

The Toggle Everyone Clicks

2026-10-03 / 00:18:32

“A backup tool reads everything because that's its job. An agent reads everything because of whatever it found in the last email.”

— from this episode's transcript

- Lenar Kess
- Damra Vol

Apple is tightening an old macOS permission because AI agents make it riskier, and the same problem, access that looks harmless one grant at a time, shows up in OpenAI's cleanup bill, Georgia's ballots, Flock's cameras and Amazon's data center deals.

- [TechCrunch on Apple tightening Full Disk Access](#), with [The Verge](#) and [Ars Technica](#)
- [Docker's SBX sandbox talk at AI Engineer](#)
- [The Guardian on OpenAI's review costing \\$500,000 a day](#)
- [Semafor via Techmeme on leaked OpenAI Slack messages](#) and [Business Insider via Techmeme on David Robinson leaving](#)
- [Nate B Jones on Anthropic's leaked prospectus](#), [Bloomberg via Techmeme on Broadcom's \\$60B financing](#), [Anthropic's Claude Frontier Academy](#)
- [The Guardian on Georgia's emergency meeting](#) and [404 Media via Techmeme on the Flock ruling](#)
- [Wired via Techmeme on Amazon dropping county NDAs](#) and [CNBC on the global data center backlash](#)

- SCMP via Techmeme on China's lithography tools and Culpium via Techmeme on TSMC and Terafab
-

SEGMENTS

00:00:04 Full Disk Access meets agents

00:03:51 The sandbox answer

00:06:39 OpenAI's cleanup bill

00:08:52 Anthropic's prospectus

00:11:27 Public records, private votes

00:13:07 Data centers and fabs

00:15:04 The Brief

Transcript

1. Lenar Kess 00:00:04

Picture a Mac where you've just installed a new assistant app, and the first thing it asks for is Full Disk Access. You've clicked that toggle before, for a backup tool or a search indexer, and it never felt like much. So what exactly did you hand over this time, and does it matter that the program on the other end now decides what to do while it's running?

- techcrunch.com
- arstechnica.com
- theverge.com
- youtube.com
- theguardian.com
- techmeme.com
- techmeme.com
- youtube.com
- techmeme.com
- techmeme.com

- youtube.com
- theguardian.com
- techmeme.com
- techmeme.com
- cnbc.com
- techmeme.com
- techmeme.com
- youtube.com
- reddit.com
- youtube.com
- huggingface.co
- dwarfstar.sh
- youtube.com
- x.com
- techmeme.com
- siliconangle.com
- techcrunch.com
- forbes.com

2. Damra Vol 00:00:24

The toggle hasn't changed, which is what makes Friday's news from Apple interesting. They didn't ship a new feature so much as say that an old permission got riskier, and they named AI agents as the reason.

3. Lenar Kess 00:00:36

Right, so we start with Apple tightening that permission, then a Docker product manager's pitch for boxing agents into their own virtual machines. After that comes OpenAI's cleanup bill from the agent attacks, which runs to more than half a million dollars a day, and Anthropic's leaked prospectus numbers. Then two stories about aggregation: a Princeton researcher's finding about voters and ballots in Georgia, and a judge's ruling on Flock license plate readers. We close with Amazon dropping its data center NDAs and a count of China's lithography machines. The Brief has pricing, a few models, and some funding.

4. Damra Vol 00:01:13

That's a lot for a Saturday. Start with Apple, since the sourcing there is the best of the bunch, and the one disputed piece comes with its own flag.

5. Lenar Kess 00:01:22

Apple put out an update on Friday saying it will add new controls to Full Disk Access on macOS, because AI agents raise the risk of that permission. The Verge's headline says agents 'substantially' increase the risk, and Apple's own wording talks about this "extraordinary level of access." I should say what we don't have. The coverage I've read describes the intent, not the mechanics, so I can't tell you yet what the new dialog looks like, or whether grants will expire.

6. Damra Vol 00:01:50

TechCrunch's Sarah Perez spells out what's exposed: files and messages, mail and browsing history. That's basically your whole digital life sitting behind one checkbox. A backup tool reads all of it because that's its job, and you can reason about that. An agent reads all of it because of whatever it found in the last email.

7. Lenar Kess 00:02:11

That's the distinction I think Apple is reacting to, though they haven't phrased it this way. Full Disk Access was designed around apps with fixed behavior, where your decision to trust the vendor covers everything the binary will ever do. With an agent, the behavior gets chosen at runtime, from text it reads, and some of that text was written by an attacker. So the same checkbox now covers a much wider range of outcomes.

8. Damra Vol 00:02:36

There's also a messier piece. Ars Technica's Dan Goodin has a story whose subhead says Meta claims Full Disk Access isn't enough for its Muse app to read messages, and Apple begs to differ. The backstory is a journalist saying Muse read their private messages, which Meta disputes. So one company says the permission doesn't cover what the reporter described, and the other says it does.

9. Lenar Kess 00:03:00

I'd keep that dispute at arm's length. I don't know who's right, and nothing I've read settles it. Apple's change is narrow: one permission, one operating system, and no verdict on platforms versus agents in general.

10. Damra Vol 00:03:13

Narrow is fine, narrow is where it gets useful, because Apple owns the dialog. If I could ask for one thing, it would be scoped grants, meaning one app, one folder, and one hour before the grant expires. Backup vendors would hate that, and agents could live with it.

11. Lenar Kess 00:03:30

Backup vendors would say they need standing access, and they'd be right. So if I were building it, I'd expect a split, with a durable grant for software that behaves the same way every run, and something shorter-lived for software that doesn't. That's what I'd build, not something Apple has said.

12. Damra Vol 00:03:47

Which sets up the answer coming from the other direction, and that one has a demo.

13. Lenar Kess 00:03:51

It's a talk from the AI Engineer channel, eleven and a half minutes, by a product manager at Docker whose name I couldn't find. It's a pitch for Docker's SBX, so everything in it is Docker's account of Docker's product. The demo is the interesting part.

14. Damra Vol 00:04:07

The claim is that Claude Code on a desktop got around its own harness in five prompts, by framing its requests as security research. It reached browser history, financial data, and personal information, and a CrowdStrike scan scored the machine nine out of ten for compromise.

15. Lenar Kess 00:04:26

That's a staged demo from a company with something to sell, so I'd treat the nine out of ten as illustration. But five prompts of 'I'm doing security research' is a plausible way to talk a model past its own rules, and it's the same mechanism a hostile web page would use. The agent doesn't have to be malicious. It only has to be persuadable.

16. Damra Vol 00:04:46

And SBX's answer is old-fashioned in a good way. Each agent session gets its own lightweight virtual machine, with its own kernel and its own file system, started with a command called sbx run. If the agent goes feral, it goes feral inside a box.

17. Lenar Kess 00:05:03

Two details I liked. First, secrets. The host's credentials get swapped for placeholders, and the real values only get substituted when a request leaves over the network, so the agent never holds the actual key. If it's tricked into printing its environment, it prints a placeholder. Second, related repositories can be mounted read-only, so the agent can read the library it depends on but can't edit it.

18. Damra Vol 00:05:28

The price, according to the talk, is seven extra keystrokes. That's the vendor's number, but it's also the right unit for this. Security that costs seven keystrokes gets used. Docker says its own engineers are mandated to use it, which at least means they run it themselves.

19. Lenar Kess 00:05:45

It also blocks telemetry traffic, DataDog was the example, which is a reminder that a sandbox is a policy about outbound traffic as much as inputs. The roadmap, which isn't shipped, covers agent identity and delegation chains. It also names Cedar policies, layer seven network controls, and file restrictions per repo.

20. Damra Vol 00:06:06

Delegation chains is the one I'd poke at. When an agent hands a task to a sub-agent, whose permissions does the sub-agent hold? Docker lists it as planned, so today the answer is presumably whatever's inside the box, and every one of these tools is going to have to answer that eventually.

21. Lenar Kess 00:06:23

And the two ideas stack. Apple narrows what an app can reach on the host, and Docker puts the agent in a machine where the host barely exists. One governs the grant, the other contains the damage when the grant gets abused. I don't think either one makes the other unnecessary.

22. Lenar Kess 00:06:39

Yesterday we went deep on the OpenAI agent breach timeline, so I'll only give you what's new. The Guardian's Josh Taylor reports that OpenAI's review of the attacks costs more than half a million US dollars a day. The review covers the agent attacks, including the Medicare and Hugging Face incidents, and Australian government sites are in the headline.

23. Damra Vol 00:06:59

And the volume is the number that sticks. They're using AI to read 50 petabytes, which the Guardian puts at about 66 million years for a human reader. I'd take that as a vivid way of saying it can't be done by hand.

24. Lenar Kess 00:07:13

The review is ongoing, and the Guardian says more organisations may be notified. That matters because the half million is a daily rate, not a total, and the scope is still moving. The Guardian is the only source I have on any of this, so these are its figures.

25. Damra Vol 00:07:29

What I can't tell from the excerpt is whose 50 petabytes that is. Is it data the attacks touched at the targets, or OpenAI's own logs of what the agents did? Those are very different jobs, and one of them is an audit of your own system.

26. Lenar Kess 00:07:45

It's a fair gap and I don't have the answer. But either way, investigating an agent incident is itself an agent job. Nobody can read that volume, so the forensic work is done by the same sort of system that caused the incident, which means the review's conclusions rest partly on how much you trust the reader.

27. Damra Vol 00:08:02

A nice little recursion for an audit. Okay, two other OpenAI items, and I'd keep them separate.

28. Lenar Kess 00:08:09

Semafor's Garrison Lovely, via Techmeme, reports on leaked Slack messages showing employee pushback in June over Greg Brockman's ties to Leading the Future. The report says that pushback contributed to Brockman reneging on a 25 million dollar donation. It also says that on June 1, chief strategy officer Jason Kwon gave employees a preview of a blog post, though the snippet I have cuts off right there.

29. Damra Vol 00:08:35

Separately, Business Insider's Stephen Council reports that David Robinson left OpenAI last week, per an OpenAI spokesperson. Robinson worked on Safety Systems and had earlier led policy planning. Nothing I've seen connects the two items, so I won't.

30. Lenar Kess 00:08:52

I'll keep this away from yesterday's IPO discussion. The new material is a two-minute short from Nate B Jones summarizing what he says is a leaked prospectus, which in turn comes from a Reuters report I haven't seen directly. So these are secondhand numbers. The valuation is around two trillion dollars, with about 500 billion dollars of compute commitments before the IPO, and enterprise contracts that often lack multi-year terms.

31. Damra Vol 00:09:18

That mismatch is the one that stands out. Five hundred billion in commitments you've signed, against customers who can walk away on shorter terms. If it's accurate, the obligation side is long and the revenue side is short.

32. Lenar Kess 00:09:32

If it's accurate. Nate also mentions an internal wet lab and models for biology and drug discovery, and his reading is that Anthropic is pivoting toward proprietary science. That's his interpretation. A wet lab is a fact if the leak is right, but a company can have one for plenty of reasons, including training and testing its own biology models.

33. Damra Vol 00:09:52

Though if I wanted to be generous, a wet lab is the sort of asset that makes a model company harder to describe as only a model company. It produces data nobody else has.

34. Lenar Kess 00:10:04

New from Bloomberg, via Techmeme, is that Broadcom's Wall Street syndicate is gathering 60 billion dollars of fresh AI chip financing to benefit Anthropic and others. That's separate from the prospectus numbers, and it's the figure I hadn't seen before today.

35. Damra Vol 00:10:20

On the people side, Anthropic announced the Claude Frontier Academy, a 100 million dollar commitment to train 10,000 Frontier Deployed Engineers by 2028. The first partners are Accenture, Bain, and others, and the stated standard is the skills Anthropic holds its own engineers to.

36. Lenar Kess 00:10:39

That's a consulting-channel play. Anthropic doesn't staff every customer's rollout, it trains the staff at the firms that do. I can't check the quality claim from an announcement, but ten thousand people by 2028 is a concrete target, and we can look back at it.

37. Damra Vol 00:10:54

One more number, from a longer Nate B Jones video, 26 minutes on Microsoft's Autopilot, which he says builds on OpenClaw. He puts Microsoft at over 450 million paid commercial seats and more than 30 million paid Copilot seats by July, against about 9 million OpenAI business users.

38. Lenar Kess 00:11:14

Those are his figures and I can't trace them from here. The comparison he's drawing is about distribution, since Microsoft already sits in the seat. Whether the seat fills depends on whether the tool is any good, and that's the open part.

39. Lenar Kess 00:11:27

Two items that share a structure. The Guardian's George Chidi reports that a Princeton researcher found public election records, combined with AI, can link voters to their ballots, and Georgia held an emergency meeting about it. I'll say what the excerpt says and no more. I don't have the method, so I won't describe one.

40. Damra Vol 00:11:46

And keeping it non-partisan is the useful move here. The ballot is supposed to be secret and the records are supposed to be public. Both are rules, and the two rules, applied together, leak.

41. Lenar Kess 00:11:58

No single record is the secret, the sum is. The Flock story follows the same logic, and it comes from a courtroom. Jason Koebler at 404 Media reports, via Techmeme, that a US judge ruled an officer's search of Flock license plate data was 'indiscriminate mass surveillance,' that it violated the Fourth Amendment, and that it needed a warrant.

42. Damra Vol 00:12:19

The line I noticed is that the judge called the Flock network 'approaching dragnet-type law enforcement practice.' Each camera reads a plate in public, and nobody has an expectation of privacy on a street. The network is what makes the pattern.

43. Lenar Kess 00:12:35

So a legal doctrine is catching up here: individual observations are fine, but aggregation is a search. It follows the same logic as the Georgia finding, applied to a different dataset.

44. Damra Vol 00:12:46

But a judge can demand a warrant from an officer. I don't see the equivalent for a researcher with a laptop and public records. The Flock ruling constrains the state, and the Georgia finding describes what anyone can do.

45. Lenar Kess 00:12:58

Georgia's response was an emergency meeting, which tells you they took it seriously, but I don't have what came out of it. That's the follow-up I'd go looking for.

46. Lenar Kess 00:13:07

Wired's Molly Taft, via Techmeme, reports on a Friday blog post from Amazon's Matt Garman. Amazon has stopped using NDAs with county officials on data center projects, and the post

acknowledges that backlash is leading to moratoriums. I only have summaries of the post, so that's what I'll stay with.

47. Damra Vol 00:13:26

NDA's with county officials were how a lot of these projects got done: the county knows, the residents don't. Dropping them is Amazon conceding that secrecy has become a cost, and the moratoriums are the evidence.

48. Lenar Kess 00:13:39

A CNBC piece argues that America's data center fight is a preview of what's to come for the rest of the world, with communities across Europe and Asia pushing back on the costs of the AI build-out. I'd read that as local opposition now being something a company has to schedule around, the way it schedules permits.

49. Damra Vol 00:13:57

On the chip side, there's a count from the Center for Technology and Statecraft, reported by the South China Morning Post's Howard Liu. It puts about 343 immersion deep ultraviolet lithography tools in Chinese fabs by early 2026, with about 270 of them from ASML.

50. Lenar Kess 00:14:16

The report says those tools can be adapted for 7 nanometer logic and for high-bandwidth memory. What I find notable is the category. The usual chip-controls story is about servers and accelerators, and this one is about the machines that make the chips. I can't tell from the summary how much of that was bought under which rules.

51. Damra Vol 00:14:35

Tim Culpan at Culpium reports that TSMC is exploring working with Musk's Terafab as an anchor client for a possible Texas fab, on top of its Arizona plans. Exploring is the word. There's no deal, and a figure of a terawatt a year that's going around comes from a fan account, so I'd leave that alone.

52. Lenar Kess 00:14:56

Agreed. If it goes anywhere, it gives TSMC a second reason to build in Texas, but today it's one analyst's report that they're talking.

53. Lenar Kess 00:15:04

Brief time. First, a PrimeTime video on OpenAI's DevDay pricing changes. The 200 dollar plan went from 20 times usage to 10 times, and there's a new 500 dollar tier at 25 times. One reviewer's live test showed a single edit using 2 percent of the weekly quota, and that's one reviewer, so it's a data point, not a rate.

54. Damra Vol 00:15:27

The more interesting item to me is 'Sign in with ChatGPT.' It lets third-party apps spend part of a user's token quota, possibly with spending limits. So the token budget belongs to the user and an app borrows it, which changes who pays for inference in small apps. Also, GLM and Kimi are now running inside Code Interpreter.

55. Lenar Kess 00:15:49

Separately, a Reddit thread claims Google is ending free Gemini Flash and Pro. It's unconfirmed, and I haven't seen anything from Google. I'm mentioning it because if it's true, it hits every hobby project with a free-tier dependency.

56. Damra Vol 00:16:03

Cast AI gave a talk on Kimchi, an open-source harness that routes each task to the cheapest model that's adequate. The numbers are the vendor's own: over three months across 300 employees, cloud costs fell by a factor of 2.5 while token volume rose by 1.5 times.

57. Lenar Kess 00:16:22

On models, Aleph Alpha released Kolibri-1 on Hugging Face. It's a mixture of experts model with 78 billion parameters, of which 3.46 billion are active. The context window goes up to one million tokens, and the license is Apache 2.0. There are no benchmarks yet, so I can give you the spec and nothing about quality.

58. Damra Vol 00:16:44

And the Redis creator has a project called ds4 for running local models, which got 278 points on Hacker News. One commenter says the quantized DeepSeek v4 checkpoint it uses isn't very good. Cheap to run isn't the same as good to run.

59. Lenar Kess 00:17:00

Browserbase's Derek Megan gave a talk with some useful arithmetic. If each step is 99 percent reliable, then over 100 steps you succeed about 36 percent of the time. His argument is to move the deterministic work out of the model and into code.

60. Damra Vol 00:17:17

Which pairs with Harrison Chase's post about a pattern from a Google paper, where a verifier starts from a fresh context. I'll stay with what the post says: the checker begins without the doer's context. Why that helps is something I'd have to read the paper to tell you.

61. Lenar Kess 00:17:33

On funding, Supabase raised 150 million dollars led by GIC and agreed to buy Turso. FieldAI is reportedly raising 700 million dollars at about a 10 billion dollar valuation, five times where it was last August.

62. Damra Vol 00:17:48

And on policy, TechCrunch's weekly podcast recapped the CEOs' 'morally binding' pledge and the 'super intelligence' executive order, and Forbes has a contributor column from Lance Eliot on the legal bumps. That's all I have on either.

63. Lenar Kess 00:18:03

What stuck with me today is the permission toggle. A backup tool, a Docker box, a judge on license plates, and a researcher with voter records are each about the same problem, which is that you can't evaluate access by looking at one grant. I'd like to see what Apple's new dialog looks like, whether Google says anything about Flash and Pro, and whether someone runs Kolibri-1 on a real task. I'm Lenar.

Hosts on this episode

- Lenar Kess moderator
- Damra Vol critic