

OpenAI withdraws three math papers

OpenAI revenue at \$50bn, not \$68bn, a free Anthropic bug scanner, Microsoft loses green cards, TypeScript ported to Rust

Friday, 9 October 2026

IN THIS ISSUE

1. OpenAI withdrew three math papers: a sign in a formula broke three proofs
2. Mathematicians against the release: "we did not ask for this"
3. OpenAI showed investors \$50 billion in annual revenue, and the market was disappointed
4. Anthropic starts a free vulnerability scanner for open source
5. The White House blocked Microsoft and seven other companies in the green card program
6. A Rust port of the TypeScript compiler: \$420 thousand with OpenAI and \$24 thousand with Opus
7. "Why is the industry not in panic about DeepSeek 4.1 Flash?"
8. GPT-6.1 Sol got an Ultrafast mode: 8 times faster
9. AI for science: \$150 million for the US government, an ultraviolet sky map and a planet
10. Whistle: speech recognition in 16.9 MB

TOPIC 1

OpenAI withdrew three math papers: a sign in a formula broke three proofs

SOURCES OpenAI, repository change history, 07.10 [GitHub](#) · Dan Roberts, @danintheory, 08.10 [@danintheory](#) · HN discussion, 08.10 [Hacker News](#) · Ethan Mollick on the extension of one of the results, 08.10 [@emollick](#)

Yesterday this digest reported that mathematicians were on their second day with OpenAI's 722 manuscripts. Now OpenAI itself admits the first serious error. A paper on the algebraicity of Weil classes on abelian varieties of dimension eight had a sign error. The error broke the argument that two terms cancel each other. Two more papers used the same construction. OpenAI withdrew all three: on Weil classes, on Kuga-Satake correspondences for K3 surfaces and on the rational Hodge conjecture for products of K3 surfaces. The last one concerned the same Hodge conjecture that supporters of the release listed among the Millennium Problems with clear progress.

The company corrected 14 more manuscripts. Among them are four papers on Lipschitz heights, six on Kähler minimal models and two on deformations. Another 13 papers got updated references to the corrected versions. The withdrawn texts stay in the archive with an explanation of the gap.

At the same time OpenAI added six new formalizations in Lean. By its count, Lean now verifies 300 of the 719 main results, about 42%. Dan Roberts of OpenAI promised to continue to publish formalizations and the list of errors. One reply under his post asks directly: why did you not check this before publication?

There is also an opposite signal. Doug Colkitt reported that a community member improved one of OpenAI's bounds: the constant κ went down from 2^{182} to 2^{15} . In his words, this is a 500 thousand times improvement over the previous step. Mollick sees this as a sign that some of the results already started fast collaborative work.

Why it matters. A withdrawal by itself is normal in science: people also make sign errors. The more important fact is different. The error was in a paper that two other papers depended on. Lean did not catch it, because these papers had no formalization. 42% of results with a formal check means that 58% rest only on text that almost nobody read. A person who quotes this release as "AI solved problem X" must now check two things: is there a formalization, and is the paper still not withdrawn.

TOPIC 2

Mathematicians against the release: "we did not ask for this"

SOURCES Association for Human Mathematics (AHM), 07.10 [Ahmath](#) · Asaf Karagila, 08.10 [Karagila](#) · Terence Tao on "Mathematics 2.0", 06.10 [Mathstodon](#) · HN discussion of Tao, 08.10 [Hacker News](#) · Ethan Mollick, 08.10 [@emollick](#)

The Association for Human Mathematics published a statement on the OpenAI release. Its main thesis:

"Mathematicians did not ask to do this work". AHM recalls advice from the math and AI advisory group that OpenAI cites. The group told labs not to test hard math problems on internal models. The statement calls the publication of over 700 files at once "a show of force, not scientific work". At the end AHM asks mathematicians to stop work with OpenAI. Mollick thinks that university courses about our era will read this text. Joscha Bach replied with a joke about an "association of organically grown mathematics".

Asaf Karagila, an expert on the axiom of choice, examined one specific paper. It proves that the partition principle does not follow from the axiom of choice. The problem is over a hundred years old, and Karagila promised a bottle of whisky to the person who solves it. OpenAI will not get the whisky. In his assessment, the preprint is "unclear and confused", and the terminology is "slightly wrong". Some references come from unpublished lecture notes. He compares the release to a denial-of-service attack.

Hundreds of hard texts demand that experts leave their own work and check them.

Terence Tao wrote his post on 06.10, before our window. It reached HN on 08.10 and got 593 points. Tao describes how a normal breakthrough works: talks, seminars, new collaborations, textbooks. In his words, AI operators now solve problems and cannot explain the result. Nobody can make a solved problem open again, thus the field becomes "less

fertile". Tao sees the way out in "Mathematics 2.0". It values explanation and community growth in addition to the first solution.

Why it matters. The dispute now is about who pays for verification. The question whether the proofs are correct is secondary. The lab gets the headline at once. The community, which did not order the work, gets the human work of verification. The same conflict exists in software development: an agent can open a hundred pull requests in one night, but there is still only one reviewer. Item 4 below shows how security people try to handle this.

TOPIC 3

OpenAI showed investors \$50 billion in annual revenue, and the market was disappointed

SOURCES CNBC, 08.10 [Cnbc](#) · Financial Times, 08.10 [FT](#) · HN discussion, 08.10 [Hacker News](#)

The FT reported it first, and CNBC confirmed it. In an investor presentation OpenAI gave an annual revenue run rate of about 50 billion dollars at the end of September. At the end of September the press widely quoted 68 billion dollars. A CNBC source says that the larger figure included the gross revenue of OpenAI's partners. That figure is easier to compare with Anthropic. The company also showed 77% growth of the revenue run rate in the third quarter and 107% in the enterprise segment.

The market read the difference as bad news. Nvidia shares fell 3%, Oracle almost 6%, CoreWeave almost 8%, AMD 5%. All these companies depend on OpenAI compute contracts. For comparison, in August Anthropic told investors about \$65 billion in annual revenue at the end-of-July run rate. According to CNBC, investors value OpenAI at \$852 billion. It filed IPO documents confidentially in June. It is also in early talks about a new round of about 30 billion dollars.

Why it matters. "Annual revenue run rate" is not an accounting term. Each company calculates it in its own way. The difference between \$50 and \$68 billion shows how wide this range is. Before the two largest IPOs in AI history, these figures will become the main argument of both sides. Each time, ask what the number includes: net revenue, gross revenue, or partner money.

TOPIC 4

Anthropic starts a free vulnerability scanner for open source

SOURCES Anthropic, Cyber Mission, 08.10 [Anthropic](#) · Anthropic, OSS Scanner, 08.10 [Anthropic](#) · [@AnthropicAI](#), 08.10 [@AnthropicAI](#) · Clem Delangue of Hugging Face, 08.10 [@ClementDelangue](#)

Anthropic announced Cyber Mission, a long program to protect the systems that everyone depends on. It starts with two directions. The first is the Critical Infrastructure Defense Program for power grids, water utilities and transport. Anthropic gives models, its own engineers on site and threat data to companies that already serve such operators. The

partners include CrowdStrike, Palo Alto Networks, Dragos, Rockwell Automation, Hitachi, Deloitte and PwC.

The second direction is OSS Scanner. It gives free regular checks of open projects with the company's strongest models, Claude Mythos among them. A project applies itself, and the selection rules are similar to Google OSS-Fuzz. The figures in the post are frank. In half a year the models found over 29 thousand vulnerability candidates, and people had time to check only about 6 thousand. At the request of maintainers, Anthropic already sent them almost 5 thousand unchecked reports. Thus the scanner reports will be fully model-generated, without a human check.

How reliable is this? Anthropic pentesters checked 97 critical and high findings in 48 projects. 85 of them (88%) passed the disclosure threshold, 11 were true duplicates, and only one was false. Maintainer feedback in the post: wolfSSL got 74 reports, 72 of them real, and 5 became CVEs. An OpenSSL representative writes that the reports are "as good as, and sometimes better than" reports from people.

Some maintainers complain that the scanner rates severity too high and does not understand the threat model of the project. All these figures and feedback come from Anthropic itself.

The same day Delangue of Hugging Face asked people to publish traces of attacks on AI agents and of their defense: "defenders cannot learn from what they cannot see".

Why it matters. The post names the bottleneck honestly: now it is easy to find a vulnerability, and it is hard to verify and fix it. Anthropic now moves verification to the maintainers. For a small project this can be harder than the vulnerability itself. The same figures give a useful practice for your own code. A report with a working reproduction and a patch takes minutes to verify. A report without them only adds work.

TOPIC 5

The White House blocked Microsoft and seven other companies in the green card program

SOURCES AP, 08.10 [Apnews](#) · BBC, 09.10 [BBC](#) · Semafor, 08.10 [Semafor](#) · Financial Times, 08.10 [FT](#) · HN discussion, 08.10 [Hacker News](#)

Vice President JD Vance accused Microsoft of fraud in the PERM program. In this program an employer proves to the Department of Labor that it found no American for a position. Then it can file a green card application for a foreign worker. Labor Secretary Keith Sonderling said that the government will not process new or current PERM applications from Microsoft, Adobe, Cognizant, Infosys, Tata, Wipro, HCL and Capgemini.

According to Vance, last year Microsoft laid off about 6 thousand Americans and filed over 6 thousand H-1B visa applications. AP and BBC give different application counts: 6,300 and 6,800. Microsoft replied that 80% of the applications were status extensions for workers who already work at the company.

According to the BBC, the median salary of an H-1B worker at Microsoft is \$155 thousand. Former USCIS adviser Doug Rand called the decision "incoherent". A green card frees a worker from dependence on the employer. Thus a block on green cards makes people even more dependent.

A few hours later Trump gave Microsoft CEO Satya Nadella the National Medal of Technology and Innovation. Elon Musk, Jensen Huang, Lisa Su, Sergey Brin and Michael Dell also got medals.

Why it matters. For engineers with an H-1B visa at these eight companies, the path to permanent status in the US is now closed. For the industry this is a signal that immigration programs became a tool of pressure on specific companies. The HN thread got 1,433 comments, and it shows how divided the industry is on this question.

TOPIC 6

A Rust port of the TypeScript compiler: \$420 thousand with OpenAI and \$24 thousand with Opus

SOURCES [ts-rust on GitHub](#), 08.10 [GitHub](#) · HN discussion, 08.10 [Hacker News](#)

Theo Browne of pingdotgg published ts-rust: a Rust port of the TypeScript compiler, type checker and language server. The port copies Microsoft's native compiler in Go, with the same tsc command. The author claims 100% compatibility on all real projects that he checked. He also writes: "I have not read a single line of this code" and "I have no idea if this will actually work".

The most interesting part of the README is the accounting. First the author ran OpenAI models, GPT-5.6 Sol and GPT-6 Astra, for several months. They wrote over 1.3 million lines of Rust and burned over \$400 thousand at API prices. They never got past 84% compatibility. Then he tried Opus 5.5. The model started from zero, had a working version in 10 hours and went further in two weeks. The cost was about \$24,047 at API prices, or from 925% to 983% of the weekly limit of a \$200 subscription.

Why it matters. A port with a reference and tests is the best task for agents. The reference behavior exists, and an automatic check can compare each answer. The 17 times cost difference between the two approaches shows that the choice of model for a specific task is more important than the token volume. The caveat is also important: "I did not read the code" means that only the tests carry the responsibility for correctness.

TOPIC 7

"Why is the industry not in panic about DeepSeek 4.1 Flash?"

SOURCES [dgt.is blog](#), 07.10 [Dgt](#) · HN discussion, 08.10 [Hacker News](#)

A developer worked intensively with DeepSeek 4.1 Flash for a month on about ten projects. His conclusion: in the middle of a session he cannot tell this model from Opus. With an

OpenCode Go subscription at \$10 a month, DeepSeek is almost unlimited for him. A full-day session rarely costs more than \$1. A task such as "sort the files on the desktop" costs \$0.003 instead of \$1.

The author explains the price with technology. DeepSeek made the KV cache about 437 times smaller than in its first model. This cache takes GPU memory in long coding sessions. He keeps Opus 5.5 for the final code review. It catches a few edge cases, and DeepSeek makes the fixes again. The post got 458 points and almost 400 comments on HN.

Why it matters. This is one source and a subjective experience. Thus check the quality claims on your own tasks [single source]. The work pattern is familiar and cheap: a cheap model does the work, an expensive model checks it. Yesterday this digest reported on the price war between Haiku 5.5 and GPT-6 Luna. The post shows that for some developers the "good enough" line is already even lower.

TOPIC 8

GPT-6.1 Sol got an Ultrafast mode: 8 times faster

SOURCES @OpenAIDevs, 08.10 @OpenAIDevs

OpenAI started an Ultrafast mode for GPT-6.1 Sol in the API, Codex and ChatGPT Work. The company says it gives intelligence close to Astra up to 8 times faster than standard Sol. The API price is \$12 per million input tokens and \$60 per million output tokens. In Codex and ChatGPT Work, the Pro 500 plan and some enterprise plans have access. An administrator must turn it on. The mode works with data residency in the US and the EU. As examples OpenAI names incident analysis, agents in applications and interactive services.

Why it matters. Speed changes how an agent works. With a slow model, ten short "test and fix" cycles can take longer than one large answer. The price is high, thus a sensible pattern is this: Ultrafast for interactive debugging, the normal mode for background tasks. The "8 times" figure comes from OpenAI itself [single source].

TOPIC 9

AI for science: \$150 million for the US government, an ultraviolet sky map and a planet

SOURCES Anthropic, Genesis Mission, 08.10 Anthropic · Anthropic, UV sky map, 08.10 Anthropic · The Rundown AI on the planet, 08.10 @TheRundownAI

Anthropic promised \$150 million over three years for Genesis Mission, a federal program to make science faster with AI. The money will pay for Claude access for over 15 agencies, among them NASA, NIH and NSF. Anthropic will give Claude, Claude Code and API credits to several hundred research projects. It named two priorities: fusion energy and quantum computing. It made the announcement at a White House science summit.

The same day the company showed an example. Astrophysicist Brice Ménard of Johns Hopkins University, who also works at Anthropic, assembled the first full map of the sky in ultraviolet. Before this, the main source was data from the GALEX mission for 2003-2013. That mission avoided bright stars and the plane of the Milky Way on purpose. Claude Science agents found all public UV surveys, aligned them with each other and filled the gaps. The model predicted about a third of the map and did not measure it.

Each pixel has a "measured" or "predicted" label and an error estimate.

Separately, Pavel Rabtsevich reported that he used Claude Code and Codex to find a planet in TESS telescope data. NASA software missed this planet. According to The Rundown, NASA plans to observe the star in November [single source, no NASA confirmation].

Why it matters. The map example shows well where AI helps science now: in postponed routine work that everyone thought necessary, but nobody had weeks for. An honest "predicted" label on each pixel is the correct standard. Without it, a user of the map cannot tell observation from model.

TOPIC 10

Whistle: speech recognition in 16.9 MB

SOURCES Cactus, 08.10 [Cactuscompute](#) · HN discussion, 08.10 [Hacker News](#)

Cactus released Whistle, an open speech recognition model for phones, wearables, robots and microcontrollers. It is one 16.9 MB file that runs on the CPU without dependencies. The model gives the first token in 11 ms. It transcribes up to 30 seconds of audio at a time in seven languages: English, German, French, Spanish, Italian, Dutch and Polish. Ukrainian is not available yet. In addition to text, the model gives the time of each word and a vector representation of the sound.

According to the author's table, Whistle is more accurate than Whisper base on LibriSpeech, SPGISpeech, Earnings-22 and on average on FLEURS. Whisper base wins on TED-LIUM, AMI and MLS, but its size is 145.3 MB. The model loads into the same engine as their small model Needle. Thus one binary turns voice directly into tool calls.

Why it matters. On-device speech recognition without a network solves two problems at once: latency and privacy. A size of 16.9 MB lets you put the model into an application without a separate download.

The benchmarks come from the author, thus test it yourself on your own language and noise.

misc

- **Step 5 Preview from StepFun** is on OpenRouter: a 600 billion parameter MoE with 27 billion active, \$1 per million input tokens and \$2.70 per million output tokens, with a

focus on agents and finance.

Openrouter

- **JetBrains released Mellum2.1**, an open model for agentic coding with stronger RL. The weights are in HF and GGUF formats (@nv_pavlichenko). [@nv_pavlichenko](#)
- **Perplexity published pplx-embed-v2-late** under the MIT license: ColBERT-like multimodal embeddings at 0.6 and 9 billion parameters. They search document images without OCR, and both models share one vector space. [Hugging Face](#)
- **TermGrade from ai&**: 1,004 terminal environments for agent training and all 36,144 attempts of six models, 14,234 failed attempts among them. Training on tasks that a model solves in about half of cases gave +3.1 on Terminal-Bench 2.1. [Aiand](#)
- **Harvard published the largest dataset of agent traces**: 12,002 sessions, 206 billion tokens, over 1.2 million tool calls in 16 weeks. [Hugging Face](#)
- **Let's Encrypt cuts certificate lifetime from 90 to 64 days** from 10.02.2027, with tests from 14.10.

In 2028 the standard will be 45 days. Manual certificate renewal has only a few months left. [Ars Technica](#)

- **@levelsio checked who opens his links**: thousands of bots visit the page within seconds after publication, among them a military intelligence firm and social media monitoring services.

@levelsio

- **Opus 5.5 against GPT-6 Astra in design**: Quesma gave both models one prompt for a 3D visualization of Calvino's "Invisible Cities". Astra finished in 53 minutes for about \$10. The author thinks the Opus result has the better design. [Quesma](#)
- **The FTC prepares an AI safety investigation**: according to Semafor, its requests to companies have dozens of questions about what they said about the dangers of their products.

Semafor