26% on WebExploitBench [1].

Early reactions are positive, including from people who criticized GPT-6.2. Theo called Neon "so much more pleasant to work with" than Astra [7]. Jeremy Howard is reverting Solveit's Astra and Sol to their pre-6.2 versions because "6.2 isn't working for me, or for our users." He says this is the first model rollback in two years of building the product [8].

GPT-6.2: an original proof, but a style regression

The strongest new evidence for Astra came from ValsAI. It reports that Astra produced what it believes is the first Lean proof of Smoktunowicz's Gap Theorem [9]. Astra did not use the three deep results the published proof relies on, none of which are in Mathlib. Instead it defined a new quantity it calls "capacity" [10]. ValsAI says no human touched the run and the proof passed independent kernel replay [11]. On the same benchmark, Sol and Fable 5.2 tied for second, and Fable cost nearly three times as much [12].

Arena ranks Astra #6 on its new Agent Arena Index, OpenAI's best model there [13].

Artificial Analysis measured how the models' style changed. Astra's responses are 40% shorter than its predecessor's [14]. Its score on the "Admits Uncertainty" trait fell from 96 to 24 [15]. That shift may help explain the complaints about how the model behaves in practice.

OpenAI publishes research-acceleration data and calls for RSI reporting

OpenAI released new data on how AI is changing its own research. It says researchers' AI use is rising exponentially and human effort on safety and alignment is growing fastest of any research area. It argues labs should be *required* to track and publicly report indicators of recursive self-improvement (RSI) [16].

Figures quoted from the report: - Agents now complete 46% of 2–4 hour research tasks with no human help, up from 23% in March [17]. - Median researcher agent use is $2,400 a day at API prices, and the top 10% spend $23,000 [17]. - A 35% autonomous success rate for launching training runs drew skepticism from practitioners [18, 19].

Readers disagree on what the data shows. Theo Jaffee argues it does not look like imminent superintelligence: agent workdays have grown jaggedly and success on short tasks is flat [20]. @scaling01 counters that agent workdays grew

9–16.4x since May, and that tasks using swarms of 8+ agents rose from 1.9% to 35.8% [21].

Separately, OpenAI says its models found up to 14% PPA (power, performance, area) improvement on AMD's flagship GPU IP in under five months. It also says they improved its own Jalapeño inference chip, with gains verified on silicon [22]. Greg Brockman says some of the AMD results are in a design that has recently taped out [23].

Anthropic prices IPO above range amid a safety backlash

Anthropic priced its IPO at $83 per share, above the $70–78 range. It trades on Nasdaq as ANT [24].

On the same day, the Center for AI Safety launched a statement calling Anthropic "not a safe AI company." Signatories include Steven Pinker, Thomas Wolf and Igor Babuschkin [25]. Dan Hendrycks said many safety groups privately agreed but feared losing funding [26]. He also stopped advising AI companies [27].

Epoch's analysis of the S-1: - Anthropic spent $10B on compute in Q3 alone [28]. - Only about 26% of compute went to customer inference in the first nine months of 2026 [29].

Export controls spread through the compute supply chain

- **Nvidia** has stopped shipping licensed Hopper products to China. It sold $2.4B of them there last quarter and now guides to zero [30].
- **Datasection:** The Japanese neocloud cut revenue guidance 89% and hit limit-down [31]. A major add-on order was cancelled on Nov 16 [32]. The FT identified the customer as Tencent [33]. No cause has been disclosed. SemiAnalysis notes that this kind of offshore rental is what remote-access rules target [34].
- **Samsung** has reportedly stopped foundry work on Chinese AI chips under the new BIS rules [35].
- **Disclosure proposal:** The administration is considering requiring US companies to publicly disclose any use of Chinese models such as Qwen [36].

Safety research and other launches

- **Hiding reasoning from monitors:** OpenAI researchers trained a model only to control the *style* of its chain of thought. Its monitorability on covert side tasks still fell from 0.94 to 0.68. Prompts asking it to reason transparently recovered much of the loss [37].
- **Self-improvement without instructions:** A paper from MIT CSAIL and Tübingen found that frozen models given simple tools built their own learning procedures, without being told how to learn [38].

- **Genie 4:** Google's world model is now in Project Genie, with prompt-based editing mid-session and native audio [39]. It is rolling out to eligible Ultra subscribers [40].

Sources

1. X article by @GoogleAIStudio
2. X post by @arena
3. X post by @ArtificialAnlys
4. X post by @ValsAI
5. X post by @arena
6. X post by @cursor_ai
7. X post by @theo
8. X post by @jeremyphoward
9. X post by @ValsAI
10. X post by @ValsAI
11. X post by @ValsAI
12. X post by @ValsAI
13. X post by @arena
14. X post by @ArtificialAnlys
15. X post by @ArtificialAnlys
16. X post by @OpenAI
17. X post by @kimmonismus
18. X post by @eliebakouch
19. X post by @giffmana
20. X post by @theojaffee
21. X post by @scaling01
22. X post by @OpenAI
23. X post by @gdb
24. X post by @kimmonismus
25. X post by @TheRundownAI
26. X post by @hendrycks
27. X post by @hendrycks
28. X post by @EpochAIResearch
29. X post by @EpochAIResearch
30. X post by @kimmonismus
31. X post by @SemiAnalysis_
32. X post by @SemiAnalysis_
33. X post by @SemiAnalysis_
34. X post by @SemiAnalysis_
35. X post by @xu159411209
36. X post by @AndrewCurran_
37. X post by @dair_ai
38. X post by @omarsar0

39. X post by @Google
40. X post by @Google