

Apple restricts full disk access because of AI agents

Georgia secret ballots, a \$300M Nvidia smuggling case, 62% vs 33% across harnesses, and Supabase buys Turso.

Saturday, 3 October 2026

IN THIS ISSUE

1. Apple restricts Full Disk Access on macOS: agents got too much
2. Ballot secrecy in Georgia: a \$20 agent matched voters to their ballots
3. The US arrests a CEO accused of smuggling \$300M of Nvidia chips into China
4. Same model, 62% in one harness and 33% in another. Hugging Face trained it to work in four
5. llama.cpp can now run decision models locally
6. Supabase is buying Turso: databases for agents, by the million
7. Amazon promises \$1B to data center communities and draws a new wave of criticism
8. Tokens get cheaper, bills get bigger: consultants warn about the cost of agents
9. Trillium Labs: a non-profit for open science on post-training
10. Underdog 27B: "Opus 4.6 level on a laptop," but the numbers are from the base model

TOPIC 1

Apple restricts Full Disk Access on macOS: agents got too much

SOURCES Apple Developer News, 02.10 [Apple](#) · Ars Technica, 03.10 [Ars Technica](#) · HN discussion, 02.10 [Hacker News](#)

Apple announced it will change the Full Disk Access permission in macOS. In its note to developers, the company says the permission "largely sidesteps" the usual private data controls and exists so that backup apps can work on the Mac. Now, Apple says, "some developers are using Full Disk Access in ways that could put users at risk," exposing files, mail, messages and even browsing history "without users' full knowledge and understanding." Then the key sentence: "As AI agents become increasingly capable and autonomous, the risks associated with this level of access will grow substantially." Granting an app this permission will require "very explicit user action." Apple gave no details of what that will look like.

The company names nobody, but Ars Technica ties the statement to the story around Muse, Meta's personal agent. Two weeks ago columnist Jason Aten wrote that Muse sent him a notification referencing his conversation with a co-worker in Apple Messages, even though he never gave the agent access to his messages. Meta CTO David Singleton replied that Muse reads Messages only when the user has enabled both the system-level Full Disk

Access and a separate connector inside the app. macOS security researcher Patrick Wardle disagreed: with Full Disk Access any app can read any of the user's files, including browsing history, cookies and chats. Apple's statement effectively sides with Wardle.

On 23.09 this digest covered a configuration hole in Muse, found by the same Wardle, and Amazon blocking the agent on its platform. Now the owner of the operating system has stepped in.

Reaction on HN was split. Some point out that the settings pane already warns explicitly about access to Mail, Messages and Safari and requires an administrator password, so it is unclear what more Apple can add. Others complain that macOS is already drowning in permission prompts.

A third group went to check their own lists and found Spotify there.

Why it matters. Desktop agents need broad access to be useful, and the easiest way to get it on macOS has been a single switch that opens everything. If Apple makes that switch harder, agent developers will have to move to narrow permissions for specific data, and users should already review who has this access: it lives in System Settings, under Privacy & Security.

TOPIC 2

Ballot secrecy in Georgia: a \$20 agent matched voters to their ballots

SOURCES The Guardian, 02.10 [Guardian](#) [single source]

Max Springer, a postdoctoral fellow at Princeton's Center for Information Technology Policy, took Georgia's public election records, obtained through an Open Records Act request, and a \$20 subscription to a language model. "Within a couple of hours, I had a pipeline to analyze and identify secret ballots across the state of Georgia and the agent told me exactly what further information it would need to identify real voters' ballots," Springer wrote. "At no point did the agent refuse to comply or raise concerns over implementing the exploit." He has never been to Georgia.

The mechanism is simple. When a voter feeds a ballot into the scanner, a digital record of the vote is created and the scanner assigns it an identifier. According to cryptographer Ben Adida, founder of the non-profit VotingWorks, those identifiers are not random enough: the order in which ballots were cast can be recovered from them. Match that order against the early-voting list and you can see who voted how. Springer recovered the voting order for 1.52 million ballots, 98.9% of in-person ballots in 114 of the 139 counties he examined. In tiny Heard county, the agent matched most of the 650 early in-person voters to a specific ballot, and the rest to within a single swap.

Security researchers found the underlying flaw in the software about four years ago. According to the Guardian, other states that use Dominion software have either patched it

or withhold the vulnerable data from the public. The Georgia secretary of state's office says it asked legislators for money to upgrade the system for three years and got "a big fat goose egg." On Thursday the state election board held an emergency meeting. Secretary of State Brad Raffensperger ordered the ID numbers redacted from released data, while part of the board argues that this violates the law on transparency of election records. Early voting in the midterms starts in less than two weeks.

Why it matters. The attack itself is old, and what changed is its price. Linking records used to take a skilled researcher and weeks of work; now it took a few hours and the cheapest subscription. The shift applies to any "anonymized" data: if one dataset can be joined with another public one, an agent will do it quickly and without objection. Anyone publishing data should ask what a person with an agent and a free weekend could do with it.

TOPIC 3

The US arrests a CEO accused of smuggling \$300M of Nvidia chips into China

SOURCES [Ars Technica](#), 02.10 [Ars Technica](#) · [The Guardian](#), 02.10 [Guardian](#)

The US Department of Justice accuses Greg Lui, 38, CEO of Earthmade Computer, of using false paperwork to mask shipments of servers worth more than \$300 million that he knew were bound for China. According to the FBI, the scheme ran from October 2023 to August 2026: servers with A100 and H100 GPUs were declared as destined for Malaysia and moved through freight forwarders in Malaysia and Singapore, then via Hong Kong. One batch of 92 servers allegedly reached a company in Hangzhou. For one shipment, prosecutors say, Lui used identity documents of another person that he had bought three years earlier. In 2024 alone his firm allegedly received more than \$176 million from the scheme. He faces three counts, the most serious carrying up to 20 years in prison.

Nvidia told Bloomberg that "less than one half of one percent" of its products have allegedly been diverted to China and called it "a drop in the bucket" compared with China's own domestic compute. Ars summarizes a Bloomberg investigation: officials in several countries believe a shadow market moves hundreds of thousands of chips, and ask why Nvidia missed obvious red flags, such as orders that could not fit into the buyer's facility. Nvidia replies that a startup's growth plan in a friendly nation is an opportunity, not a red flag.

Why it matters. Export controls rest on checking the end buyer, and every case like this shows that the check relies on paperwork that is easy to forge. As long as the seller is responsible for the check and earns money on every order, expect more arrests than closed channels.

TOPIC 4

Same model, 62% in one harness and 33% in another. Hugging Face trained it to work in four

SOURCES Hugging Face, multi-harness RL guide, 01.10 [Hugging Face](#) · @huggingface on X, 02.10 [@huggingface](#)

A Hugging Face team published a large open guide to training agent models with reinforcement learning inside several harnesses at once. The starting number comes from their own measurement:

Liquid AI's LFM2.5-2.6B, with identical weights, solves 62% of tasks under Mini-SWE-Agent and only 33% under Claude Code. Almost 30 points of difference come from the wrapper around the model.

The trick is to leave the harness alone. Instead of the model, the harness talks to a proxy that speaks all four API formats coding agents use: OpenAI Chat Completions and Responses, Anthropic Messages and the Gemini format. The proxy records the exact tokens and log-probabilities vLLM sampled, and the model is trained on that. Not a single line of Claude Code, Codex or OpenCode changes.

Results on held-out tasks: after training in four harnesses at once (OpenCode, Claude Code, Codex and Mini-SWE-Agent), the model went from 42.2% to 54.2% on average and made 31% fewer tool calls on tasks it already solved. A model trained only in OpenCode is stronger in OpenCode itself (58% vs 50%), but the multi-harness one is better in Claude Code (49% vs 42%) and Codex (54% vs 43%).

The shortcut, fine-tuning on 3,189 successful trajectories from the larger Qwen3.8-27B, plateaued at 47.5%, below both RL runs. Everything is open: the proxy in OpenEnv, the trainer in TRL, the tasks, the data and seven trained models.

Commenters on X pointed out the weak spot: all four harnesses were both in training and in the test. Whether the skill transfers to a harness the model has never seen is not tested for their own model.

Why it matters. Agent benchmark numbers that do not name the harness say little: the gap between wrappers can be larger than the gap between models. When comparing models for your own tasks, keep the harness fixed and run them in the one you actually use.

TOPIC 5

llama.cpp can now run decision models locally

SOURCES Hugging Face blog (ggml-org), 02.10 [Hugging Face](#) · @gggerganov on X, 02.10 [@gggerganov](#) · The Wall Street Journal, 02.10 [WSJ](#)

The llama.cpp server gained a `/v1/systemone` endpoint for decision models. The API follows the System One format TypeSafe introduced with its Jev model, so existing clients only need a new base URL. Decision models do not generate text: they read a state once (text, JSON or a screenshot) and return a probability for each answer option. Typical uses are

routing a request, moderating content, checking whether an agent's step worked, or choosing its next action.

Five models are supported at launch: from Julia-1 with 144 million parameters (3 ms per answer)

to OpenJev with 27 billion, which reads images (43 ms, non-commercial CC BY-NC 4.0 license). Times were measured on one NVIDIA RTX PRO 6000. A useful tip from the guide itself: Julia-1 routed "I was charged twice" to shipping when the options were bare labels, and to billing with probability 0.99 once each option had a description. Cloudflare's Clef is promised next.

Yesterday this digest covered Clef, Cloudflare's open decision model, and on 26.09 Ollaya, a local Jev-style server. Now support has landed in the most widely used engine for local models.

The same day the WSJ ran a headline saying Jev has sparked copycats and talk of alternatives to large language models; the article is behind a paywall.

Why it matters. Many decisions inside agent pipelines are still made by a large model that writes text, which is then parsed with regular expressions. A decision model handles such a task in milliseconds, locally, with a probability you can use as a threshold: act on confident answers, send the rest to a human.

TOPIC 6

Supabase is buying Turso: databases for agents, by the million

SOURCES Supabase blog, 02.10 [Supabase](#) · [@kiwicopple](#) on X, 02.10 [@kiwicopple](#) · HN discussion, 02.10 [Hacker News](#)

Supabase announced it is acquiring Turso. CEO Paul Copplestone's post is explicitly about agents:

he says Supabase already launches over a million databases a week, and agents should be able to create a database as easily and cheaply as a file. For small workloads that means not provisioning a dedicated machine for every database. Turso rebuilt SQLite in Rust and made a cloud where a single server manages millions of databases, loading them on demand and suspending them when idle. Supabase stays on Postgres, Turso continues with SQLite, and Turso co-founder Glauber Costa will lead the agentic infrastructure effort. The price was not disclosed.

On X, Charlie Coppingier wrote that 60% of the latest YC batch use Supabase, and Garry Tan reposted it. On HN reactions were mixed: some are glad the Turso project no longer depends on the fate of a small company, others, also customers, are unhappy, and one reminded readers that Turso failed to get into ClickBench several times because new bugs kept turning up.

Why it matters. Agents writing prototypes create databases by the thousand and abandon most of them within a day. Economics built around a database as a long-lived server do not

fit that, and providers are rebuilding around "a database as a file." For teams that let agents create environments, the cost of idle databases becomes a real question.

TOPIC 7

Amazon promises \$1B to data center communities and draws a new wave of criticism

SOURCES [Ars Technica](#), 02.10 [Ars Technica](#) · [Semafor](#), 02.10 [Semafor](#) · [The New York Times](#), 02.10 [NYT](#)

Amazon pledged more than \$1 billion over five years to communities near its data centers: free community college, trades training and home energy-efficiency upgrades.

Communities will decide how the money is spent. The company will also publish its energy and water use every year, promises that data centers will not raise power bills, aims to be "water positive" by 2030 and is dropping nondisclosure agreements when building. AWS CEO Matt Garman, in a roughly 3,000-word post, went through what he called "myths" about water, energy and pollution, and said US rivals are trying to "trick us into slowing down."

Environmental group Stand.Earth called it "corporate propaganda" in a comment to Ars. Its main argument: Amazon is building a power plant in Pecos, Texas that, the group says, will become the largest source of climate emissions in the US, and the commitments say nothing about it. The second is scale: \$1 billion over five years against the \$220 billion the company is investing in data centers in 2026 alone. Garman himself said Amazon gave the same amount to communities over the previous three years. Stand.Earth did praise the end of NDAs.

Why it matters. Opposition to data centers is already blocking billions of dollars of projects, and hyperscalers have moved from ignoring communities to negotiating with them. For the industry that is a new line in the cost of compute; for local governments it is leverage: a building permit can now come with public commitments attached.

TOPIC 8

Tokens get cheaper, bills get bigger: consultants warn about the cost of agents

SOURCES [Semafor](#), 02.10 [Semafor](#) · [Financial Times](#), 02.10 [Ft](#)

Semafor brings together two new reports. Bain & Company estimates that the industry's compute demand will require \$6 trillion in annual revenue by 2031, about \$4.2 trillion of it from markets and products that do not exist yet. In a separate report on US government buyers, McKinsey warns that the per-token price is collapsing while the total bill rises, because agents consume far more compute. "Usage might have been essentially near-zero cost," McKinsey senior partner Tim Ward told Semafor. Very soon, "it won't be." The same

day the FT published an interactive titled "AI got smarter. The bills got harder to control"; it is paywalled and only the headline is available.

Why it matters. The conversation inside companies is moving from "does AI work" to "can the budget take it." An agent making dozens of model calls per task multiplies cost even with cheap tokens. Practical measures such as step limits, caching and routing small decisions to small models (see items 4 and 5) stop being optimization and become budget discipline.

TOPIC 9

Trillium Labs: a non-profit for open science on post-training

SOURCES Trillium Labs blog, 01.10 [Trilliumlabs](#) · [@natolambert](#) on X, 02.10 [@natolambert](#) · WIRED, 02.10 [Wired](#)

Nathan Lambert, who previously co-led the Olmo models at Ai2, co-founded the non-profit Trillium Labs with Tom Zick. The goal is fully open post-training recipes: data, code, evaluations and intermediate checkpoints, so outside researchers can study how model behavior develops. Next they promise open infrastructure for research on recursive self-improvement, reward hacking and multi-agent systems. Initial support comes from Halcyon Futures and Schmidt Sciences; the lab is fundraising, hiring and looking for compute. Advisors include Thomas Wolf of Hugging Face and Hanna Hajishirzi.

The founders' argument: post-training has become central to reasoning and agentic capabilities, yet almost no complete recipes remain public, because their commercial value has grown. A non-profit structure, they say, lets them publish failed runs too.

Why it matters. Most open models today open the weights, but not how they were post-trained.

Without the recipe an outside researcher sees the result but cannot check which decision led to it. If Trillium finds the compute, the open ecosystem gets one more place where post-training can be reproduced in full.

TOPIC 10

Underdog 27B: "Opus 4.6 level on a laptop," but the numbers are from the base model

SOURCES Sigil Wen's article on X, 02.10 [@0xSigil](#) [single source]

Sigil Wen's startup Underdog went public with a manifesto and a list of backers including a16z, Khosla Ventures, Naval and several well-known researchers. The pitch: "6 months ago, the #1 AI was Claude Opus 4.6. Our model, Underdog 27B beats it while running on your Mac or iPhone." The product is a personal assistant for mail, calendar, notes and messages where both the model and the data stay on the device, encrypted, with no company servers for personal data. Separately the company claims its own inference engine, Husky, is up to 4.5 times faster than Apple's MLX on the same weights.

The fine print is in the article itself. Underdog 27B is built on Qwen 3.8 27B, and the comparison table against Opus 4.6 is for the Qwen base model, as stated outright: "These are base-model results." So the headline claim rests on someone else's benchmarks for a different model, and the article gives no independent measurements of Underdog 27B.

Why it matters. The privacy argument is strong: data that never left the device cannot leak from a server. But a local model with access to mail and messages is the same broad access as in item 1, just without the cloud. A claim like "our model beats the frontier of six months ago" is worth testing on your own tasks, because here it rests on another model's numbers.

misc

- **1,313 vulnerabilities in one Linux kernel update.** Debian issued a security advisory covering 1,313 CVEs, 1,295 of them from this year. The advisory is dated 29.09; it took off on HN on 02.10 with 550 points, and the first question in the thread was whether these are AI findings.
[Lwn](#) Alongside it, Greg Kroah-Hartman's talk "Security in the LLM age" from Kernel Recipes, 189 points. [Youtube](#)
- **DeepSeek Harness shipped as a desktop app for macOS and Windows**, with plugins, scheduled tasks and a bundled dsh command. 383 points on HN; on 14.08 this digest covered the launch of the harness itself. [Deepseek](#)
- **@lennysan:** five minutes before a live DevDay demo, a Dot agent noticed production was down and offered to fix it. "I don't think you're there yet, little Dot, but thank you for trying."
[@lennysan](#)
- **@karpathy:** a simple eval - ask a model "land or water?" for 16,200 coordinate pairs and plot the answers as an image. "The models know. From compressing the internet."
[@karpathy](#)
- **A court agreed with the EFF:** Utah's VPN law demands a technical impossibility. 530 points on HN. [Eff](#)
- **"Frog and Toad and the Increasingly Capable Machines"** - a story in the style of Arnold Lobel's books about the two friends and machines that keep getting more capable. 543 points on HN. [Frogandtoad](#)
- **@emollick** on Alex Imas and Jacob Schaal's review of research on AI and the labor market (29.09): the best evidence supports only a narrow claim that AI may be affecting hiring of junior staff in the most exposed occupations, and even that is contested.
[Substack](#)
- **Trump is expected to name Jay Clayton as AI czar, according to the WSJ.** Headline only, the text is paywalled [single source].
[WSJ](#)

