AI DIGEST

OpenAI found 948 paths to root in Linux for $1.2M

US voters blocked data centers, Meta and OpenAI bought agent tools, Kimi K3.1 Pro weights

Thursday, 5 November 2026

IN THIS ISSUE

1. OpenAI: agents found 948 paths to root in the Linux kernel for $1.2 million
2. US voters blocked data centers in almost all ballot measures
3. Deal day: Meta, OpenAI and Factory bought tools for work with agents
4. IEEE Spectrum: AI forces us to pay 40 years of technical debt
5. Open weights: Kimi K3.1 Pro, the American Ripple-65B and Tencent Hy4
6. Gemini trains on chats, even of paid subscribers
7. Anthropic: Claude helped to find the first structure of a bacterial vault particle
8. Layer3: residential proxies exit through the White House network
9. RTX Spark laptops on Linux give five times fewer tokens than on Windows
10. OpenAI released Images 2.6: GIF, SVG and generation in 6 seconds

TOPIC 1

OpenAI: agents found 948 paths to root in the Linux kernel for $1.2 million

SOURCES OpenAI, 04.11 OpenAI · @OpenAI, 04.11 @OpenAI · @gdb, 04.11 @gdb · HN, 04.11 Hacker News

The OpenAI security team built a pipeline with five stages. Scout agents select kernel code that a normal user can reach.

Analyst agents look for bugs in it. Other agents reproduce each finding in a virtual machine, then write an exploit and a patch. At the last stage, a person checks the result.

In mid-October, one run on the new cyber model gave these numbers. The scouts selected 5,034 targets, or 2.8 million lines and ~7% of the kernel. The analysts found 2,308 candidates. The agents reproduced 1,724 vulnerabilities, which is ~75%.

Wrong analysis by the model was only 3.1%. The output was 948 unique vulnerabilities with a working root exploit. The targets ran on a normal desktop distribution with SMEP, SMAP and KASLR.

The full run cost $1,213,430 in API fees, or $1,279 per exploit. @gdb rounds it to "about $1,000". For comparison, OpenAI uses old Pwn2Own and TyphoonPWN rewards: $20,000

and $70,000 for such an exploit. In July, GPT-5.6 Sol on the same pipeline found 166 vulnerabilities. The new model found 5 times more, although developers patched the kernel for three months.

Two observations are useful outside Linux. First, in an early run, 20 of 200 "successful" exploits got root through a different vulnerability than the intended one. Thus OpenAI now checks the path to the result as well as the result. Second, a heavy harness almost does not help a stronger model. A set of 103 techniques increased the success of GPT-5.5 Cyber from 76.2% to 88.9%. For GPT-5.6 Cyber, the same set gave only +1.5 p.p.

OpenAI writes that the agents used at least 10 new kernel exploitation techniques. It does not publish the details for security reasons. Hundreds of reports and patches already went to the maintainers. After 10 days of scans, 61 vulnerabilities already had a patch in upstream, and 43 had patches in staging.

Why it matters: OpenAI says it directly: access to an unpatched system with an account without privileges now equals access to root. This gives a practical conclusion for any infrastructure. Kernel updates become urgent, and sandboxes such as gVisor and Firecracker become a mandatory layer. In the HN thread, the main question is this: is the kernel simply too large to be free of a thousand such bugs?

TOPIC 2

US voters blocked data centers in almost all ballot measures

SOURCES The New York Times, 04.11 NYT · Ars Technica, 04.11 Ars Technica · Semafor, 04.11 Semafor · Stratechery, 04.11 Stratechery · The Wall Street Journal, 04.11 WSJ · @AndrewCurran_, 04.11 @AndrewCurran_ · HN, 04.11 Hacker News

Yesterday this digest mentioned that Democrats took the House. Today the role of data centers in this result became clear. The NYT counts that in 40 of 41 local votes, voters limited or banned data centers. More than half of them were in Ohio. Ars cites Redbud Consulting and writes about 38 approved measures of 44 in 10 states. The numbers differ because the count continues. But each source names one exception: the city of Harwood in North Dakota.

At the state level, the picture is the same. Democrat Gina Hinojosa became governor of Texas, and she promised a moratorium on data centers. According to Ars, Texas was to give more than 41% of new data center capacity in 2028.

Democrats also won the governor races in Arizona, Georgia and Ohio. In Ohio, senator-elect Sherrod Brown made $2.5 billion of tax breaks for 226 data centers his main argument.

The industry answers in different ways. Meta president Dina Powell McCormick said at a WSJ conference that construction will become "harder". The industry group Chamber of Progress calls it an "anti-Trump wave" and says it is no mandate against technology. NBC polls show that 3 of 4 voters oppose data centers. More than half said that AI and data centers were an important factor in their vote.

Ben Thompson in Stratechery states the result like this. The limit for AI moved from chips, memory and energy to the permission to build. Hyperscalers and companies that already have permits win from this. Neoclouds that need new sites lose.

Why it matters: before this week, the main risks of AI construction were credit and energy. Now local votes are also a risk, and banks already count them as credit risk. For the inference market, this means a longer road to cheaper tokens in the US. For the rest of the world, this is a chance: a Natixis economist says that China gets an opportunity to sell cheap inference.

TOPIC 3

Deal day: Meta, OpenAI and Factory bought tools for work with agents

SOURCES Conductor, 04.11 Conductor · @charlieholtz, 04.11 @charlieholtz · OpenAI, 04.11 OpenAI · @gdb, 04.11 @gdb · @dexhorthy, 04.11 @dexhorthy · @bentossell, 04.11 @bentossell · HN, 04.11 Hacker News · HN, 04.11 Hacker News

Three deals closed in one day. @bentossell called it "holy acquisition day". All three companies made tools for coding agents, and all three disappear as separate products.

Conductor, an app for parallel work of several agents, moves to Meta Superintelligence Labs. Conductor no longer accepts new users. Conductor Cloud becomes read-only on 13.11 and closes on 20.11. The local version will get models and critical fixes until early 2027.

OpenAI buys Pierre Computer Company. Its libraries Trees and Diffs are already in the Codex app for file navigation and change review. Pierre also makes Code Storage, a service that stores code through Git. OpenAI connects the deal to Pando, its Git backend, which this digest covered on 03.11. The open projects of Pierre stay open.

HumanLayer moves to Factory. The founder Dex Horthy writes that large teams need a product that does not depend on a specific lab. HumanLayer will open its code and keep the platform for at least 30 days. On the same evening, Cat Wu announced that she moves to OpenAI to lead the product Core Products. Before that, she led the Claude Code product at Anthropic.

Why it matters: the labs take the layer above the model for themselves: change review, agent orchestration, code storage. This layer has fewer independent tools now. For teams that build their process on such tools, this is a risk:

Conductor gave cloud users only 9 days before read-only. Keep a data export and a migration plan ready in advance.

TOPIC 4

IEEE Spectrum: AI forces us to pay 40 years of technical debt

SOURCES IEEE Spectrum, 04.11 Ieee · HN, 04.11 Hacker News

This is the most popular HN article of the day: 1,113 points and 648 comments. The occasion is familiar: PHP2Shell, which this digest covered on 22.10. Somebody reported the out-of-memory bug in PHP in 2024. People considered it safe, because the program simply crashed. An auditor from Calif, together with AI, turned it into arbitrary code execution on the server in two weeks.

Columbia University professor Suman Jana explains what changed. An exploit "is not rocket science", but it needs a lot of boring and exact reasoning, and AI works well here. Jana does not say that AI does everything by itself. He says that the top 0.1% of experts now do more, and people slightly below them reach their level.

The main problem is asymmetry. A bug is now fast to find, but it is still slow to fix. Four engineers of the PHP Foundation wrote the patch for two months. Jana hopes for formal verification of patches: an LLM writes the patch and the proof, and a formal checker verifies them. Matthew Green from Johns Hopkins is less optimistic. He expects a really bad incident "maybe sometime in early 2027", and only after it he expects money for defense.

Why it matters: item 1 shows the same asymmetry in numbers. Findings become cheaper, but human time for a patch and a release does not. Old reports about a "simple crash" in the tracker of any project are now worth a new review as potential vulnerabilities.

TOPIC 5

Open weights: Kimi K3.1 Pro, the American Ripple-65B and Tencent Hy4

SOURCES Hugging Face, Moonshot AI, 04.11 Hugging Face · HN, 04.11 Hacker News · Ricursive Intelligence, 04.11 Ricursive · HN, 04.11 Hacker News · Tencent, 04.11 Tencent · HN, 04.11 Hacker News

On 24.10, Moonshot promised the Kimi K3.1 Pro weights by 04.11. It kept the promise. The model has 2.8 trillion parameters, and 104 billion of them are active. The context is 1 million tokens, and the weights are in MXFP4. In the Moonshot table, the model gets 65.7 on Terminal-Bench 4.0. Claude Fable 5.1 gets 57.9 there, GPT-6 Astra 58.2, Claude Opus 5.5 66.4. These are the numbers of Moonshot itself.

The license has a condition for providers. A service with revenue above $20 million must sign a separate agreement with Moonshot. In the HN thread, people argue if this is still "open weights". One commenter writes that his team runs the previous Kimi on a server with 4 AMD MI300A for ~$110,000.

Ricursive Intelligence released Ripple-65B Preview, a dense model with 65 billion parameters. It gets 80.52 on TerminalBench 2.1, and the company calls this a record among US open models. Ricursive says that the full cost of development from zero was $32 million. The model runs locally on NVIDIA DGX Spark. On the same day, Tencent released Hy4 with vision and computer use. Tencent writes that Hy4 reduced the hallucination rate from 15.76% to 5.45% in its internal tests.

Why it matters: open models already stand next to closed ones on agent benchmarks, although by the numbers of the authors themselves. Companies that must keep data in-house get more choice. But you must read the license: "open weights" now often means conditions for a commercial service.

TOPIC 6

Gemini trains on chats, even of paid subscribers

SOURCES HN, Tell HN, 04.11 Hacker News · Google, Gemini help Google

The author of an HN post thought that a paid subscription protects his chats from training. The Google help page says otherwise. For personal accounts, activity in the Gemini app goes to service improvement, "including by training generative AI models". This applies while the Keep Activity setting is on.

People read part of the chats. Such chats stay when you delete the activity, and they remain for up to 3 years. You can turn off training only together with Keep Activity. Then the chats disappear after 72 hours. The thread got 635 points.

Commenters remind that all large providers have human reviewers.

Why it matters: yesterday the misc section had a case where the training switch in Codex turned itself on. Today the story is wider: payment by itself does not turn off training on your data. Work information must go through corporate accounts with separate terms. Check the privacy settings in each product separately.

TOPIC 7

Anthropic: Claude helped to find the first structure of a bacterial vault particle

SOURCES Anthropic, 04.11 Anthropic · @AnthropicAI, 04.11 @AnthropicAI

Scientists discovered vault particles in 1986. These are large hollow protein barrels in animal and plant cells, and their function is still unknown. Anthropic biologists showed that similar proteins of bacteria and phages also assemble into vaults. They got the first 3D structure of such a particle: 96 copies of one phage protein L05, about 9.5 megadaltons.

Claude searched for relatives of the protein in a database of ~15 billion predicted proteins, designed DNA and processed the cryo-EM data. The purification took ~90 minutes instead of the usual three days. The team had the first model of the structure after 34 hours. It estimates that manual work would take weeks. The full project took less than four months.

Anthropic describes the errors separately. The Claude protocol had an additive that blocks the delivery of DNA into cells, and this cost two weeks. In the cryo-EM processing, Claude skipped a standard step and got a map mixed with its mirror image. The preprint went to bioRxiv, but at the morning of 05.11 no link to it exists yet.

Why it matters: a team that did not work with such large protein structures before did the work of structural biologists. The second half of the report is also valuable: AI errors cost the most where the person knows the subject poorly. Human review stays mandatory exactly there.

TOPIC 8

Layer3: residential proxies exit through the White House network

SOURCES Layer3 Intel, 04.11 Layer3intel · HN, 04.11 Hacker News

Residential proxies sell access to normal devices, so that traffic looks like traffic from their network. Layer3 Intel saw such exits from AS2650. This network belongs to the Executive Office of the President. From 5.10 to 3.11, Layer3 requests went out through it 236 times, from 6 addresses, on 26 days of 30. 28 different proxy services sold the access.

Layer3 compared the addresses with Spur and Synthient, and they also mark them as proxies. Half of the activity falls on 17:00-00:00 Eastern time. Layer3 thinks that this is an infected device of a specific person, and not guest Wi-Fi. The data does not show who it is or which device. The HN thread got 777 points [single source].

Why it matters: anybody can buy access to an exit from such a network for a few dollars. Badly protected proxy SDKs already gave access to internal networks, and the Kimwolf botnet spread this way. The conclusion for any company: check if your IPs appear in residential proxy databases.

TOPIC 9

RTX Spark laptops on Linux give five times fewer tokens than on Windows

SOURCES MindStudio, 04.11 Mindstudio · HN, 04.11 Hacker News

Laptops on the RTX Spark platform have the same GPU as the desktop DGX Spark, and up to 128 GB of shared memory. They ship with Windows on ARM. MindStudio installed Linux on an MSI Prestige and measured GPT-OSS-120B. Windows gave 65.4 tokens per second, WSL 65.1, DGX Spark 61, and Linux on the same laptop only 13.9.

The cause is in the firmware. On mains power, Linux keeps the GPU at ~780 MHz instead of ~2.2 GHz. The power table in the firmware is encrypted, and only a Windows driver component decrypts it. The temperature is 40-47 °C, thus no overheating occurs. Only Nvidia can give a solution, if it moves this logic into the Linux driver [single source].

Why it matters: vendors sell these laptops to AI developers, and their tools, vLLM and SGLang, target Linux first.

For now, the only practical path is WSL. Before you buy an "AI laptop" for Linux, look for measurements on Linux.

TOPIC 10

OpenAI released Images 2.6: GIF, SVG and generation in 6 seconds

SOURCES OpenAI, 04.11 OpenAI · @OpenAI, 04.11 @OpenAI · HN, 04.11 Hacker News

Images 2.6 appeared in ChatGPT, Codex and the API. OpenAI says that generation became 50% faster. The API has two models.

GPT-Image-2.6 Flare at low quality gives an image in 6 seconds. GPT-Image-2.6 Sunburst gives higher quality and keeps a character more consistent between edits.

New features: animated GIFs, examples from reference images, markup in the API and export to SVG. The complexity parameter sets how much the model plans the scene. According to OpenAI, users create 4 billion images per week.

Why it matters: SVG export and markup in the API make generation a part of the design workflow. It stops being a separate toy. For interfaces and marketing, this means less manual work between generation and the finished file.

misc

- **AMD vs Intel, continued:** AMD expects ~$17 billion of revenue in Q4. Tom's Hardware writes that this is more than the Intel forecast for the first time. Data center revenue in Q3 was $8.55 billion, +97% year over year.

  Tomshardware · Hacker News
- **Debian apt:** Max Justicz found a bug in APT with an HTTP redirect and resumed download. A man-in-the-middle attacker can replace a package after the check and execute code as root. Debian fixed this in apt 3.0.4, CVE-2026-109518.

  Debian · Hacker News
- **Fridges, continued:** hackers raised the temperature in three storage sites of the French blood service EFS. The alarm went off, and technicians acted before the products went bad.

  Guardian
- **Google and schools, continued:** Kagi writes that Google let schools enable search separately from Gemini, but schools still cannot remove AI Overviews from search.

  Kagi · Hacker News
- **Hinton:** Geoffrey Hinton told the Guardian that people would possibly be better off without AI. He expects a large incident, for example a shutdown of financial or energy systems, this year or next year.

  Guardian · Hacker News
- **Grok Bot makes calls:** the xAI bot now makes phone calls for the user in the US, except in the states of Washington and Illinois. It waits on the line, goes through voice menus and gives a transcript.

@TheRundownAI

- **AI pause:** Congresswoman Yassamin Ansari wants Democrats to put a pause on advanced AI development and a ban on superintelligence on the first day. @AiBreakfast answered sharply against it [single source].

  @AiBreakfast
- **AI semiotics:** @emollick promotes a paper that says visual symbols strongly affect models, and the model does not always see this effect in its chain of reasoning.

  @emollick · arXiv
- **New X web:** X engineer Yagiz Nizipli writes that on a slow connection the first profile post now appears in 2.7 s instead of 22 s. The release goes country by country.

  @yagiznizipli · X
- **Meta and Brazilian media:** a Brazilian court banned Meta from the collection of local media content for AI training.

  Semafor
