

# Mathematicians launch First Proof, a first-of-its-kind math exam for AI

 [scientificamerican.com/article/mathematicians-launch-first-proof-a-first-of-its-kind-math-exam-for-ai](https://scientificamerican.com/article/mathematicians-launch-first-proof-a-first-of-its-kind-math-exam-for-ai)

Joseph Howlett

February 9, 2026

February 9, 2026

3 min read

[Add Us On Google](#)

---

Mathematicians issue a major challenge to AI: Show us your work

Frustrated by the AI industry's claims of proving math results without offering transparency, a team of leading academics has proposed a better way



Alfred Gescheidt/Getty Images

The race is on to develop an artificial intelligence that can do pure mathematics, and top mathematicians just threw down the gauntlet with an exam of actual, unsolved problems that are relevant to their research. The team is giving AI systems a week to solve the problems.

The effort, called “First Proof,” is detailed in a [preprint](#) that was posted last Thursday.

“These are brand-new problems that cannot be found in any LLM’s [large language model’s] training data,” says Andrew Sutherland, a mathematician at the Massachusetts Institute of Technology, who was not involved with the new exam. “This seems like a much better experiment than any I have seen to date,” he adds, referring to the difficulty in testing how well AIs can do math.

The AI industry has become fixated on pure mathematics. Because mathematical proofs follow a checkable sequence of logical steps, their conclusion is true or false beyond any subjective measure. And that may offer a better way to compare LLMs’ prowess than evaluating how convincing their poetry is. Start-ups dedicated to AI for mathematics have recently recruited a number of high-profile mathematicians.

These efforts have had some early successes: In 2025 an advanced version of Google’s Gemini Deep Think achieved a gold-level score on the [International Mathematical Olympiad](#), an exam for prodigious high schoolers. And in the past few months, an AI has solved multiple “Erdős problems”—a trove of challenges set by the late mathematician Paul Erdős. The start-up Axiom Math made [headlines](#) last week for successfully tackling several research-level (though far from groundbreaking) math questions.

But none of these tests were controlled experiments. Olympiad problems aren’t research questions. And LLMs seem to have a tendency to find existing, forgotten proofs deep in the mathematical literature and to present them as original. One of Axiom Math’s [recent proofs](#), for example, turned out to be a misrepresented literature search result.

And some math results that have come from tech companies have raised eyebrows among academics for other reasons, says Daniel Spielman, a professor at Yale University and one of the experts behind the new challenge. “Almost all of the papers you see about people using LLMs are written by people at the companies that are producing the LLMs,” Spielman says. “It comes across as a bit of an advertisement.”

First Proof is an attempt to clear the smoke. To set the exam, 11 mathematical luminaries—including one Fields Medal winner—contributed math problems that had arisen in their research. The experts also uploaded proofs of the solutions but encrypted them. The answers will decrypt just before midnight on February 13.

None of the proofs is earth-shattering. They’re “lemmas,” a word mathematicians use to describe the myriad of tiny theorems they prove on the path to a more significant result. Lemmas aren’t typically published as stand-alone papers.

But if an AI were to solve these lemmas, it would demonstrate what many mathematicians see as the technology's near-term potential: a helpful tool to speed up the more tedious parts of math research.

"I think the greatest impact AI is going to have this year on mathematics is not by solving big open problems but through its penetration into the day-to-day lives of working mathematicians, which mostly has not happened yet," Sutherland says. "This may be the year when a lot more people start paying attention."