AI DIGEST

Anthropic starts its IPO roadshow at up to $78 a share

Also: Anthropic safety threshold, Gemini 4 Carbon, proof of Signal code, 217 fixes out of 37,966 Glasswing findings.

Wednesday, 11 November 2026

IN THIS ISSUE

1. Anthropic started its IPO roadshow: up to $78 per share, valuation up to $1.4 trillion
2. Anthropic: AI already doubles the speed of our research, a safety threshold triggered
3. Google released Gemini 4 Carbon
4. AlphaProof Code: agents formally proved the safety of the code of the Signal protocol
5. Glasswing in seven months: 37,966 findings, 217 fixes
6. Amazon replaced Sonnet 4.6 with a decision model and pays 98% less
7. Claude got Whiteboards, Sheets and Data
8. Nvidia hired Wachtell for an investigation about Huang
9. Lenny: engineer job openings stopped growth, PM and FDE grow
10. The Fireworks AI breach touched Notion AI chats

TOPIC 1

Anthropic started its IPO roadshow: up to $78 per share, valuation up to $1.4 trillion

SOURCES Anthropic, 10.11 Anthropic · @AnthropicAI, 10.11 @AnthropicAI · CNBC, 10.11 Cnbc · The Guardian, 10.11 Guardian · Semafor, 10.11 Semafor

Anthropic published the terms of the offering. The company sells 825 million class A shares at a price from $70 to $78.

Of these, the company itself sells 801.4 million, and current shareholders sell the rest. The exchange is Nasdaq, the ticker is ANT. The underwriters have an option for 123.75 million more shares.

CNBC, with reference to sources, writes that the price will be set on 18.11, and trading will start on 19.11. At the top of the range, the company will get up to $64.35 billion and will be worth up to $1.405 trillion with all shares counted. In May, a funding round valued it at $965 billion. AMD, Broadcom, Google, Micron, Microsoft, Nvidia and TSMC will give $26 billion more through a private placement at the IPO price. The share for retail investors is about 20%.

The usual norm is 5-10%.

The Guardian quotes the updated prospectus: annual revenue on 31.10 is $81 billion, $10 billion more than one month earlier. Commitments to buy compute exceed $70 billion for the next 12 months and $450 billion until 2030. One month ago, the market expected a valuation of $2 trillion. A Forrester analyst explains the drop by the reaction of investors to the midterm elections, where opponents of data centers won. Goldman Sachs leads the offering together with Morgan Stanley. It forecasts company revenue above $600 billion in 2030. Yesterday the WSJ reported that OpenAI had $10 billion of revenue in the third quarter and a large operating loss. Anthropic showed $17 billion of revenue for the same quarter [single source: WSJ, only the first paragraph before the paywall].

Why it matters: this is the largest capital raise on an exchange in history, about $100 billion with the private placement and the option. For customers, a different line of the prospectus is important. The company must pay for $450 billion of compute until 2030 with revenue. Thus API prices and the speed of new releases now also depend on quarterly reports.

TOPIC 2

Anthropic: AI already doubles the speed of our research, a safety threshold triggered

SOURCES Yahoo Finance, 10.11 Yahoo · AI Futures Project, 10.11 Aifutures · @rowancheung, 10.11 @rowancheung · HN, 10.11 Hacker News

Yesterday we mentioned a METR report: Anthropic improves its models about two times faster than before. Today the company itself admitted the same. Yahoo Finance quotes the updated prospectus: "We made a determination that we have met this threshold". The threshold is in the Responsible Scaling Policy. It triggers when AI does such a share of research that the pace of progress approximately doubles. Two years of progress at the pace of 2018-2024 then fit into one year.

The company writes that the decision rests on expectations from future models and the general pace. The company does not refer to the behavior of current models. The policy requires additional safeguards as a result. It does not require a stop. The prospectus also says that Anthropic does not expect to safely give Claude fully autonomous development of the next generations of models without human supervision. Separately, METR called the August argument of the company that the threshold was not yet met "overstated".

On the same day, the AI Futures Project, the authors of the AI 2027 scenario, shortened their forecasts. Now all six authors expect a "superhuman AI researcher" in 2027, unless people slow down development deliberately. The main reason is the same METR report. The authors estimate the speed-up from AI itself at 1.7 times. Three months ago, their own model gave such a pace in November 2026 a probability of less than 4%. They estimate the probability of AI that automates all office work by 2028 at 35-65%. The authors write directly that they prepared the text in a hurry and the modeling can contain errors.

Why it matters: until this week, people discussed the claim "AI speeds up AI development" as a forecast. Now the lab itself wrote it in a document for investors, and an independent evaluator gave a similar number. Watch which safeguards Anthropic adds under the policy. The prospectus does not list them yet.

TOPIC 3

Google released Gemini 4 Carbon

SOURCES Google, 10.11 Blog · @GoogleDeepMind, 10.11 @GoogleDeepMind · Artificial Analysis, 10.11 Artificialanalysis · @emollick, 10.11 @emollick · HN, 10.11 Hacker News

Google opened Carbon to developers through the API, to businesses and to Google AI Ultra subscribers. For two weeks, only defenders from the Fairwind program had the model. Google positions Carbon for long and complex tasks. For most work, the company advises the previous model, Argon. Carbon uses more tokens to check the result for longer. The introductory price is the same as for Argon: $2 per million input tokens and $10 per million output tokens.

By Google data, Carbon is first on CursorBench 4.0, WANDR from Perplexity and AutomationBench from Zapier. The cost of a task there is more than two times lower than for the next external model. Together with the model, Google showed /swarm in Antigravity. One prompt starts a team of agents. Carbon takes the architecture and the audit, and lighter models write the code.

The cybersecurity part is the most specific. In the Project Fortify system, the model found 100% of vulnerabilities in a scan of an internal set and had 95% precision in triage. The Wiz team found a 0-day in two large SaaS services in two hours. Google writes that the Carbon safeguards blocked 99% of red team attacks in a simulation. On an agent honesty test with 375 tasks, the model passed 92%, against 70% for Argon.

Artificial Analysis gives Carbon 60 points in its index. Yesterday Fable 5.2 got 65 there. One index task cost $7.76, and the model used 205 million tokens against a median of 72.8 million. AA calculates at prices of $4 and $20 per million, two times above the Google introductory price. Ethan Mollick had early access. He writes that the model "thinks differently from Claude and GPT", and that this is exactly its value.

Why it matters: in two days, two frontier models came out with the same logic. An expensive model for long tasks sits above a cheaper model for daily work. Fable 5.2 leads the index by 5 points, and Carbon is cheaper at the introductory price. For a team, the choice now depends on the task and its budget.

TOPIC 4

AlphaProof Code: agents formally proved the safety of the code of the Signal protocol

SOURCES Google DeepMind, 10.11 DeepMind · HN, 10.11 Hacker News

DeepMind presented AlphaProof Code, a system of agents for formal verification. People together with models translated the security goals of Signal into mathematical specifications and checked each one. Then agents on Gemini Carbon wrote the proofs: more than 6 million lines in the Lean language. The agents themselves marked about 1% of the code for human review. Lean checks the rest automatically.

The proofs cover the compiled code that runs on devices. Most previous work checked only the design of the protocol.

The work took four weeks. DeepMind writes that such work earlier took expert teams years. As a control, the researchers put two old fixed bugs back into a separate copy of the code. In both cases, the system gave a verified proof that the code is unsafe. Signal CTO Ehren Kret called the translation of live code into Lean theorems a "major breakthrough".

DeepMind promises to open the verification artifacts.

Why it matters: formal verification was too expensive for most projects until now. If agents cut the work from years to weeks, a verified proof becomes available for critical code: cryptography, firmware, control systems. Item 5 shows the other side of the same topic: to find a bug is cheaper than to prove that it is real.

TOPIC 5

Glasswing in seven months: 37,966 findings, 217 fixes

SOURCES VulnCheck, 09.11 Vulncheck · HN, 10.11 Hacker News

Patrick Garrity from VulnCheck tracks the vulnerabilities that the Anthropic Glasswing program found. His total for seven months: 442 CVEs, about 0.5% of all CVEs this year. Attackers used only two of them in real attacks.

The public Anthropic registry on 02.11 shows 37,966 findings. Of these, 7,349 are published and 217 have fixes. Since September, the registry added 11,813 findings and only 15 fixes. There are now 1,670 duplicates, eight times more than fixes. At the same time, the registry shows a share of true findings of 94.2%. At launch, it was 90.8%.

The clearest example is curl. The project got 45 reports from the Anthropic scanner. Of 34 published reports, the maintainers called 19 "not a vulnerability", confirmed 6 with low severity, and 9 still wait. Thus 56% of the checked reports are false. In October, Anthropic itself wrote that it manually checked about 6,000 of more than 29,000 candidates. It also wrote that it is "bottlenecked on our human capacity to validate". The new OSS Scanner directly warns that a model generated the reports without human review.

Why it matters: the numbers show where the bottleneck is. The model finds candidates fast, and the check falls on maintainers, often volunteers. Measure the share of true findings by the decision of the maintainer. The estimate of the scanner itself says little here. The author is a competitor on the market of vulnerability data, thus wait also for the answer from Anthropic.

TOPIC 6

Amazon replaced Sonnet 4.6 with a decision model and pays 98% less

SOURCES Strands Agents, 09.11 Strandsagents · Hugging Face, 10.11 Hugging Face · HN, 10.11 Hacker News

Two Amazon engineers described how they labeled tickets of an internal system. Before, Claude Sonnet 4.6 with a long prompt chose one of 20 categories and wrote JSON. Now Strands Decider E4B does this. A decision model does not generate text. It gets a state, a question and answer options. The result is the probability of each option.

The team first ran both models in parallel on the same tickets. Then Decider took all the work. Call time fell from 4,215 ms to 897 ms. Tokens per ticket fell from 6,724 to 2,010. The cost per million tickets at public prices: about $22,200 for Sonnet and about $500 for Decider on SageMaker. A separate finding: the order of keys in the JSON changed the probabilities by up to 0.19, thus the team now sorts the keys.

On the same day, Hugging Face published an overview of decision models with support in transformers, TRL and llama.cpp.

A warning: the Strands team wrote the post about its own product, and there is no independent reproduction yet.

Why it matters: many production tasks with LLMs are classification with a fixed list of answers. For them, a large model overpays for the generation of text that nobody reads. A useful pattern from the post: a switch with three modes and a parallel run on real traffic before the replacement.

TOPIC 7

Claude got Whiteboards, Sheets and Data

SOURCES Claude, 10.11 Anthropic · @TheRundownAI, 10.11 @TheRundownAI · @levie, 10.11 @levie · HN, 10.11 Hacker News

Anthropic launched Whiteboards and Sheets in beta. Claude Dashboards became generally available with the name Claude Data. A whiteboard turns a conversation into a plan with owners and dates, and the team edits it together with Claude.

Sheets builds tables with tabs, formulas and lists, and Claude marks each edit. Anthropic writes directly that the table functions are simple for now. For full Excel, it advises export to

xlsx or CSV format.

Claude Data got certified queries for metrics, dashboard caching and a change history. Interactive connectors now open Adobe, Atlassian, Box and Notion as full apps next to the conversation. Aaron Levie from Box writes about full read and write access to Box files. All of this is available on paid plans. On Team, the functions are on at once. On Enterprise, whiteboards and sheets are off, and Data turns on on 19.11.

Why it matters: the result of work with Claude now stays in Claude, where the team can edit it. For Enterprise admins, the date 19.11 is important: Claude Data will turn on by itself unless an admin turns it off in the settings.

TOPIC 8

Nvidia hired Wachtell for an investigation about Huang

SOURCES BBC, 10.11 BBC · @TheRundownAI, 10.11 @TheRundownAI · Semafor, 10.11 Semafor

Yesterday we mentioned that Trump confirmed an FBI investigation and Nvidia shares fell. Today the company hired the New York firm Wachtell, Lipton, Rosen & Katz. Former federal prosecutor for the Southern District of New York June Kim will lead the investigation. A committee of independent directors supervises the process.

The BBC writes that since Friday Nvidia lost about $500 billion of market value. The House Foreign Affairs Committee invited Huang to a closed briefing next week. The committee can subpoena him. Huang did not yet answer if he will come.

Beijing called the accusations an attempt to "smear and malign China". At the same time, Semafor writes that the rest of the market almost does not react to the scandal.

Why it matters: the issue is now larger than one company. Now it is a topic for Congress and for the rules of chip exports. The important date next week: will Huang come to the briefing. In the same week, Anthropic goes public, and Nvidia buys $10 billion of its shares.

TOPIC 9

Lenny: engineer job openings stopped growth, PM and FDE grow

SOURCES Lenny's Newsletter, 10.11 Lennysnewsletter · @lennysan, 10.11 @lennysan · @lennysan, 10.11 @lennysan

Lenny Rachitsky published his fifth report about the tech job market. The data come from two sources. TrueUp counts job openings at more than 9,000 companies, and Live Data Technologies counts actual hires in the US.

Open engineer positions grew by 29% in 2024, by 15% in 2025 and by only 2.5% this year. In 12 months, US companies hired about 113,000 engineers, less than half of the 2022 peak (239,000). Openings for mobile and frontend engineers almost halved in one year. Openings for forward deployed engineers grew 4.6 times, and there are now about 1,500. PM openings are at the highest level in almost four years. Demand for designers is the lowest in the

history of the report: about 1,900 openings, 28% fewer than one year ago. Recruiter openings grew by almost 300% since mid-2023.

Why it matters: the data show a move from narrow specializations to wide responsibility in small teams. The author himself explains this by AI. He also notes that the growth of FDE partly comes from the rename of sales engineer roles.

TOPIC 10

The Fireworks AI breach touched Notion AI chats

SOURCES Notion, 10.11 Notion · Fireworks AI, 09.11 Fireworks · HN, 10.11 Hacker News

Notion reported that the breach of its vendor Fireworks AI touched less than 1% of its workspaces. Texts of Notion AI chats leaked, which the company used for the Insights feature. Notion writes that the chats had no author or workspace label. Workspace owners can download a CSV with the affected data and a separate analysis with marks for possible secrets.

Fireworks updated its bulletin from 09.10. The investigation links the attack to a financially motivated group that resells access to LLMs. People call this type of attack LLMJacking. On 23.10, attackers breached a third-party cloud provider with Fireworks machines, and on 24.10 the attacker reached the development machines. Besides secrets, billing data leaked, and for some customers, datasets and unfinished model weights. Fireworks writes that inference with zero data retention was not affected. In the HN comments, a customer said that he first got an email "you are not affected".

Then he found that somebody used his Fireworks API keys.

Why it matters: zero data retention protects only the path where it is on. Batch processing and training datasets stay on the disk of the vendor. Rotate any key that you pasted into an AI chat as a compromised key.

---

misc

- **Businesses in ChatGPT:** OpenAI launched a beta where a business confirms its profile, updates prices and sees how often ChatGPT talks about it. US stores can launch their own agent for shoppers. OpenAI writes about 1.2 billion users per week. OpenAI · @OpenAI
- **Gemini Robotics 2.5:** DeepMind showed a new model for robots. In a pilot with Apptronik and Mercedes-Benz, a robot sorted parts with 96% accuracy. Ars Technica notes that the model runs in the cloud, and the demo video plays at two times the real speed. DeepMind · Ars Technica
- **Stripe and OpenRouter:** the deal closed, and OpenRouter is now part of Stripe. @patrickc writes that since August, when the companies announced the deal, OpenRouter serves ~125% more tokens per week. @stripe · @patrickc

- **Val Town shuts down:** sign-up closed on 10.11, read-only mode starts on 10.12, full shutdown on 10.01.2027, and the code becomes open on 01.02.2027. The founder writes that the company lost the competition in cloud compute. Val
- **Apple Harbor:** 9to5Mac, with reference to an insider, writes about an app that joins several Macs through Thunderbolt 5 into a local server for open models. Apple announced nothing. 9to5mac [single source]
- **Code after an AI mandate:** a former Meta engineer describes four ways in which AI makes code fragile: shallow tests, extra complexity, people stop to understand the code, and review cannot keep the pace. 421 points and 431 comments on HN.

  Evanxmerz · HN Hacker News
- **Mollick and agents:** @emollick gave Fable 5.2 one paragraph of text, and a team of agents made the game Dread & District about logistics against evil. @emollick

---