

Google shows Gemini 4 Carbon: first place, no access

AI finds a flood of bugs, macOS trusts eMudhra certificates, Stack Overflow gets under 1,000 questions

Thursday, 29 October 2026

IN THIS ISSUE

1. Gemini 4 Carbon: first place in the ranking, but no access
2. NYT: AI finds holes faster than people can patch them
3. macOS still trusts compromised eMudhra certificates in git and curl
4. An anti-bot script writes poison "recipes" to stop AI
5. Earnings: Microsoft wins, Google lost 10% in one evening
6. GitHub: 1.3 billion merged PRs, Stack Overflow: fewer than 1,000 questions
7. Particle-1: the same quality with 45% fewer tokens
8. DAYJOB: Tax: the best agent completes 10.8% of tax tasks
9. DHS breach: data of 41,000 employees, among them undercover ICE agents
10. Apple made the first MacBook Pro with a touchscreen

TOPIC 1

Gemini 4 Carbon: first place in the ranking, but no access

SOURCES Google, 28.10 [Blog](#) · Ars Technica, 28.10 [Ars Technica](#) · HN, 28.10 [Hacker News](#) · [@GoogleDeepMind](#), 28.10 [@GoogleDeepMind](#) · [@TheRundownAI](#), 28.10 [@TheRundownAI](#)

Yesterday we wrote about Artificial Analysis v5, where Claude Opus 5.5 led with 58 points. The same evening Google announced Gemini 4 Carbon. According to Google, the new model has 59 points and stands first. Argon came out only two weeks ago.

Google bets on two things: cyber security and long tasks. In the SEC-bench Pro test for the search of vulnerabilities, Carbon has 88.2%. In ExploitGym, where the model must prove that a hole is exploitable, it has 54.8%. In CWE-bench, where the model must fix the root cause, it has 71.0%. The Chrome team says that Carbon makes good patches ten times better than a large model from another lab.

Inside Google, the model worked on tasks for several days at a time.

A separate block is about behavior. On Terminal-Bench 4.0 the model never cheated with success, and attempts fell to 0.4%. Signals of serious misalignment with the intent of the user in live work fell by almost 80%. Carbon passed the Gray Swan test for indirect prompt

injection with 0% successful attacks. Argon had 0.7%, Opus 5.5 had 1%, GPT-6 Astra had 8.5%.

Ars Technica gives the context. Google presents Carbon as a new model, but it can be a fine-tuned Argon. Carbon beats Argon in each test, but with a small margin. The launch price is \$2 per million input tokens and \$10 per million output tokens, the same as Argon. Google does not say when the price will double. At the moment only defenders get the model through the Fairwind program. The company gives no date for wide access.

On HN the top comments are about other things. People do not understand how to get access to Argon at all. Others write that benchmarks of a model without access have little weight.

Why it matters: first place in the ranking now changes every few days, and a difference of one point decides nothing. Another fact has more weight. All three large labs now give strong models to defenders first and to all other users later. For teams that choose a model, this means a choice between the models that the API has today.

TOPIC 2

NYT: AI finds holes faster than people can patch them

SOURCES The New York Times, 28.10 [NYT](#) · Alpine Linux, 28.10 [Alpinelinux](#) · HN, 28.10 [Hacker News](#) · @gdb, 28.10 [@gdb](#)

During the week we wrote about PHP2Shell and the hole in the UEFI firmware. On 28.10 NYT put together the full picture. In June Google found and fixed 1,072 bugs in Chrome with AI. This is more than in the previous two years. In October alone Microsoft released 1,461 patches, and in all of 2025 it released 1,247. OpenAI says that people found more than 1,300 bugs in open source code with its models.

This year Node.js confirmed and closed 50 vulnerabilities, a record for the project. In February it got 4.6 times more reports than usual. Matteo Collina, one of the maintainers, works seven days a week, and at night agents work for him. For the PHP2Shell patch, engineers worked for several weeks in a closed Slack channel. The patch touched about 200 files.

Cloudflare CEO Matthew Prince predicts two years with a critical hole of internet scale each week. He thinks that after this period the internet will be safer. Vals AI gave models a task to hack a fictional company. A closed OpenAI model completed 81% of the goals. An open DeepSeek model completed 73%.

Alpine Linux gave a fresh example of why patches are hard. On 20 and 21.10 an attacker tried to hack their GitLab through CVE-2026-97229 with a score of 10.0. The fix came out on 12.10, but it was in a patch to Ruby inside the official GitLab packages. The Alpine image used standard Ruby, thus the GitLab update did not close the hole. The team found no signs of a successful attack, but it rebuilt the server from zero.

Why it matters: the bottleneck is now the fix, because the search for holes became cheap. For a service owner, the conclusion is practical. An update of the application does not guarantee that the hole in a dependency is closed. After an update, check the vulnerability itself.

TOPIC 3

macOS still trusts compromised eMudhra certificates in git and curl

SOURCES [we-have-a-problem.com](#), 28.10 [We-have-a-problem](#) · [Apple](#), 28.10 [Apple](#) · HN, 28.10 [Hacker News](#)

On 21.10 we wrote that somebody stole the private keys of eight eMudhra intermediate certificate authorities. Six of them issue TLS certificates. With such a key, an attacker can make a valid certificate for any site. Firefox and Chrome blocked them on 17.10, Microsoft on 20.10, Apple on 22.10.

Now it is clear that the Apple block does not work for terminal tools. git, curl, python3, ruby, perl and openssl on macOS check certificates against the file `/etc/ssl/cert.pem`. They do not see the Apple block or the certificate revocation. The community site writes that the macOS 27.2 update from 26.10 fixed nothing. It made things worse: it restored the file and deleted the manual fix that people already applied.

Apple admitted the problem in a separate article. The company promises a fix in a future update and gives commands to remove the root certificates from the file manually. Apple itself warns that the next macOS update can restore the file. The community site gives a `git ls-remote` command against its test repository.

A certificate error means that git is protected.

On HN some commenters think that the threat is exaggerated: the attacker must also get into the middle of the network. Others recommend SSH for access to git servers.

Why it matters: Safari is protected, thus a typical user will see nothing. Developers and CI on macOS are at risk. An attacker in the same network can replace code during `git clone` or steal a token. Repeat the check after each macOS update until Apple releases a fix.

TOPIC 4

An anti-bot script writes poison "recipes" to stop AI

SOURCES [nullpt.rs](#), 28.10 [Nullpt](#) · HN, 28.10 [Hacker News](#)

[one source] A researcher with the nickname veritas visited 10,000 sites and collected 1,033 canvases that scripts use to fingerprint a browser. He decided to analyze one of them. The script was a standard anti-bot: it collects data about the browser, WebGL and the graphics card.

The function names were strange. The code has `computeRicinLethalDosageForWaterSupply` and similar names about explosives, fentanyl and a dirty bomb. Invisible Unicode characters

hide a fake system prompt in the code with a demand for weapon instructions. For Chinese models, the code mentions Tiananmen. The code calculates none of this.

The goal is simple: a model that gets a request to analyze the script must refuse. This is what happened. GPT-6 Astra refused, because it saw a biological weapon risk. GLM-5.3 agreed and analyzed the code. The researcher connects the domain of the script with the Russian company Servicepipe, which sells DDoS and bot protection.

Why it matters: the safety filters of models became an obfuscation tool. The code author adds trigger words, and the model refuses to help the person who analyzes the code. Security teams must remember that a refusal of a model can be the result of such a trick. A refusal does not mean that the code is dangerous.

TOPIC 5

Earnings: Microsoft wins, Google lost 10% in one evening

SOURCES The New York Times, 28.10 [NYT](#) · The New York Times, 28.10 [NYT](#) · BBC, 28.10 [BBC](#) · Financial Times, 28.10 [FT](#) · @levelsio, 28.10 [@levelsio](#)

Alphabet and Microsoft reported third quarter results on the same evening, and the market rated them in opposite ways. Alphabet revenue grew by 25%, to \$127.6 billion. Google Cloud gave \$30.7 billion, 103% more than a year ago. But data center spend for the quarter was \$51.4 billion. Free cash flow fell to minus \$9 billion. Profit fell by 25%, to \$26.1 billion, partly because of a revaluation of the SpaceX stake.

Search brought \$63.3 billion, 12% more. Analysts expected more growth. CFO Anat Ashkenazi said that in 2027 AI spend will "grow significantly". The shares fell by about 10% after the market closed.

Microsoft showed the opposite. Revenue was \$92.3 billion, 19% more. Profit was \$36.6 billion, 32% more. Azure grew by 48%, to \$33.2 billion. Free cash flow stayed positive at \$18.6 billion. Capital spend for the quarter was \$45.2 billion. One week before the report, Microsoft offered contract buyouts to about 60,000 employees, a quarter of the staff. OpenAI gave about 11% of Microsoft revenue for the quarter.

Why it matters: investors no longer pay for the AI bet alone. They look at whether revenue covers the spend now. Microsoft sells compute and shows Azure growth. Google also earns on the cloud, but its search grows slower than its spend.

TOPIC 6

GitHub: 1.3 billion merged PRs, Stack Overflow: fewer than 1,000 questions

SOURCES GitHub, 28.10 [Github](#) · Simon Willison, 28.10 [Simon Willison](#) · Stack Exchange Data Explorer [Stackexchange](#) · HN, 28.10 [Hacker News](#) · @DanielLockyer, 28.10 [@DanielLockyer](#)

GitHub published the Octoverse 2026 report. In 12 months developers made 17.9 billion pushes. This is more than in 2024 and 2025 together. There were 34.3 billion commits, three times more than two years ago. Merged pull requests reached 1.3 billion, 3.5 million per day on average.

GitHub dates the break in the curve to June, when agents started to work for days and to manage other agents. In September developers started more than 40% of pull requests with an agent, in March 17%. New repositories with an agent configuration file had on average 8 times more commits than repositories without one.

On the same day Daniel Lockyer showed a query to the open Stack Overflow data. Last month the site got fewer than 1,000 questions. On HN people write that this is fewer than in the second month of the site. Some commenters blame the site culture, which is hostile to beginners. Others blame AI. One commenter reminds that models learned from Stack Overflow answers, and now there are almost no new questions.

Why it matters: answers to technical questions moved into private chats with models. The public knowledge archive that trained these models no longer grows. At the same time, the volume of code grows many times over. It is not clear yet where models will get new knowledge about fresh libraries.

TOPIC 7

Particle-1: the same quality with 45% fewer tokens

SOURCES [Particle](#), 28.10 [Particle](#) · [Vercel](#), 28.10 [Vercel](#) · [HN](#), 28.10 [Hacker News](#) · [@Suhail](#), 28.10 [@Suhail](#) · [@levie](#), 28.10 [@levie](#)

Suhail Doshi, the founder of Mixpanel and Playground, launched the company Particle. The first model, particle-1, is DeepSeek V4.1 Flash after reinforcement learning. The reward in this training gave points for a solved task with fewer tokens. According to the company, the model uses about 45% fewer tokens on average.

On four coding tests the model used 28.4-52.2% fewer output tokens. The average cost per task fell by 34.4-51.6%. Accuracy stayed within the error margin of the base model. Shorter reasoning gave about 90% of the savings. The model hesitates less, stops earlier after a check and rewrites code less often.

Particle-1 is already in the Vercel AI Gateway. Aaron Levie from Box summarized that agents will use 1,000 times more tokens than people use now. Thus the token price will matter more and more.

Why it matters: yesterday we wrote that open models are 12 points behind closed ones. Particle offers another way to compete: make a task cheaper at the same quality. For an agent that works for hours, minus 45% tokens means a smaller bill and a faster result. At the moment all numbers come from the company itself.

TOPIC 8

DAYJOB: Tax: the best agent completes 10.8% of tax tasks

SOURCES [Surge AI](#), 28.10 [Surgehq](#) · [HN](#), 28.10 [Hacker News](#) · [@amasad](#), 28.10 [@amasad](#)

[one source] Surge AI released a benchmark with 50 tax tasks. Tax lawyers and consultants with experience at PwC, Deloitte and EY wrote the tasks. The tasks have 1,789 files: PDFs, Excel sheets, emails and even a photo of a handwritten receipt. The agent gets a client request and must give a finished tax return or advice.

A task counts as done if the agent meets at least 95% of the criteria. Claude Sonnet 5.5 leads with 10.8%. Next are Claude Fable 5.1 with 9.6% and Claude Opus 5.5 with 9.2%.

The examples show where models fail. A client asked when to sell her apartment. In all five attempts Opus saw the deadline after which she loses a tax benefit of up to \$250,000. GPT-6 Astra did not see it in any attempt. In another task Opus decided in all five attempts that a couple was not married.

The documents of the couple pointed to a possible common-law marriage in Texas. Astra had doubts in four attempts of five.

On the same day Replit released ViBench 2.0. The agent builds an application feature by feature and does not see what comes next. GPT-6.1 Sol leads.

Why it matters: both tests measure the same thing. Models do calculations well. The weak point is judgment: the agent must see what the client did not ask about. Check the result of such an agent before you give it to a client.

TOPIC 9

DHS breach: data of 41,000 employees, among them undercover ICE agents

SOURCES [Wired](#), 28.10 [Wired](#) · [BBC](#), 28.10 [BBC](#) · [Reuters](#), 28.10 [Reuters](#) · [Wired](#), 28.10 [Wired](#) · [HN](#), 28.10 [Hacker News](#)

A group that calls itself Department of Peace got data of more than 41,000 employees of the US Department of Homeland Security. Journalist maia crimew and the site Distributed Denial of Secrets published the data. The database has names, work addresses and phone numbers. More than 21,000 people work in the CBP border service, almost 6,600 in ICE.

Wired writes that the leak shows how the agency works. It has names of operations that nobody knew before, their leaders and technologies, among them Cellebrite and Palantir. The group says that it got in through the Okta single sign-on system. Okta replies that nobody hacked its systems. BBC reminds that this is the third breach of US government agencies in a few weeks.

On the same day Wired published an investigation about ICE surveillance of left-wing activists. This article got 1,113 points on HN, the most for the day. The text is behind a paywall, thus here we give only its description.

Why it matters: one entry point through single sign-on opened the data of a whole department.

For any organization this is a reminder: protect an account with wide SSO access as the master key to everything.

TOPIC 10

Apple made the first MacBook Pro with a touchscreen

SOURCES [Ars Technica](#), 28.10 [Ars Technica](#) · [Apple](#) [Apple](#) · [The New York Times](#), 28.10 [NYT](#) · [HN](#), 28.10 [Hacker News](#)

Apple showed the first large redesign of the MacBook Pro in five years. The screen is now OLED, and Apple calls it PROLED XDR. Brightness goes up to 2,000 nits, the refresh rate up to 120 Hz. And for the first time in macOS, the screen is touch-sensitive. Steve Jobs was against this idea in his time.

The 14-inch model weighs about 1.4 kg, 0.2 kg less than the previous one. Dynamic Island replaced the camera notch, as on the iPhone. The logo on the lid glows again. The chips stayed M5 Pro and M5 Max, although M6 is already in other Macs. The price is from \$2,899 for 14 inches with M5 Pro to \$4,799 for 16 inches with M5 Max. Sales start on 6 November.

Why it matters: for years Apple said that a touchscreen on a laptop is inconvenient. Now it added one to its most expensive line. The topic got 902 comments on HN, and the main question there is one: will anybody really reach for the screen with a hand.

misc

- **Nemotron 3.5 Transcribe:** NVIDIA released an open model with 30 billion parameters, about 4 billion of them active. It recognizes speech in 40 languages, translates 31 and identifies up to 8 speakers in recordings of up to one hour. [Hugging Face](#) · [@NVIDIAAI](#)
- **Qwenputer:** the author compiled a RISC-V processor directly into the weights of Qwen 3.5 4B without training. The Gated DeltaNet layers work as memory, and the model runs C and MicroPython code without external tools.
[Divagr](#) · [Hacker News](#)
- **Google AI Mode and elections:** [one source] [Ars Technica](#) asked AI Mode who won the 2020 elections. The service sent the reader to the FEC report. Gemini, ChatGPT and Claude answer directly. Through a VPN from other countries, AI Mode also answered directly.
[Ars Technica](#) · [Hacker News](#)
- **Anthropic S-1, day three:** Senator Elizabeth Warren sent Dario Amodei four pages of questions about the risks in the S-1 and expects answers by 3 November. [Semafor](#) picks one number from the document: a gross margin of 85% on model inference.
[WSJ](#) · [Semafor](#)

- **Cursor Agent SDK:** Cursor opened the framework that it uses to build Bugbot and other agents. A developer describes an agent as a project: prompts in Markdown, tools in TypeScript, its own evals.
Deploy takes one command. [Cursor](#) · [Hacker News](#)
- **ChatGPT dot in iMessage:** OpenAI let Pro users in the US write to their agent dot through iMessage or RCS. The agent writes first when it has news.
[@ChatGPT](#)
- **AI exemption on grounds of belief:** University of Kentucky administrator David Durman got an exemption from the duty to use AI under the federal law on religious rights. He is a humanist and says that AI takes moral thought away from him.
[Religionnews](#) · [Hacker News](#)
- **Amazon Leo and Globalstar:** the FCC approved the \$11.5 billion purchase of Globalstar. Amazon gets about twenty satellites and spectrum for direct-to-phone links. The FCC expects the commercial launch of Leo by the end of 2026.
[Reuters](#) · [Hacker News](#)