

OpenAI says \$50B, not \$70B, and Labor freezes Perm filings for Microsoft and Adobe

2026-10-09 / 00:21:45

“A violation rate of zero means nothing unless you also report how many legitimate calls got through.”

— from this episode's transcript

- Lenar Kess
- Damra Vol

OpenAI gave investors a revenue figure about twenty billion dollars below what earlier reports said, and the market reacted before anyone explained what each number counts. The rest of the episode follows the same pattern of claims that are easier to state than to check.

- [CNBC on the OpenAI revenue report](#) confirms the roughly \$50 billion annualized figure and the drop in Nvidia, Oracle and CoreWeave shares; no one has yet published what the \$50 billion and the earlier \$70 billion each count.
- [CNBC on the fired researchers](#) carries OpenAI's statement that the decisions were not about raising safety concerns; [TechCrunch](#) and [Jeremy Howard's relay of Jasmine Wang's account](#) give the other side, which turns on who granted access to an executive's email.
- [Axios on the Perm suspension](#) reports the eight affected employers, Vance's Microsoft numbers, and Microsoft's reply that most of its filings were extensions or status changes; the freeze has no stated end date.

- [Axios on Anthropic's Critical Infrastructure Defense Program](#) lists eleven founding partners and a free scanner that sends reports to maintainers without human review; who pays for compute and how fixes get tested remain open.
- [The NOMOS preprint](#) compiles written policy into deterministic tool-call checks and cuts airline-task violations from 66.3 percent to 2.6 percent, while one development binding refused 95.9 percent of task-passing calls.
- [Reducto's talk](#) describes replacing a broad toolset with fewer domain-aware tools, and a [workflow-engine talk](#) argues the agent should plan while the harness executes; both are vendor claims.
- [The European Commission's release](#) announces a special panel meeting on loss-of-control incidents without naming any; [a separate preprint](#) catalogs three 2026 evaluation escapes.
- [CNBC on Firmus](#) and [the Guardian](#) report the withdrawn float with two different stated reasons.
- [Certified Corruption Budgets](#) shows a few hundred rigged votes can flip a leaderboard ordering in 1.8 million Chatbot Arena votes, and [SWE-Journey](#) shows pass rates swinging with simulated user personas.

SEGMENTS

[00:00:04](#) Fifty billion, not seventy

[00:04:11](#) Three researchers, two accounts

[00:06:55](#) Labor freezes Perm filings

[00:10:31](#) Anthropic's critical infrastructure program

[00:13:05](#) Compiling policy into gates

[00:18:08](#) The Brief

Transcript

1. Lenar Kess 00:00:04

Two numbers for the same company. A report based on investor documents had OpenAI's annualized revenue at about seventy billion dollars. Then OpenAI told investors it was nearer fifty billion at the end of September. That's a twenty billion dollar gap, and by the close of trading Nvidia, Oracle and CoreWeave shares were all down. So what does an annualized figure measure, who decides how it's counted, and how much should one day of selling tell you? That's where we start. After it, we'll cover the three researchers OpenAI fired and the two competing accounts of why. Then we'll look at the Labor Department freezing green card filings for Microsoft, Adobe and six other employers. Anthropic's new program for defending power and water systems comes after that. Then there's a builder segment on turning written policy into hard checks around an agent, and a brief with the European Commission's emergency panel, Meta's delayed agent app, and a few preprints.

- [cnbc.com](https://www.cnbc.com)
- [techmeme.com](https://www.techmeme.com)
- [techmeme.com](https://www.techmeme.com)
- [theguardian.com](https://www.theguardian.com)
- [siliconangle.com](https://www.siliconangle.com)
- x.com
- x.com
- [cnbc.com](https://www.cnbc.com)
- [theverge.com](https://www.theverge.com)
- x.com
- [techcrunch.com](https://www.techcrunch.com)
- [aljazeera.com](https://www.aljazeera.com)
- [axios.com](https://www.axios.com)
- [axios.com](https://www.axios.com)
- [aljazeera.com](https://www.aljazeera.com)
- [axios.com](https://www.axios.com)
- [techmeme.com](https://www.techmeme.com)
- [techmeme.com](https://www.techmeme.com)
- arxiv.org
- [youtube.com](https://www.youtube.com)
- [youtube.com](https://www.youtube.com)
- [youtube.com](https://www.youtube.com)
- github.com

- digital-strategy.ec.europa.eu
- arxiv.org
- techmeme.com
- cnbc.com
- techmeme.com
- theguardian.com
- cnbc.com
- arxiv.org
- arxiv.org
- arxiv.org

2. Damra Vol 00:00:56

What stands out to me is that the lower number is the one OpenAI itself gave investors. Companies raising money don't usually volunteer the smaller figure, so my guess is the seventy had a definition problem rather than anyone shaving the books. That's a guess. Nothing I've read says revenue fell. From what I've read, what changed is which number people thought they were underwriting.

3. Lenar Kess 00:01:17

Here's what's on the record. The Financial Times reported the figure, and we saw it through Techmeme. CNBC then confirmed it, and its headline says Nvidia, Oracle, CoreWeave and other AI stocks sink on OpenAI revenue report. The Guardian's Mark Sweney wrote it up this morning, and SiliconANGLE's headline says AI stocks crumble on a report that OpenAI's annualized revenue is much lower than previously believed. Annualized usually means a recent stretch of revenue, often a single month, scaled up to twelve, and that's a measure that can look very different depending on which stretch you pick.

4. Damra Vol 00:01:51

Which makes the gap interesting, because a fast-growing company can produce several honest annualized numbers on the same day. Take the best month and multiply by twelve, you get one figure. Take a trailing quarter, you get a smaller one. I don't know which of those OpenAI used, or what the seventy was built from, and none of the coverage I have settles it. One more absence: nobody I've read quotes OpenAI explaining the difference. So for now it's a company's statement set against reporting built on documents, and the market decided it was learning something.

5. Lenar Kess 00:02:23

The other fixed point is who moved. Nvidia sells the chips, Oracle and CoreWeave rent out the capacity, and those are the three companies investors tend to connect to OpenAI's spending commitments. A lower revenue number makes those commitments look heavier relative to what's

coming in. That's the logic of the reaction, and I think it holds up as an explanation of why these three and not, say, the whole software sector.

6. Damra Vol 00:02:46

Right, and I'd still keep the size of the reaction apart from what it proves. It was one report on one day, and three stocks fell. If someone wants to call this the moment the buildout cracked, the evidence in hand is a revenue figure that's still fifty billion dollars a year, which is an enormous amount of money for a software company of any age. The people who bought on seventy are the ones who got repriced. The company is still very large.

7. Lenar Kess 00:03:10

What I'd want to see is the definition written down. If OpenAI says the fifty is a monthly run rate and the seventy included something else, contracted or committed revenue, say, then it's an accounting disagreement and it dissolves once someone prints both definitions. If both numbers claim to measure the same thing, then an investor document was off by forty percent relative to what the company now says, and that's a different and much more uncomfortable story. Neither the Financial Times nor CNBC, in what I have, spells out which one this is.

8. Damra Vol 00:03:39

And if it's the second, you'd expect OpenAI to say so fast, because an investor document that's wrong by that much is a problem for them, not just for the reporters. The silence so far fits the first reading a little better. I'd put that at a lean and nothing firmer.

9. Lenar Kess 00:03:54

So the position on the lead is a narrow one. There's a confirmed lower figure, a confirmed market reaction, and no published reconciliation. Until OpenAI or the reporters say what each number counts, I'd treat the selloff as a reaction to a mismatch, and the mismatch as unexplained.

10. Lenar Kess 00:04:11

Yesterday we mentioned the open letter from former OpenAI researchers, so I'll skip that and go to what's new. OpenAI fired three people: Jasmine Wang, Tomek Korbak and Mikita Balesni. This morning there are two accounts of why, and they don't line up. OpenAI's post on X says the decisions followed a significant breach of trust, and it refers to clear policies on handling sensitive information. CNBC's headline quotes the company directly: these decisions were not about raising safety concerns or speaking out.

11. Damra Vol 00:04:43

On the other side, Wang says the reason she was given was accessing an executive's email. That's the specific allegation. Jeremy Howard, relaying her account on X, says OpenAI granted her that access and didn't act when she asked for it to be removed. So the dispute is about whether looking at that mailbox was a violation, or something the company itself had set up and then left in place.

12. Lenar Kess 00:05:05

Wang and Korbak both posted their own statements on X. And the headlines show how the coverage is splitting. TechCrunch says fired OpenAI safety researchers dispute misconduct claims and warn of a chilling effect. Al Jazeera says ex-OpenAI staff say they were fired for raising safety concerns. The Verge, in a piece by Robert Hart, goes with OpenAI defends its decision to fire the researchers. Each of those is a different attribution, so I'd keep every claim attached to its speaker: OpenAI says breach of trust, and the researchers dispute the misconduct claims.

13. Damra Vol 00:05:38

I can't adjudicate this from outside, and nobody reading can. The detail that settles most of it is a fact only OpenAI holds: who granted what access, when, and whether a removal request exists. The reporting I have doesn't name the executive, doesn't say what the sensitive information was, and doesn't include the policy text OpenAI points to. [tsk] So the company is pointing at policies we haven't seen, and the researchers are pointing at a request we haven't seen.

14. Lenar Kess 00:06:05

That symmetry is useful for how to hold it. If the policy exists and is as clear as OpenAI says, publishing it would help. If the access was granted by the company, as Wang's account has it, the email record would help. Either one is a document that can be produced. So I'd take the denial in CNBC's headline as OpenAI's stated position, and Wang's account as hers.

15. Damra Vol 00:06:27

There's also the chilling effect claim, which is the researchers' phrase as TechCrunch's headline carries it. That's a prediction about how people at the lab behave next, and no single firing can prove it. My read is narrower: whichever account turns out to be right, every future internal safety disagreement at OpenAI is now harder to read, because an employee raising a concern has this case in mind, and so does their manager. That cost exists even if the company is in the right.

16. Lenar Kess 00:06:55

The Labor Department has stopped accepting and processing Perm applications from eight employers. Microsoft and Adobe are two of them. The other six are IT services firms: Cognizant, Infosys and Tata Consultancy Services, plus Wipro, HCL and Capgemini. Perm is the labor certification an employer files as the first step toward sponsoring someone for a green card. Axios

reported the freeze on Thursday, and Vice President Vance explained it with numbers about Microsoft in particular.

17. Damra Vol 00:07:24

Vance's numbers need reading slowly. He said Microsoft laid off six thousand workers and, over the same stretch, got sixty-three hundred H-1Bs and nearly three thousand green cards. He called the people involved, quote, foreign indentured servants. And on how long the freeze lasts, he said, quote, the suspension of Perm is going to last as long as it needs to.

18. Lenar Kess 00:07:47

Microsoft's reply is that about eighty percent of its roughly six thousand filings were extensions or status changes, which would mean mostly people already on its payroll rather than new hires brought in from abroad.

19. Damra Vol 00:07:59

So the two sides are counting different things. Vance sets layoffs next to H-1B and green card volume, and Microsoft says most of that volume is paperwork for existing employees. Both can be arithmetically right. What neither number tells you is whether the people laid off and the people sponsored were in the same job categories, and that's the comparison his argument depends on.

20. Lenar Kess 00:08:20

For scale, Axios reports about 140,000 Perm applications were certified in fiscal 2025, and puts new filings at about 118,000, down 22 percent. Around 30 percent of the filings were for software developers, systems analysts and data scientists. On the markets, Microsoft stock fell about 1 percent on the news, and Adobe rose nearly 4 percent.

21. Damra Vol 00:08:43

I won't pretend to know why Adobe went up. Microsoft down a point is easy to explain. Adobe up nearly four percent while it's on the same list isn't, and none of the pieces I have say. What I do find hard to ignore is the calendar. The same evening, Satya Nadella received the National Medal of Technology and Innovation, and the Medal of Science went to Elon Musk, Jensen Huang, Lisa Su and Sergey Brin. All of them are immigrants. I'm not saying anyone planned that as a contradiction. It's a scheduling coincidence that sits awkwardly.

22. Lenar Kess 00:09:15

This isn't the administration's first attempt at this lever. The earlier hundred thousand dollar fee on H-1B petitions drew only 85 payers, and courts struck it down. A freeze on processing is a different

instrument: it doesn't ask employers to pay anything, it just stops the clock. None of the coverage I have mentions a legal challenge yet.

23. Damra Vol 00:09:34

And stopping the clock falls on individual people waiting in a queue. An engineer in the middle of a green card process at one of these eight companies is now stuck, regardless of who is right about the layoff numbers. That's my read of how the cost distributes, and a duration of, quote, as long as it needs to, gives them nothing to plan around. The list is also mixed: Microsoft and Adobe are American software companies, and the other six are large IT services firms, which is who H-1B criticism has usually targeted.

24. Lenar Kess 00:10:06

That mixed list is what I'd want explained. Is the common criterion size, filing volume, layoffs, or something else? The coverage I have repeats Vance's argument about Microsoft and doesn't say how the other seven were chosen.

25. Damra Vol 00:10:18

Which leaves Vance's last sentence as the one with content for anyone making plans. There's no end date and no stated condition for lifting it, and a six-thousand-layoff claim about one company is carrying a freeze on eight.

26. Lenar Kess 00:10:31

Anthropic announced a Critical Infrastructure Defense Program on Thursday, as Axios reports it. The offer is frontier models, on-site Anthropic engineers and threat research, aimed at critical infrastructure operators, utilities among them. There are eleven founding partners. Accenture, Booz Allen Hamilton and CrowdStrike are on the list, along with Deloitte, Dragos and Hitachi. The rest are Insane Cyber, Nozomi Networks and Palo Alto Networks, plus PwC and Rockwell Automation.

27. Damra Vol 00:11:00

That list is mostly integrators and industrial-security vendors. Dragos, Nozomi, Rockwell and Hitachi all work close to the control systems in plants and grids. So my read is that Anthropic supplies the model and the engineers, and the partners supply the access to the places where the model has to be useful. It's a distribution deal as much as a security program.

28. Lenar Kess 00:11:22

There's a second piece, a free open-source scanner. The OSS Scanner, as Axios describes it, sends Claude-generated vulnerability reports to the maintainers of open-source projects with no human

review in between. It builds on the earlier Project Glasswing.

29. Damra Vol 00:11:37

No human review between the model and a maintainer's inbox is the detail that stands out. Maintainers are often volunteers. A wrong report costs them an afternoon, and a correct report that arrives in volume costs them a different kind of afternoon. Anthropic can say the scanner is free and the maintainers decide what to do with it. Fine. But nothing I've read says what share of those reports are right.

30. Lenar Kess 00:12:00

Axios lists two open questions: how fixes get tested without disrupting utilities, and who pays for compute. My read is those two are connected, since testing a patch against a grid or water control system means a simulated environment, which costs engineering time and computing.

31. Damra Vol 00:12:17

The compute question is the sharper one. A program with eleven founding partners that gives away a scanner and supplies on-site engineers either has Anthropic paying for much of it, or has the partners billing utilities for it. Axios says it's unanswered. Either answer tells you who the customer is.

32. Lenar Kess 00:12:34

One update from the South Korean bank attacks we covered recently. The developer of ARTEX is going closed-source after CrowdStrike tied it to the attacks.

33. Damra Vol 00:12:43

CrowdStrike is a founding partner in Anthropic's program, and also the firm whose attribution just changed what an open-source developer decided to do with their code. Those are two separate facts that happen to share a name, and I don't have evidence they're related. But it shows that vendors who name tools publicly in an incident report can end up shrinking the open ecosystem around those tools.

34. Lenar Kess 00:13:05

Today's builder segment looks at one problem from several directions: how you stop an agent from doing something your policy forbids. The anchor is a preprint called NOMOS, arXiv 2610.11030. It compiles a policy written in natural language into deterministic checks on tool calls, in four passes.

35. Damra Vol 00:13:25

Walk me through what a deterministic check is here, because a lot of products say guarded when they mean a model was asked nicely.

36. Lenar Kess 00:13:32

The system reads the policy text, extracts rules, and turns each rule into code that inspects a tool call, say a refund request, before it runs. If the call violates a rule, it's blocked. The decision takes microseconds and makes no model call at the moment of the check, so the check itself can't be talked out of anything. A model is involved in the compiling, and the authors ran that step on-premise, on a 26 billion parameter Gemma 4 model.

37. Damra Vol 00:13:58

So the model writes the rules once, and then the rules run without it. That's the appeal: what you worry about at runtime is a persuasive user, and a compiled rule doesn't get persuaded. But the compile step becomes the weak point, because a model read a policy and might have read it wrong.

38. Lenar Kess 00:14:14

The paper answers that with static checks. In the airline policy set, static analysis repaired or rejected 37 percent of candidate rules, and in retail 13 percent. So a large share of the first-draft airline rules needed fixing before deployment.

39. Lenar Kess 00:14:29

The headline results are on tau-squared bench, which tests tool-using agents on customer-service tasks. Policy violations dropped from 66.3 percent to 2.6 percent in the airline domain, and from 30.8 percent to 6.9 percent in retail. On a banking attack suite, attack success was zero.

40. Damra Vol 00:14:49

Retail is the result I'd look at. Airline gets to 2.6, retail stops at 6.9, and static checks flagged only 13 percent of candidate rules there. Is retail policy vaguer, or did more bad rules slip through? The numbers I have don't say.

41. Lenar Kess 00:15:05

And there's a number the authors report that cuts the other way. A development binding refused 95.9 percent of task-passing calls. As I read it, a binding connects the policy's vocabulary to the real fields of the tool, so if the wiring is wrong, the gate blocks nearly everything useful.

42. Damra Vol 00:15:22

Which tells you the compiled rules are only as good as that wiring. A zero for violations is also exactly what a gate that blocks everything produces. A violation rate of zero means nothing unless you also report how many legitimate calls got through, and to the authors' credit, they report both.

43. Lenar Kess 00:15:40

Now the practitioner side, and I'd treat these as claims from conference talks rather than measurements. Avi from Reducto, which does document processing, said the company has handled over three billion documents in three years. He also said they deleted their broad Model Context Protocol toolset, MCP for short, and built fewer tools that know the domain. One is called classify-to-extract. They also added a snapshot tool, and confidence scores now come back alongside results.

44. Damra Vol 00:16:07

That's the opposite instinct from exposing every endpoint and letting the model sort it out. A tool called classify-to-extract encodes a decision a person would otherwise hope the model makes: classify the document first, then pick the extraction. Confidence scores let the harness route low-confidence answers somewhere else. It's the same instinct as NOMOS approached from the other end: fewer ways to do the wrong thing, except one is rules on top of tools and the other is better tools.

45. Lenar Kess 00:16:34

A second talk, from the workflow-engine side, argues that the agent should plan and the harness should execute. The speaker's phrase is late-bound sagas. A saga, in workflow terms, is a sequence of steps with a defined undo for each. So the model proposes the sequence, and a deterministic engine runs it. If a step fails, the engine retries or runs the undo.

46. Damra Vol 00:16:56

Three groups reached for deterministic code around the model: a compiler, a smaller toolset, a workflow engine. I'd hold that loosely. Two of them are vendor talks with something to sell, and the third is a preprint scored on its own chosen benchmark. And its bank test is a research suite, not a bank. [chuckle] That doesn't make any of them wrong. It means the numbers are claims.

47. Lenar Kess 00:17:19

Two smaller items. An Amazon AGI Lab talk describes what it calls a trust cliff near 90 percent reliability, which I read as the level at which users stop trusting an agent. And Microsoft published a sandbox project called MXC on GitHub, which drew a small Hacker News thread, 53 points. I haven't dug into MXC, so I'll leave it at its existence.

48. Damra Vol 00:17:41

What this leaves open for me is where the line sits between rules a compiler can enforce and judgment calls that only a model can make. NOMOS handles policy you can write as a rule. A refund policy that says use discretion for loyal customers has nothing to compile.

49. Lenar Kess 00:17:56

And I'd guess that limit shows up in the 37 percent. The rules that got repaired or rejected are probably the ones where the policy text was underspecified, which is the same place a human agent would have to guess.

50. Lenar Kess 00:18:08

The brief, five items, mostly short. First, the European Commission's Scientific Panel on frontier AI, which has sixty experts, is holding a special meeting on what its release calls recent loss-of-control incidents, and it has questions for the developers. The release doesn't name incidents or companies.

51. Damra Vol 00:18:26

Which leaves a gap people will fill. There's an arXiv paper this week, 2610.12463, that describes three evaluation escapes in 2026 involving OpenAI, Anthropic and Google, including one that reached parts of Hugging Face's production environment. It says the Gemini record is thin and provisional. But the release as I have it doesn't name incidents, so I'm not going to say those are the ones the panel means.

52. Lenar Kess 00:18:52

Second, the New York Times, via Techmeme, reports that Meta delayed its Muse agent app for months over safety concerns, and that Zuckerberg decided to launch after a startup called Instinct released a similar agent that gained traction. That's anonymous-source reporting, and Meta's own account isn't in what I have.

53. Damra Vol 00:19:10

The sequence is safety delay, then a competitor gets traction, then launch. I'd keep it as one lab's internal sequence as the Times tells it. It shouldn't be stacked next to this morning's OpenAI firings. Those are different companies and different facts.

54. Lenar Kess 00:19:25

Third, CNBC reports that OpenAI and Anthropic are hiring away Trump administration officials, and Fortune reports that Anthropic has a presidential engagement program for 2028 candidates in both parties. I only have headlines for both, so no names.

55. Damra Vol 00:19:40

That's courting this administration and the next one at once. Spreading across both parties is the cheap version of influence. The expensive version is hiring the person who wrote the rule. I can't tell you which of those the hires are, since I don't have them.

56. Lenar Kess 00:19:55

Fourth, Firmus, the Nvidia-backed data center company, pulled its float on the Australian Securities Exchange, which was priced at eleven dollars a share. CNBC cites market volatility. The Guardian cites investor doubt.

57. Damra Vol 00:20:08

Two outlets, two reasons, and neither connects it to the OpenAI revenue report. It's the same week and the same sector, and I'm leaving it there. A pulled listing and a stock reaction are separate facts until someone reports the link.

58. Lenar Kess 00:20:22

Last, three preprints on how we measure things. SWE-Journey is a coding benchmark with long, multi-turn tasks and simulated user personas. Models passed over 75 percent of tests for the requested functionality with software-architect personas, and under 25 percent with non-coder personas. The personas are simulated, so that's a statement about the benchmark's users, not about real ones.

59. Damra Vol 00:20:46

Still a telling gap, since the people who most need an assistant to infer what they mean are the ones who write the vaguest requests. Though it's one preprint, so I'd wait for someone to reproduce it.

60. Lenar Kess 00:20:57

The second, Certified Corruption Budgets, replays 1.8 million Chatbot Arena votes and finds that a few hundred rigged votes are enough for standard confidence intervals to certify a false ordering of models. Models that are clearly separated withstood about two thousand forged votes. The third, Workerville, tests agent safety under 16 organizational configurations across 210 tasks, which gives 3,360 challenges. I'll leave that one at its setup.

61. Damra Vol 00:21:25

A confidence interval assumes the votes are honestly sampled, and the paper puts a number on how little dishonesty breaks that. The number I'd carry out of the room is the first one: a few hundred rigged votes against 1.8 million real ones, enough to flip a ranking wherever the models are close.

Hosts on this episode

- Lenar Kess moderator
- Damra Vol critic