AI DIGEST

Jensen Huang resigns as Nvidia CEO

Also: DOJ probes Nvidia, Blackwell fingerprint in Chinese weights, Gemini 4 Hydrogen, Anthropic IPO numbers.

Friday, 13 November 2026

IN THIS ISSUE

1. Jensen Huang left Nvidia, Colette Kress is interim CEO
2. Washington after Huang: the DOJ examines Nvidia, Congress prepares export controls
3. Researchers found a fingerprint of Blackwell chips in the weights of six Chinese models
4. Google released Gemini 4 Hydrogen: a fast model, stronger than Argon in the AA index
5. Anthropic opened its IPO site: $17 billion of revenue per quarter and a $23.3 billion loss
6. Global Microsoft 365 outage: Outlook, Teams and Copilot were down for hours
7. Critical RCE in Apache HTTP Server, found by Anthropic and Google
8. Thinking Machines released the weights of Inkling 2 Small under Apache 2.0
9. OpenAI launched Daybreak Defense Factory and Daybreak Red Team
10. Agents with a wallet: Replit buys for business through Link and Visa

TOPIC 1

Jensen Huang left Nvidia, Colette Kress is interim CEO

SOURCES Nvidia, 12.11 Nvidia · The Guardian, 12.11 Guardian · The New York Times, 12.11 NYT · Reuters, 12.11 Reuters · WIRED, 12.11 Wired · @TheRundownAI, 12.11 @TheRundownAI · @naval, 12.11 @naval · HN, 12.11 Hacker News

Yesterday the WSJ wrote that the Nvidia board prepares to remove Huang. Today the board announced that Huang left the posts of CEO, president and board member. The decision is effective immediately. CFO Colette Kress became interim CEO.

She also stays CFO. For the first time in the company history, the board appointed a separate chair, Harvey Jones.

Jones is on the board since 1993 and earlier led Synopsys.

The press release says nothing about the scandal. The board thanks Huang for his "immeasurable contributions". It will look for a permanent CEO among internal and external candidates. Nvidia will publish its third-quarter report on 17.11, as planned. On social media, Trump proposed Tim Cook for the post. Semafor also names Johny Srouji from Apple and Lisa Su from AMD as candidates.

After the announcement, Reuters reported a third affair of Huang, again with an Nvidia employee. The agency does not give her name. At the open, the shares fell by 13-14%

(Guardian, NYT), and they closed about 4% lower. By Guardian data, the company lost more than $180 billion of value in one day and about $490 billion in one week. The NYT writes that the shares fell by about one third from the peak in early October. Bernstein and RBC lowered their price targets.

Inside the company, the reaction is different. WIRED saw an internal post that asks the board to consider the opinion of employees. The post has more than 1,600 comments, and most of them support the return of Huang. On X, some investors also disagree with the board. Naval recalls the stories of Steve Jobs and Travis Kalanick. These are opinions, and they contain no new facts.

Why it matters: Nvidia does more than sell chips. It finances its customers, it put more than $50 billion into AI startups, and it guarantees data center leases for OpenAI and others. Many of these deals relied on the personal weight of Huang. A new CEO will probably have less freedom from the board. Thus the cloud companies that lived on Nvidia money can be the first to suffer. On 12.11, Nebius shares fell by almost 17%, and CoreWeave shares by almost 7%.

TOPIC 2

Washington after Huang: the DOJ examines Nvidia, Congress prepares export controls

SOURCES Semafor, 12.11 Semafor · Semafor, 12.11 Semafor · BBC, 13.11 BBC · @amasad, 12.11 @amasad

Semafor cites two officials and writes that the US Department of Justice runs several investigations. One is an antitrust investigation, and it is about Nvidia itself. Two others are about the companies Megaspeed and Supermicro and the role of Alice Huang. US intelligence suspects that she works for China. There is no sign yet that the DOJ wants to bring criminal charges against Nvidia or Huang. Attorney General Todd Blanche said that the department has no conclusion about espionage yet.

Brian Mast, chair of the House Foreign Affairs Committee, said that Huang agreed to a closed briefing. The briefing must occur before 20.11, when Congress goes on recess. A Senate staff member said that Huang can get a subpoena if he refuses other committees.

The main change is in export policy. By Semafor data, Nvidia sold chips for $4 billion to Megaspeed, and officials suspect this company of chip resale to China. Senate Majority Leader John Thune estimates new restrictions as a "high probability". Speaker Mike Johnson and Minority Leader Hakeem Jeffries also support them. Three bills can go into the defense budget law. The Chip Security Act requires a check of the physical location of chips after the sale. The MATCH Act closes holes in the rules for equipment. The AI Overwatch Act gives Congress oversight of licenses for chip sales to "countries of risk". Anthropic publicly supports all three, and OpenAI also leans to support.

Why it matters: Last year, lawmakers removed a similar bill, the GAIN AI Act, from the final defense law under pressure from Nvidia and the White House. Now the main lobbyist

against restrictions is gone. If the laws pass before the end of the year, the sale of powerful chips abroad will become similar to the sale of weapons. This will hit data centers in Singapore, Malaysia and the Gulf, where part of the supplies go now.

TOPIC 3

Researchers found a fingerprint of Blackwell chips in the weights of six Chinese models

SOURCES Unemployment Labs, 12.11 Unemploymentlabs · HN, 12.11 Hacker News

Researchers published a forensic analysis of open weights. Their conclusion: DeepSeek, Moonshot AI, MiniMax and Ant Group got access to Nvidia Blackwell chips. A sale of these chips to China requires a US license, and nobody reported such a license. The researchers found the fingerprint in six models released since April 2026: DeepSeek-V4-Flash, DeepSeek-V4-Pro, Kimi K3, Kimi K3.1-Pro, MiniMax-M3 and Ling-3.1-flash. Older models of the same labs show a Hopper fingerprint.

The mechanism is simple. Before training, PyTorch fills the weights with pseudorandom numbers from a "seed". The GPU divides this work between threads, and the number of threads depends on the number of processor units in the chip.

Hopper has 132 units, B200/B300 has 148, and GB200/GB300 has 152. Thus one seed puts the same numbers in different positions on different chips. Part of the weights stays almost unchanged after training. From this part, the researchers recover both the seed and the number of units.

The researchers tested the method on 24 models where the developer named the hardware. All 24 matched. For DeepSeek-V4-Flash and Kimi K3, they repeated the result on physical Blackwell chips. On B300, the DeepSeek seed reproduced 445 rows of weights bit for bit. Other labs, among them Alibaba (Qwen), Zhipu (GLM) and Tencent, show only Hopper.

The authors state the limits of the method directly. The fingerprint shows the hardware that created the initial weights. It does not prove that all training ran on the same hardware. It also does not show where the chips were.

Anyone can rent Blackwell in data centers outside China. Thus the method alone does not prove a violation of export controls.

Why it matters: Until now, the dispute about Chinese chips relied on leaks and statements of officials. Now there is a check that anyone with the open weights can repeat. The work came out on the day when Congress discusses export controls. Thus supporters of the law about chip location checks will use it as an argument.

TOPIC 4

Google released Gemini 4 Hydrogen: a fast model, stronger than Argon in the AA index

SOURCES Google, 12.11 Blog · Ars Technica, 12.11 Ars Technica · @GoogleDeepMind, 12.11 @GoogleDeepMind · @levie, 12.11 @levie · HN, 12.11 Hacker News

On 11.11, Google released the flagship Gemini 4 Carbon. Today the fast model of the same family came out, Gemini 4 Hydrogen. It replaces Gemini 3.8 Flash in all Google products. In the Artificial Analysis Intelligence Index, Hydrogen got 51 at the high level and 54 at the new xhigh level. The older model Argon has 50. The price did not change: $1.50 per million input tokens and $7.50 per million output tokens, for the full window of 1 million tokens. Carbon and Argon cost $2/$10, but from next month they will cost $4/$20.

Google numbers: Hydrogen completes tasks 2.5 times faster than 3.8 Flash. On Terminal-Bench 4.1, it has 57% against 18.2% for the predecessor. By the Google estimate, it makes 5 times fewer invented claims. Aaron Levie from Box writes that accuracy on their test increased by 5 percentage points.

Together with the model, Google added Orchestrator to Antigravity. It switches models in one session without help:

Carbon plans, and Hydrogen executes. On Terminal-Bench 4.0, this pair almost reached the result of Carbon, by Google data. It also used 24% less quota and worked 19% faster.

Why it matters: Google sells the cheap model at the price of its predecessor, and the flagships become two times more expensive. The market logic changes: teams call the expensive model only for planning, and the fast model does the main work. All numbers so far come from Google and its early access partners.

TOPIC 5

Anthropic opened its IPO site: $17 billion of revenue per quarter and a $23.3 billion loss

SOURCES Anthropic, 12.11 Anthropic · @AnthropicAI, 12.11 @AnthropicAI · HN, 12.11 Hacker News

On 11.11 we reported $17 billion of Anthropic revenue for the third quarter from one source, the WSJ behind a paywall.

Now the company published this number itself on the IPO site. The site has a founders letter, a roadshow video and the prospectus. Quarterly revenue grew more than 12 times in one year. More than 11,900 customers pay more than $100,000 per year.

The net loss for the quarter is $23.3 billion. The company explains $21.9 billion of it with a non-cash revaluation of convertible notes and SAFEs, because its own valuation grew. The operating margin is minus 3%. The adjusted operating margin is plus 15%. Compute efficiency, which is revenue per dollar of infrastructure cost, grew more than 4 times since the first quarter of 2025, by company data.

Why it matters: Until 18.11, when the share price is set, investors will see the difference between two numbers.

Minus 3% comes from standard accounting rules, and plus 15% comes from the company's own calculation. The large loss in the report is mostly technical, but the headlines will quote it.

TOPIC 6

Global Microsoft 365 outage: Outlook, Teams and Copilot were down for hours

SOURCES BBC, 13.11 BBC · The New York Times, 12.11 NYT · HN, 12.11 Hacker News

On 12.11, in the afternoon US East Coast time, Outlook, Teams, Copilot and Microsoft Store went down. By Downdetector data, complaints about Outlook reached almost 140,000 at about 18:00 GMT. Microsoft explained that part of its servers became unavailable without warning, and the rest of the infrastructure could not handle the load. At 17:00 New York time, the company wrote that previous actions "have not provided the anticipated relief". The BBC writes that there are no signs of a cyberattack.

The outage also hit AI companies: OpenAI investigated problems with connections to Microsoft services. The NYT gives effects outside IT. The sheriff of Arapahoe County in Colorado could not accept bail for detainees, because email did not work. Unity Health hospitals in Toronto asked patients to call by phone.

Why it matters: This is the second large Microsoft outage in a month, after the Azure outage in October. Agents that work with email and calendar through Microsoft 365 break together with it. Teams that build automation on one cloud provider must plan for a scenario where that provider is down for half a day.

TOPIC 7

Critical RCE in Apache HTTP Server, found by Anthropic and Google

SOURCES Wiz, 11.11 Wiz · HN, 12.11 Hacker News

The vulnerability CVE-2026-119618 is in the HTTP/2 module (mod_http2) of Apache HTTP Server. It is a race condition in the function assign_headers. A remote attacker without authentication can corrupt heap memory and execute own code with the rights of the web server. Versions 2.4.55 to 2.4.70 are vulnerable, and 2.4.71 has the fix. Anthropic and the Google Cloud vulnerability research team found the bug independently.

By Wiz data, 64% of cloud environments have vulnerable resources, and in 51% these resources are open to the internet.

Wiz itself notes that it did not take configuration into account. At the time of publication, there was no public exploit. If an immediate update is not possible, Wiz advises to disable

HTTP/2 and keep only HTTP/1.1.

Why it matters: On 11.11, Anthropic reported 37,966 findings of the Glasswing program and 217 fixes. Now its researchers, together with Google, found a critical CVE in a server that runs a large part of the web. Wiz does not say if this is a Glasswing finding. Race conditions are among the bugs that are the most difficult to find manually.

TOPIC 8

Thinking Machines released the weights of Inkling 2 Small under Apache 2.0

SOURCES Hugging Face, 12.11 Hugging Face · @thinkymachines, 12.11 @thinkymachines

On 06.11, Thinking Machines released Inkling 2 Small only through its Tinker platform and promised to open the weights in a week. The company kept the promise: the weights are on Hugging Face under Apache 2.0. It is an MoE with 594 billion parameters, and 17 billion of them are active. Each token goes to 8 of 384 experts. The model accepts text and images, with a window of up to 1 million tokens.

The company trained the model directly in the NVFP4 format, and the main version of the weights is also NVFP4. It requires a cluster with at least 500 GB of video memory. The MXFP8 version requires 800 GB and, by company data, gives only 1-2 points. In the company's own table, Inkling has 30.6 in AA Index v5.0. This is higher than GLM-5.3-Flash (27.8), but lower than DeepSeek-V4.1-Flash (32.6) and Claude Haiku 5.5 (36.5).

Why it matters: An American lab released an open model of this size with a permissive license. Until now, most strong open weights came from China. The model is too large to run at home, but a company can keep it in its own data center without license conditions.

TOPIC 9

OpenAI launched Daybreak Defense Factory and Daybreak Red Team

SOURCES OpenAI, 12.11 OpenAI · OpenAI, 12.11 OpenAI · @gdb, 12.11 @gdb

Greg Brockman announced two cybersecurity products. Daybreak Defense Factory accepts findings from scanners and the red team. Agents trace each vulnerability in the code and prepare fixes. Engineers review and test the fixes before the merge. The service also scans new commits on a schedule.

Daybreak Red Team starts a group of agents against web applications. The agents read the code, look for attack paths and confirm vulnerabilities. Then they check if the fix really stops the exploit. An isolated environment and control of the addresses that the agents can reach set the limits of the test.

Why it matters: Item 7 shows that models find critical bugs faster than people can fix them. OpenAI sells the closure of this cycle: finding, fix, check. The product pages give no prices and no results on real customers.

TOPIC 10

Agents with a wallet: Replit buys for business through Link and Visa

SOURCES @Replit, 12.11 @Replit · @link, 12.11 @link · @amasad, 12.11 @amasad · WSJ, 13.11 WSJ

The Replit agent now buys what the business of the user needs. Payment goes through the Link agent wallet and Visa, and the agent runs ads in Google Ads. An example from Link: "find coffee pods for the office, under $35". The agent looks for options and asks for confirmation before payment.

The WSJ writes that such agents, among them Muse from Meta, got the right to buy on behalf of the card owner in recent weeks. Stripe builds its business on this, and JPMorgan Chase is skeptical. Amazon blocked Muse from purchases on its site. We saw only the first paragraph of the WSJ article, and the rest is behind a paywall.

Why it matters: On 12.11 we reported that United Airlines banned AI agents from ticket purchases. Now two strategies are visible. Payment companies give agents a card, and sellers close the door until they understand who is responsible for a wrong purchase.

---

misc

- **Humble Bundle after the redesign:** users of the SpaceBattles forum and Reddit write that the new Humble Bundle site disabled two-factor authentication. By their words, anyone can change the password and email without confirmation.

  The authors blame vibe coding. Humble Bundle did not comment at the time of the posts, and we did not check this ourselves. Spacebattles · HN Hacker News
- **Rust 1.100:** after 10 years and 5 failed attempts, the team stabilized the never type (`!`). The main parts of allocator_api and the new `core::mem::DropGuard` are also stable. Rust-lang · HN Hacker News
- **Primer from Wharton:** the lab of Ethan Mollick opened a system where a teacher describes a lesson in a conversation with agents and gets a simulation for students. The system builds nothing until the teacher approves the plan. The lab published the code, the prompts and the paper openly. Upenn · @emollick
- **Mirendil, $1 billion Series A:** the startup of Behnam Neyshabur, which builds AI for scientists, raised $1 billion led by Kleiner Perkins and Clearlake. @eladgil
- **HubSpot:** co-founder Brian Halligan returned to the CEO post 20 years after the company start. The company now calls itself the "Customer Harness Company". @dharmesh

- **Retell:** the voice agent startup has $100 million of annual revenue. The path from $10 million to $100 million took 17 months, and the company raised only $4.6 million. @_bingwu
- **Lockfile for GitHub Actions:** Hugo van Kemenade shows a new lockfile for Actions, `.github/workflows/actions.lock`, and a CI check that it is up to date. Earlier advice was to pin Actions to a SHA.
  Hugovk · HN Hacker News
