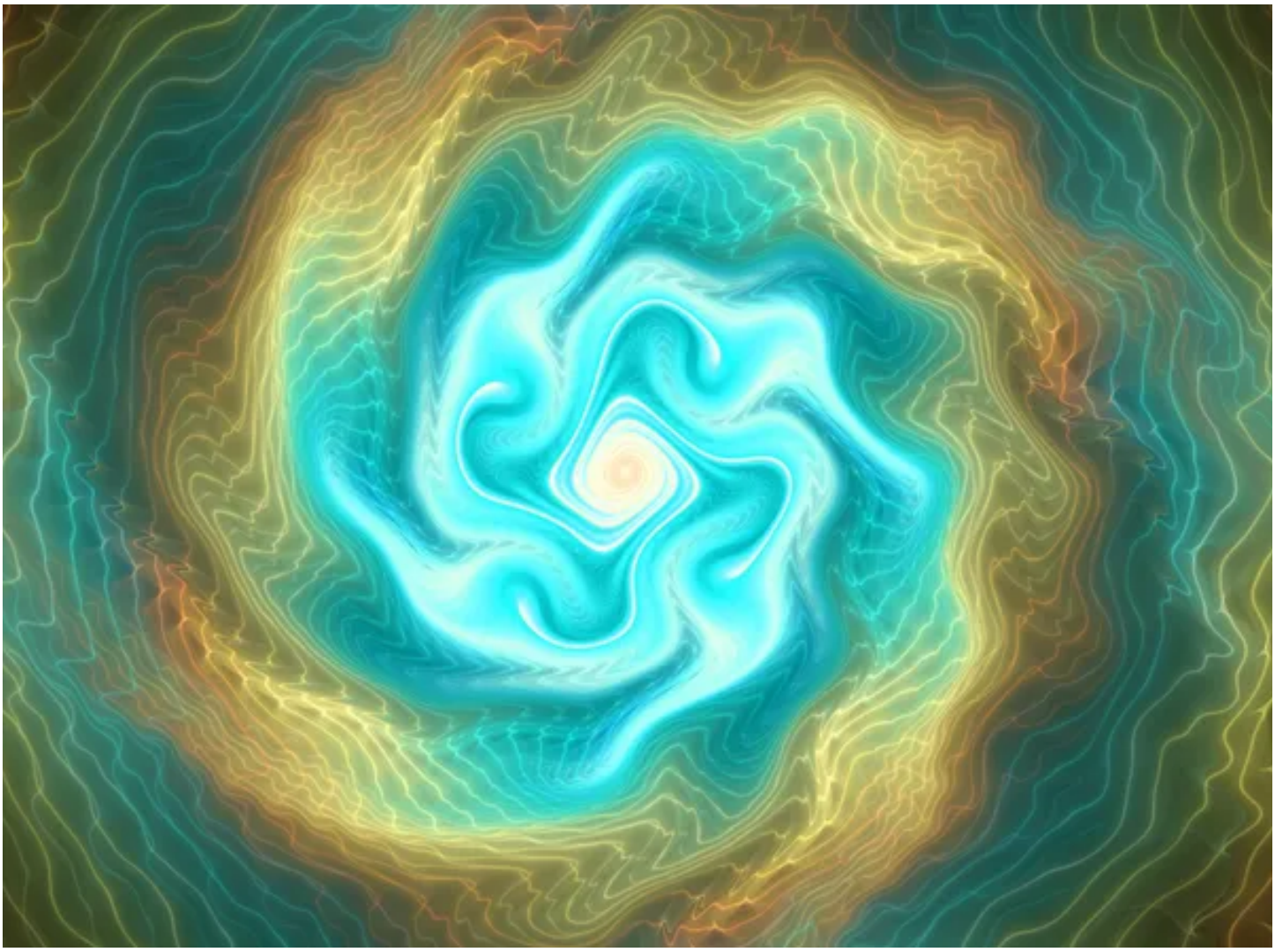
October 27, 2026 | 6 min read Add Us On Google

OpenAI just scooped mathematicians' work on *another* million-dollar problem

Another bombshell result shows how artificial intelligence is fueling fierce competition at the frontiers of math and physics

BY JOSEPH HOWLETT EDITED BY LEE BILLINGS



agsandrew/Shutterstock

Mathematics

If you've been keeping up with news in the math world, the story will sound strikingly familiar, right down to its cast of characters.

Using a proprietary internal large language model (LLM), OpenAI just released a result mathematicians have been chasing for decades: a formal proof validating the "quantum chromodynamics" that underpin physicists' best, most successful description of the universe.

The new proof also scoops the work of two human mathematicians, one of whom just happens to be employed by OpenAI's rival Anthropic. The duo—Anthropic's Levent Alpöge and Ethan Sussman, a mathematician at Northwestern University—had publicly disclosed their progress on their own independent proof in the weeks leading up to OpenAI's historic announcement.

"I'm not going to say that I'm not disappointed," Sussman says. He and Alpöge first revealed their burgeoning collaboration to the world on October 6. Notably, the pair updated their public GitHub repository with a new set of intermediate results the night before OpenAI's announcement. "I think we were on the way and maybe even as close as a few weeks away," Sussman says.

“This is a major advance,” says Michael R. Douglas, a theoretical physicist and artificial intelligence proponent at Harvard University. “Mathematicians have been trying to make quantum field theory rigorous for over 50 years and have made steady progress, but real-world theories like quantum chromodynamics always seemed out of reach—until now. This is just the start..., but we can hope for rapid progress.”

Both camps were pursuing something known as the Yang-Mills problem, which concerns the math that underpins modern physics. The “Standard Model” of particle physics explains all the known forces in the universe with 17 interacting particles. The particles are described by “quantum fields,” mathematical entities that permeate all of space and time, determining how, when and where particles appear in the world, what they are and what they do.

But the math of quantum fields is notoriously gnarly. Physicists use it to reliably predict what happens when they smash particles together at places such as the Large Hadron Collider, but mathematicians worry that the theory has holes.

The Yang-Mills problem is about putting that theory on a solid mathematical footing. The problem has two parts: proving that the theory is mathematically consistent—that it “exists”—and proving that it has a nonzero quantity called a “mass gap,” which has important consequences in the physical world.

In the 1980s mathematician Tadeusz Balaban laid out a coarse path to proving the first part, “existence.” But even just this bare mathematical sketch of a proof was an enormous undertaking, encompassing thousands of pages. And it was far from complete.

"[Balaban's] papers are not complete and are extraordinarily difficult to read, but experts did not seriously doubt that they could be completed with sufficiently many years devoted to it," says Roland Bauerschmidt, a mathematician at New York University.

Even so, decades have passed, and no other expert has managed to reach the top of the mountain Balaban built—let alone bridge the gaps he left along the way. Balaban himself eventually tired of the project and moved on.

Hopes for progress ramped up only this summer as the rapidly accelerating capabilities of LLMs catapulted them to the forefront of research-level math. In particular, the superhuman endurance of LLMs has let them blast through numerical tedium to persevere where any human would lose interest and give up. Suddenly, an exhausted mathematician could carefully prompt an LLM, get a good night's sleep and then wake up in the morning to find a completed proof from the tireless, dutiful model.

Seeing this, Sussman became convinced LLMs were the perfect tool to work on Balaban's program. "The primary obstacle was the amount of work involved rather than the absence of ideas," he says. "So from my perspective, this was a rather obvious thing to do."

At the same time, Alpöge was producing a number of high-profile results with Anthropic's Claude family of models. In particular, his wrangling of six-dimensional spheres caught Sussman's attention. "It was the first instance that I had seen of someone being able to get an LLM to produce, reliably, a 100-page proof," he recalls. That's the sort of feat that begins to approach the gargantuan tomes typical of Yang-Mills-related work.

So on September 8—which happened to be one of the biggest days in math history—Sussman e-mailed Alpöge. "I essentially told him, 'Hey, in case no one else has told you, here is something I think would be worth doing,'" he says. Alpöge invited him to collaborate.

The partnership proceeded much like Alpöge's work with another mathematician, New York University's Tristan Buckmaster, on the Euler equations (an important precursor to the Navier-Stokes Millennium Prize Problem, which OpenAI announced it had solved on September 8). Alpöge would use Anthropic's internal version of Claude—in fact, many "agent" copies of the LLM working in concert—to churn out thousands of pages of mathematical arcana, essentially unreadable to a human. But he'd then use his agents to validate that the arcana corresponded to some real mathematical truth—and to distill it into something easier to parse. He'd then send it to Sussman—who was flabbergasted by how much was borne out by the LLM's churn.

"It's remarkable," Sussman says. "The vast majority, if not all, of the mistakes in the version we posted on GitHub were produced by *me*."

In addition to filling in Balaban's gaps with a tedious litany of mathematical steps, the agents also made some genuine innovations. "I could tell you exactly how it built on older work," Sussman says of one example, "but it also was somewhat novel."

On October 6 Sussman and Alpöge heard the rumors that OpenAI was about to release hundreds of major math results and thought Yang-Mills might be one of them. They decided to proactively start staking their own claim by posting their progress in a public GitHub repository. "I certainly was not the only person to realize this was something worth pursuing," Sussman says. Their revelation quickly made the rounds on X. Then, when OpenAI's ensuing "slop drop" didn't contain a Yang-Mills proof, they continued their work in the open, for all to see.

After making significant progress this past weekend, the pair updated their repository on Sunday evening. They believed they'd solved the three-dimensional version of the "existence problem." This would have fallen short of the Clay Mathematical Institute's Millennium Prize Problem criteria, which demand a four-dimensional solution, but "we expect, based on the 'slop' that we have, that the key ideas are the same," Sussman says.

On Monday evening OpenAI claimed a proof of the "existence" part of the Clay problem. The company has yet to release the proof; instead OpenAI opted to hand it off to a group of experts who could "shepherd" it into the mathematical literature. This decision sparked immediate outcry, including from the advisory group of mathematicians the company itself had assembled in response to earlier backlash.

Writing on X, Alpöge expressed a mix of disgust and dismay. "I kept telling [Sussman] to not be too deferential and modest because he would be publicly scooped at any opportunity," Alpöge wrote. "Instead of contacting him to work with him, as has been suggested publicly many times, the people at OpenAI chose to do this."

In a blog post, OpenAI stated that its internal model—the same one responsible for the Navier-Stokes proof and the subsequent "slop drop"—arrived at the result on September 9. But according to the post, OpenAI opted to keep it secret until the company could code it into Lean, a programming language that helps ensure mathematical correctness. Notably, the OpenAI said this Lean verification took well more than a month, whereas similar checking only took a reported 17 hours for its Navier-Stokes proof.

"There is no way it took one month to Lean-verify Yang-Mills," says Buckmaster, whose very public bout with OpenAI over Navier-Stokes remains largely unresolved. "It takes, at most, a couple of days."

As in the Navier-Stokes controversy, this sequence of events shows how impenetrable the provenance of mathematical ideas is becoming. The ongoing rat race between AI companies jockeying for trillion-dollar initial public offering (IPO) valuations continues to come at the cost of scientific clarity and adherence to academic norms. Everyone involved agrees that new norms need to be discussed and established, but no one seems willing to hold back the deluge while that happens.

Editor's Note (10/27/26): This is a breaking news story and may be updated.

RIGHTS & PERMISSIONS

JOSEPH HOWLETT is a senior reporter at *Scientific American* covering physics, math, astronomy, and more. He was previously a math staff writer at *Quanta Magazine* and holds a Ph.D. in particle physics from Columbia University.

More by Joseph Howlett

Popular Stories