

A 2015 neural network beats GPT-6 Astra and Fable 5.1 on skin cancer

Token prices fall up to 80%, Qwen3.8-27B matches Fable 5, a Samsung TV listens to the home, 11 Tailscale flaws

Monday, 26 October 2026

IN THIS ISSUE

1. An old network with 20 minutes of fine-tuning beat GPT-6 Astra and Fable 5.1 in skin cancer diagnosis
2. OpenAI and Anthropic tokens became up to 80% cheaper in four months. Max Woolf asks why
3. Five copies of Qwen3.8-27B with a shared plan file matched Fable 5 on hard tasks
4. A teardown of a Samsung TV: a microphone through Bluetooth and AlexNet every few seconds
5. Tailworm: 11 findings in Tailscale, but the attacker first needs an entry point
6. Noah Smith: "the internet breaks". Yesterday's Saxe thinks differently
7. The heirs of Elvis sue Anthropic over 274 songs
8. AI errors at Deloitte again: a university merger report with a \$260 million mistake
9. "Build more ambitious things": Dan Luu and Ethan Mollick about the same idea

TOPIC 1

An old network with 20 minutes of fine-tuning beat GPT-6 Astra and Fable 5.1 in skin cancer diagnosis

SOURCES [Lynch Lab, 25.10](#) [Lynchlab](#) · [HN, 25.10](#) [Hacker News](#)

[single source] Magnus Lynch, a dermatologist and surgeon at Guy's hospital in London, took 200 skin lesions from the public set MILK10k. A biopsy confirmed the diagnosis of each lesion. A doctor must not dismiss half of them without treatment: 30 melanomas, 50 carcinomas and 20 actinic keratoses. The other half is safe: moles and seborrheic keratoses. Each lesion has two images: a normal photo and a dermatoscope image.

GPT-6 Astra, Claude Fable 5.1 and Inception v3 got these images. Inception v3 is a convolutional network from Google from 2015. Lynch fine-tuned it on 68 011 images from the ISIC archive on one H100.

He removed the 200 test lesions from the training data.

Result: the network named the correct diagnosis out of six in 125 cases (62%). GPT-6 Astra gave 84 (42%), and Fable 5.1 gave 85 (42%). The largest gap is on safe lesions. The network recognized 37 of 50 moles, GPT-6 Astra 16, Fable 5.1 33. The network recognized 26 of 50 seborrheic keratoses, GPT-6 Astra 13, Fable 5.1 only 6.

Each model had a favorite answer. GPT-6 Astra called 75 lesions basal cell carcinoma, but the set had only 25 of them. The 75 lesions included 14 melanomas.

Safety gives a different picture. The models dismissed almost no dangerous lesion: they caught 97 and 96 of 100. But they also dismissed only 32 and 22 safe lesions. At its own threshold, the network dismissed 76 safe lesions and missed 18 melanomas. After a threshold adjustment, it catches 98-99 dangerous lesions. Lynch lists the limits himself: the sample is small, 30 melanomas give a wide error, and each model ran once.

Why it matters. a general model that "sees images" is still no specialist. A network with 20 minutes of fine-tuning wins where a good labeled set exists. For a medical or industrial task with a narrow classifier, start with a comparison against a small specialized model. The frontier model is careful here, but this care gives little benefit.

TOPIC 2

OpenAI and Anthropic tokens became up to 80% cheaper in four months. Max Woolf asks why

SOURCES Max Woolf, 25.10 [Minimaxir](#) · HN, 25.10 [Hacker News](#)

Max Woolf, a former data scientist at BuzzFeed, collected a price timeline since June. On 30.07 OpenAI cut the price of GPT-5.6 Luna by 80%. On 01.09 Anthropic released Fable 5.1 with a 75% cheaper cache read. On 22.09 OpenAI released GPT-6 Sol at \$2/\$10, and Anthropic released Opus 5.5 at \$4/\$20 per million tokens. Anthropic then wrote directly that Opus 5.5 needs less compute. On 29.09 GPT-6.1 Sol cut the cache read price by half.

Woolf explains why the cache matters more than the input price. In a coding agent, cache reads make 96-99% of all tokens. Thus a 75% cache discount gives almost the same discount on all agent work.

He sees two causes. First, models became a commodity. A user changes the model with one switch in OpenRouter, and 29 providers already serve DeepSeek V4.1 Flash 1017 there. The second cause is less obvious. Both companies prepare for an IPO, and a discount on the main revenue looks strange before a prospectus. Woolf accepts that the inference margin is really large, and this margin permits the discounts.

A separate detail concerns subscriptions. Together with GPT-6.1 Sol, OpenAI cut the limit of the \$200 Codex plan by 50%. The official reason: "models became more efficient".

Why it matters. the thesis "tokens are cheap only for a time, then prices will rise" does not come true yet. Competition from open weights and dozens of providers keeps the price low. For an agent budget, the main number is the cache read price. Check it first.

TOPIC 3

Five copies of Qwen3.8-27B with a shared plan file matched Fable 5 on hard tasks

SOURCES GVS5H on GitHub [GitHub](#) · HN, 25.10 [Hacker News](#)

[single source] The repository authors call the method "self-orchestration through a journal". Fresh copies of one model divide a task into parts. They coordinate through a shared folder with a plan, notes and the current solution. The method needs no additional training.

Test: the 100 newest hard tasks of LiveCodeBench, eleven models. Qwen3.8-27B with open weights on a local server went from 66,8% to 92,4% pass@1. Fable 5 has 90,4% on the same tasks. Qwen3.8 Flash Next with orchestration gave 93,0% at 9% of the Fable 5 cost.

The gain reaches 25,6 percentage points. The authors write honestly that the gain is not universal.

Some models did not change or became worse. According to their analysis, the task split and the memory between steps give the gain.

Why it matters. this is a preprint on one benchmark, and olympiad tasks differ from work in a real repository. But the direction is clear: the organization of agent work adds as much as a step between model generations. A model on one GPU with a good scaffold becomes a real alternative for closed environments.

TOPIC 4

A teardown of a Samsung TV: a microphone through Bluetooth and AlexNet every few seconds

SOURCES @mike64_t, 25.10 @mike64_t · @freddier, 24.10 @freddier · @Tristan0x, 25.10 @Tristan0x · HN, 25.10 [Hacker News](#) · HN, 25.10 [Hacker News](#)

Yesterday this digest skipped this topic: researcher @log4j got root on his Samsung TV already on 21.10. Over the weekend, other people checked his findings, and the topic got 287 and 136 points on HN. Andrej Karpathy reposted it.

According to @log4j, the TV classifies the sound in the room before it reports to Samsung. The telemetry has event counters for "dog barks" and "baby cries". If the TV has no microphone of its own, it looks for one in a soundbar or another Bluetooth device.

Mike (@mike64_t) examined the second part. Every few seconds, the TV runs the model `alex.tflite` on 960x540 frames with INT8 quantization. This is AlexNet, and it tries to guess which game you play.

According to him, it fails in 99% of cases, but it still loads the processor.

Tristan (@Tristan0x) wrote instructions. The main function here is ACR, automatic content recognition of the screen. It also works on the HDMI signal, for example from a laptop or a console.

Step one: disable "Viewing Information Services" on Samsung or "Live Plus" on LG. Step two: move the TV to a separate guest network. Step three: filter trackers at the DNS level.

Why it matters. a cheap TV pays for itself with viewer data, and now the firmware code shows it.

A firmware teardown took weeks before, and with agents it takes a weekend. More such reports about home appliances will come. A separate network for smart devices is no longer advice for paranoid people.

TOPIC 5

Tailworm: 11 findings in Tailscale, but the attacker first needs an entry point

SOURCES Tailworm (Cyberdione Labs), 25.10 [Tailworm](#) · @levelsio, 25.10 [@levelsio](#) · Tailscale Security Bulletins [Tailscale](#)

Erica Windisch from Cyberdione Labs published an audit of Tailscale v1.104.2 and headscale v0.29.5.

The tests ran on 22-24.10. The report has 11 findings, and six of them have high severity. The main class concerns Tailscale SSH. The built-in SSH server runs as root, and unix socket forwarding does its checks as root. A normal session can replace or open a privileged socket.

According to the report, in the lab the worm moved from node to node in less than one second.

A community note on X set the limit of the attack. The attacker needs a device that the network already approved and an SSH rule that permits access. Thus this is movement inside the network after a breach. It gives the attacker no entry from outside. Jamieson O'Reilly wrote that none of the 11 findings applies to a standard Tailscale installation.

The report authors did not wait for fixes on purpose. They think that networks are vulnerable already now. On the morning of 26.10, the last entry on the Tailscale security bulletins page is TS-2026-013 from 13.10. The page says nothing about this report. The report recommends three actions: SSH rules only for specific device pairs, disabled forwarding and key expiry for nodes on servers.

Why it matters. a mesh network that trusts each device turns one breach into access to the full fleet. After the week of PHP2Shell and Brotli, attackers have enough entry points. A check of your SSH rules in the Tailscale policy takes five minutes, and it needs no bulletin.

TOPIC 6

Noah Smith: "the internet breaks". Yesterday's Saxe thinks differently

SOURCES Noahpinion, 25.10 [Noahpinion](#) · @levie, 25.10 [@levie](#)

Yesterday this digest covered Joshua Saxe. He thinks that people, money and the risk of arrest limit the losses from cybercrime. For him, the number of found vulnerabilities is secondary. On 25.10, economist Noah Smith wrote the opposite. His main thesis: open models with guardrails that are easy to remove took away the knowledge barrier.

As an example, Smith gives a Unit 42 report. A low-skill hacker with DeepSeek V4 without guardrails hacked a public university in India. He stole more than 5 200 records of students and teachers. The request to the bot was at the level of "find what they cheated on there".

Smith quotes the NSA warning about PHP from 21.10. He also quotes Jason Clinton from Anthropic, who already in August advised to isolate critical networks. The conclusion of Smith: the internet as we knew it can stop working for some time. Most of the post is behind a paywall.

Aaron Levie from Box gives a third position. Most large companies do not update libraries fast enough. They do not know the real state of their risks, because their data is scattered. But swarms of agents can already check attacks and read logs, thus the defense also has a chance.

Why it matters. Saxe and Smith argue about the scale, and both agree on the facts. Attacks became more frequent and cheaper, and the window for preparation is short. Levie gives the most practical conclusion: first an inventory of data and access, then agents.

TOPIC 7

The heirs of Elvis sue Anthropic over 274 songs

SOURCES The Guardian, 25.10 [Guardian](#)

[single source] Authentic Brands Group owns the rights to the music and the image of Elvis Presley.

The company filed a suit against Anthropic on Friday, and Billboard reported it first. According to the suit, Anthropic trained Claude on 274 Presley songs, among them "Can't Help Falling in Love". The plaintiffs want \$150 000 for each track, approximately \$41 million in total.

The suit says that Claude reproduces protected content of "practically every song from every catalog". According to the suit, the limits on song output only prove that the company knew about the problem. According to the Guardian, this is one of the first large suits from the heirs of an artist, without a label.

Why it matters. music suits against AI companies move from labels to separate catalogs. Each rights owner can now calculate its own \$150 000 per track. For labs, the price of

training data grows even without a court decision.

TOPIC 8

AI errors at Deloitte again: a university merger report with a \$260 million mistake

SOURCES The Guardian, 25.10 [Guardian](#)

[single source] The government of Western Australia paid Deloitte \$166 000 for an analysis of a merger of three universities. Earlier, critics found references in the report that AI generated.

Now the Guardian also found arithmetic errors. In one scenario, the report says that \$514 million plus \$145 million give \$399 million. The correct sum is \$659 million. The gap is \$260 million, and in another scenario the gap is \$311 million.

One reference to a WHO report has a ChatGPT tag in its address, thus somebody copied it from a chat.

Deloitte admits that AI helped with material collection and proofreading. At first, the company called the errors minor. Now the state government demands the money back. Earlier, Deloitte already returned money to the government of Australia for a report with AI errors.

Why it matters. AI in consulting gives a report faster, but it does not check the numbers for people. A client now has a simple test: recalculate the table totals and open each reference.

Contracts with consultants will get clauses about AI use and penalties.

TOPIC 9

"Build more ambitious things": Dan Luu and Ethan Mollick about the same idea

SOURCES Dan Luu [Danluu](#) · HN, 25.10 [Hacker News](#) · [@emollick](#), 25.10 [@emollick](#)

The essay of Dan Luu became the most popular HN story of the day: 759 points and 698 comments. The example of Luu is pgrust, a Rust rewrite of Postgres by Michael Malis. Luu calls most such rewrites "vibeslop". But the pgrust tests are so careful that they already found real bugs in Postgres itself. The Postgres team fixed them, thus Postgres won even if pgrust fails.

The main thesis of Luu: most projects with agents are weak, but this does not limit a careful builder. Projects that were too large before are now realistic, if they have tests and benchmarks.

On 25.10, Ethan Mollick from Wharton gave similar advice: ask yourself how to make a task even more ambitious. Thus a person either hits the limit of the model or sees that the

model can do more.

Why it matters. skepticism about AI projects in the feed is mostly justified. But the price of a large rewrite fell. Now tests and benchmarks separate such a project from garbage.

misc

- **JetBrains ThinkRail:** JetBrains released a light open IDE for the agent pi. Each task lives in a separate git worktree, and project knowledge stays in "living specifications". The project has 1,9 thousand stars on GitHub.

[GitHub](#) · [Hacker News](#)

- **Cerebras and models with 3 trillion parameters:** the company explained how it splits Kimi K3 between WSE-3T wafers with 44 GB of memory each. A layer takes approximately 16 GB, and only activations move between the wafers. The stated speed is 1000-2000+ tokens per second.

[Cerebras](#) · [Hacker News](#)

- **Nikon took back a victory:** Nikon took the title from the winner of the Small World in Motion micro-video contest. Scientists noticed that structures in the video appear from nowhere. The author admitted that he used AI in processing. [Techspot](#) · [Hacker News](#)

- **DreamHost incident:** the hosting company reset the passwords of some mail and SFTP accounts.

The file manager of the panel does not work. The company does not name the cause.

[Dreamhoststatus](#) · [Hacker News](#)

- **Rest of World can close:** the publication about technology outside the West prepares to close at the end of January. Traffic from search and social networks disappeared, and Semafor names AI summaries in search among the causes.

[Semafor](#)

- **Gumroad and Valkey:** @ershus moved all Redis of Gumroad to Valkey and cut the ElastiCache bill by 69%, approximately \$2 200 per month. The code needed no changes.

[@ershus](#)