AI DIGEST

Mathematician: judge every paper as if AI wrote it

Red Hat patches the Linux kernel, Ubuntu not yet, Intel down 29%, Binance reopens, agents in Minecraft

Monday, 2 November 2026

IN THIS ISSUE

1. Calegari on the blog of Tao: judge each paper as if an AI wrote it
2. CVE-2026-98432: Red Hat fixed its kernels, Ubuntu still has no fix, and a disabled MPTCP does not protect
3. The market doubts AI spend: Intel down 29% in one day, neoclouds pay more for money
4. Binance opened after the hack: customers withdrew $3.64 billion in one day
5. What a JavaScript framework costs an agent: Nolan Lawson built the same app 60 times
6. Claude Code edited a vlog from 58 minutes of video and chose the tools itself
7. Mollick: GPT-6 modes are confusing, and the productivity line must be January 2026
8. 50 agents in Minecraft got the task to build a civilization and built a bureaucracy

TOPIC 1

Calegari on the blog of Tao: judge each paper as if an AI wrote it

SOURCES Terence Tao blog, guest post by Frank Calegari, 01.11 Wordpress · HN, 01.11 Hacker News

On 6 October, OpenAI published more than 700 mathematics papers. This digest wrote about it on 07.10. Now mathematician Frank Calegari tells colleagues how to live with this. His text is a guest post on the blog of Terence Tao. Tao added a note that AI helped to convert the text into another format.

Calegari admits that the OpenAI papers are "terribly written", break attribution norms, and that some of them are wrong. But he says that mathematicians have more power over hiring, journals and teaching than over the behavior of trillion-dollar companies. He quotes Hugo Duminil-Copin: each large open problem that Duminil-Copin mentioned in his career now has a claimed proof.

Calegari gives six proposals. Judge each work as if someone made it with AI. Hire after a personal conversation, also for postdocs and PhD students. Ask more of journals, and leave the journals that do not change. Teach PhD students more broadly, but keep their technical skills. Make better seminars. Take responsibility for AI results. The most radical idea: a result that a machine can produce must no longer give credit, even if the author made it by hand.

On HN, the thread got 244 points and 273 comments.

Why it matters: other professions do not have this discussion yet, and mathematics already has it. The mechanism is simple: if a result is cheap, the value moves to the choice of the problem and to the explanation of why it is interesting. All people whom others judge by the number of papers, reports or commits will feel the same pressure.

TOPIC 2

CVE-2026-98432: Red Hat fixed its kernels, Ubuntu still has no fix, and a disabled MPTCP does not protect

SOURCES Atlas, updated 01.11 Atlasum · Ubuntu tracker Ubuntu · @levelsio, 01.11 @levelsio

Yesterday this digest wrote that RHEL and Ubuntu had no fix. On 1 November, Red Hat released fixed kernels for RHEL 8, 9 and 10.

On the morning of 02.11, the Ubuntu tracker shows "Vulnerable" for all releases. For 26.10, 26.04, 24.04 and 22.04 it shows "work in progress". AlmaLinux keeps fixed kernels in a test repository.

Atlas added an important detail. Anthropic says that Claude wrote a working exploit that gives root on a server with an MPTCP listener. But the bug itself is in the usual IPv4 TCP code. The Go security team writes that systems with no MPTCP listeners can also be open to similar attacks. Thus, a disabled MPTCP removes only the known attack path. It does not remove the bug.

On the evening of 01.11, Atlas counted 41 GitHub repositories that mention this CVE. Anthropic did not publish its exploit.

@levelsio advises to close all ports on a VPS, to allow SSH only through Tailscale, and to send web traffic through Cloudflare Tunnel.

Why it matters: the vulnerable code runs before the application sees the connection. Thus, only a new kernel and a reboot protect a server. On Ubuntu, the available measures for now are an nftables rule against TCP packets with IPv4 options and a minimum of open ports.

TOPIC 3

The market doubts AI spend: Intel down 29% in one day, neoclouds pay more for money

SOURCES Semafor, 01.11 Semafor · BBC, 01.11 BBC · The Wall Street Journal, 01.11 WSJ

Yesterday this digest wrote about the credit squeeze on neoclouds. Now the stock market shows it too. Semafor writes that on Friday the Nasdaq entered a correction. Intel shares fell 29%, the largest one-day fall since the 1970s. The results of Amazon and Google disappointed investors. S&P Global changed its outlook for Meta to "negative" because of large capital spend.

The BBC explains where the weak point is. Neocloud companies, such as CoreWeave and Nebius, borrow money against long contracts with large customers and against the chips themselves. CoreWeave has $35.1 billion of debt. Its shares lost a quarter of their value in a month. Nebius, IREN and Applied Digital fell by the same amount or more. Gartner predicts that by 2030, 80% of neocloud companies will stop independent operation.

When Microsoft and Google gave an AI spend forecast below analyst expectations, neocloud shares fell. The WSJ put the topic in the headline of its newsletter, "AI Spenders Get Spooked". The text is behind a paywall.

Why it matters: all AI infrastructure rests on one assumption: large customers will continue to increase their spend. The first forecast below expectations hits the links with the most debt. For buyers of compute from small providers, this is a supplier risk, and the contract must take it into account.

TOPIC 4

Binance opened after the hack: customers withdrew $3.64 billion in one day

SOURCES BBC, 02.11 BBC · Financial Times, 01.11 FT · Semafor, 01.11 Semafor

Yesterday this digest wrote that the FBI confirmed the North Korea trail. On Sunday, Binance started operations again. The exchange says that from 06:00 to 15:30 GMT, customers withdrew $4.649 billion and deposited $1.01 billion. The net outflow was $3.64 billion. This is a little more than 2% of $171 billion of assets under management.

Binance says that it already covered about 95% of the stolen funds. The rest can take more time. Early checks by the exchange show that the attack started from the hacked laptop of one engineer. The FBI published a list of TraderTraitor group wallets and asks the industry to block transactions with them.

Why it matters: the exchange itself says that the entry point of the largest crypto hack in history was a usual work laptop.

An outflow of 2% in one day shows no panic. But trust now costs the exchange billions in reserves.

TOPIC 5

What a JavaScript framework costs an agent: Nolan Lawson built the same app 60 times

SOURCES Nolan Lawson, 28.10 Nolanlawson · HN, 01.11 Hacker News

[single source] Developer Nolan Lawson tested the claim that some framework is "better for agents". He gave Opus 5.5 the same specification of a Pokemon app and asked it to build the

app on 12 frameworks. On each framework, the agent built the app 5 times, and Lawson took the median.

Hono+htmx was the cheapest: $2.67 per build at API prices. Angular was the most expensive: $4.97, and two Angular runs gave no working app at all. SvelteKit cost $4.28, because the agent read the v3 migration guide each time. Lawson explains that the new version is not yet in the training data of the model.

Lighthouse scores were almost equal: performance 97-100 and accessibility 100 for all. The agents often ignored the official MCP servers and skills. The agent found the TanStack Start skill in only 2 of 5 runs, and it never used the Nuxt MCP.

Why it matters: the question "which stack is better for agents" now has a measured answer, and a measurement costs a few tens of dollars. The first conclusion: a framework that the model does not know costs more, because the agent learns it during the work.

TOPIC 6

Claude Code edited a vlog from 58 minutes of video and chose the tools itself

SOURCES @levelsio, 01.11 @levelsio · @levelsio, 01.11 @levelsio

[single source] Pieter Levels copied clips from a DJI camera into a folder and asked Claude Code to edit a 10-minute vlog. The input was 58 minutes of video from 14 clips. He learned about the tools only when someone asked him how he edited the vlog.

The agent transcribed speech with mlx-whisper on the MacBook GPU. It made frame grids every few seconds to see the content of the clips. It separated the voice of Levels from other voices with the SpeechBrain ECAPA model.

The edit itself became a JSON list of cuts: clip, start, end. The agent chose the fragments from the transcripts and the frame grids.

Levels says that without AI he would not make a vlog, because the edit would take too much time.

Why it matters: the agent divided an unfamiliar task into usual local tools and asked nobody about each step. For routine media edits, the barrier is now the wording of the request.

TOPIC 7

Mollick: GPT-6 modes are confusing, and the productivity line must be January 2026

SOURCES @emollick, 01.11 @emollick · @emollick, 01.11 @emollick · @emollick, 01.11 @emollick

[single source] Ethan Mollick praises GPT-6 Pro for hard one-time tasks with calculations and references. He says that the model is slow, but it gives the feeling of "asking an oracle".

Pro is available only in the chat interface. In Codex and ChatGPT Work, the highest mode is Ultra, and it does a different thing: it starts many agents. There is also Max Thinking, which is also weaker than Pro.

Separately, Mollick writes about productivity. GPT-4 had a 50% chance to complete 4 minutes of independent work, and current models complete work of several weeks. Thus, he proposes to treat January 2026 as the start of agent work. He says that before this date, the large effect was mostly in code.

Why it matters: studies that find no AI effect in the economy often measure the chatbot period. The line of Mollick gives a simple check for each such report: which data after January 2026 does it use?

TOPIC 8

50 agents in Minecraft got the task to build a civilization and built a bureaucracy

SOURCES YC Labs, 29.10 Ycombinator · Tom's Hardware, 01.11 Tomshardware · HN, 01.11 Hacker News

YC Labs and Olam Labs put 50 GPT-6.1 Sol agents into a Minecraft world for 3 hours with the goal to "build a thriving civilization".

The agents had a directory of participants, a shared sandbox and private messages. They sent about 5,900 messages.

The first night killed many: 22 of the 34 deaths in the full run. Death had no penalty except the loss of items, but the agents still became careful. They stayed in shelters, shared food by a log, and wrote rules and permissions. Tom's Hardware quotes a case where four agents spent 11 minutes to appoint a "sole fertilizer operator" for one block of wheat.

In the game Mindustry, GPT-6.1 Sol agents beat Claude Opus 5.5 agents 6:1 in 7 games. The authors released the environments and the Minecraft harness as open source.

Why it matters: the authors say that the behavior reminded them of work with coding agents. A large swarm with no cost of error becomes careful and spends resources on coordination. This is a useful reference point for teams that scale agents by number.

---

misc

- **HackerBooks:** after the HackerTube of Karpathy, which this digest covered yesterday, the author did the same for books. The site collected 15,362 books from 270,334 mentions on HN, and the thread got 1,044 points.

  Hackerbooks · Hacker News
- **Flock:** the city of Plymouth cancelled its Flock contract on 22.09, but the company did not remove the cameras. The police chief ordered officers to pull the power cords.

  Thedrive · Hacker News

- **Brazil:** a court fined two lawyers about $16,000 for a prompt for the court AI, hidden in white text. The decision is from July, and it appeared on HN only now.

  Jetstream · Hacker News
- **Proton:** developer Raed writes that the 404 Media story about Proton is about its management, and AI has no part in it. The CEO called a request for a few days of pentest "loser thinking".

  Raed · Hacker News
- **Monorepo:** @tobi recalled how Shopify moved to a monorepo. Shopify engineer Burke Libbey writes that for agents this is a large advantage, and he would not go back.

  @tobi
- **Jepsen:** Simon Willison quoted Kyle Kingsbury: LLMs damage trust in documentation and software, because the signs of a mature system are now easy to fake.

  Simon Willison
- **Britain:** Openreach switches off the old PSTN telephone network on 31.01.2027 with no extensions. Alarms, payment terminals and emergency buttons for old people are at risk.

  Theregister · Hacker News
- **Teesside:** National Grid buys 29.5 hectares for the largest substation in Britain, near a planned AI data center.

  Guardian
- **Subscriptions:** @amasad warns owners of subsidized AI subscriptions about a sharp price increase. The reason: Elon Musk explained why Grok Bot users reach their limit faster.

  @amasad
