

AI uncovers solutions to Erdős problems, moving closer to transforming math

 scientificamerican.com/article/ai-uncovers-solutions-to-erdos-problems-moving-closer-to-transforming-math

Joseph Howlett

February 12, 2026

February 12, 2026

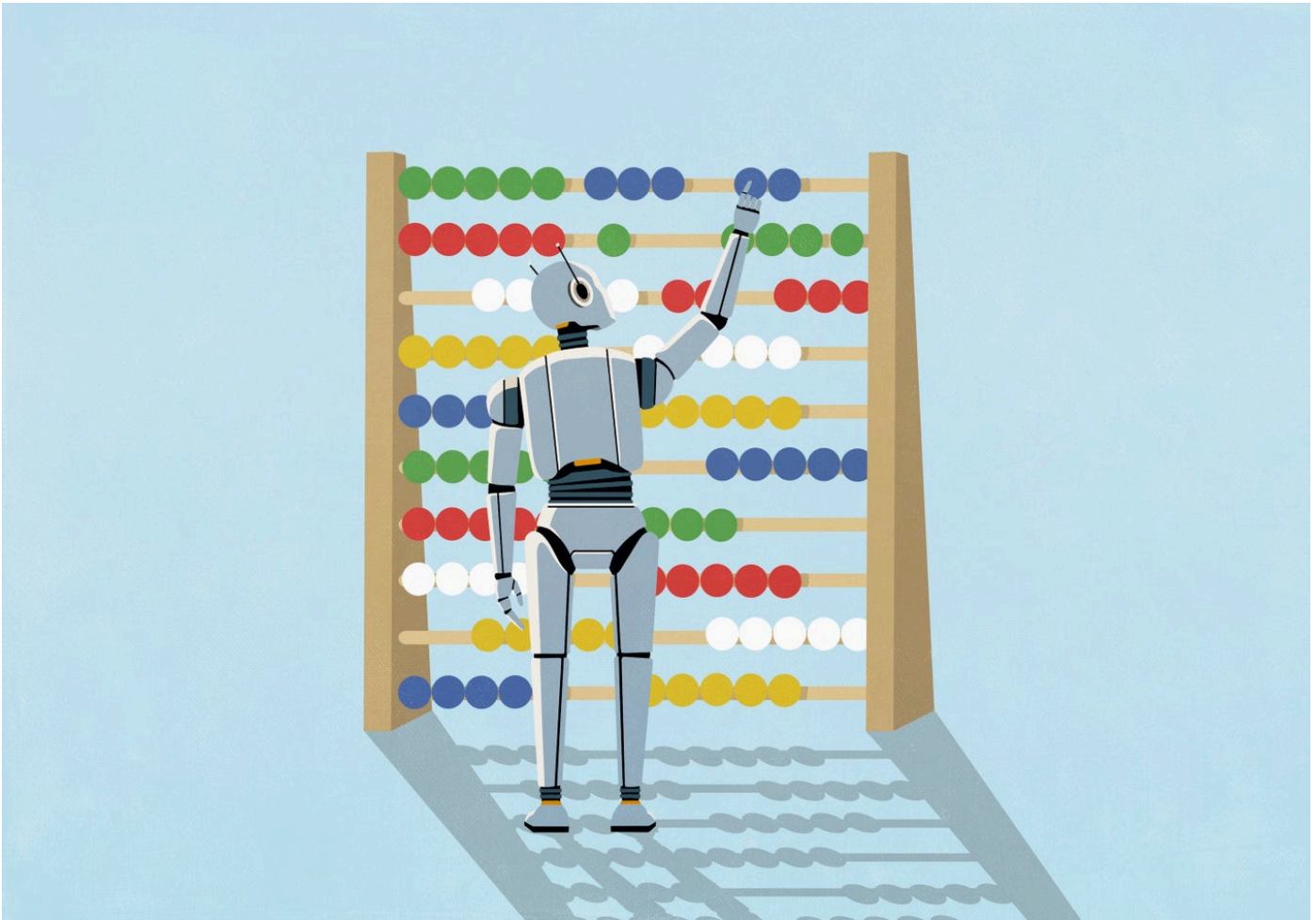
6 min read



[Add Us On Google](#)

Is AI on the precipice of revolutionizing math? It depends

LLMs have recently helped find solutions to a number of minor longstanding problems. But a new plan called First Proof is really putting them to the test



Malte Mueller/Getty Images

One idle evening last October, Mehtaab Sawhney took up an old pastime. He began perusing the website erdosproblems.com, an updated record of the 1,179 conjectures left behind by the eccentric and indefatigable 20th-century mathematician [Paul Erdős](#).

Sawhney, a mathematician at Columbia University, had always been interested in the Erdős problems, which range from minor curiosities to central open problems in number theory and combinatorics.

He came upon a problem, #339, that seemed too straightforward to still be “open” nearly two decades after Erdős’s death. He’d seen similar conjectures before. “There were a number of problems that kind of looked *too* approachable,” Sawhney says. In the past, he’d turned to Google. “And then eventually, with enough searching, I would find a reference to a solution.”

But recently he’d been playing with ChatGPT as a new way to check the literature. “I decided to plug it in, and then it just told me there was a reference,” Sawhney says.

It went so well that he reached out to a fellow mathematician, Mark Sellke, who had recently gone on leave from an academic position to work for OpenAI. Together they prompted ChatGPT to dig up lost solutions to nine other Erdős problems, plus partial solutions to 11 more.

Since then, the website's activity has skyrocketed. According to a webpage started by the mathematician Terence Tao, [AI tools have helped transfer about 100 Erdős problems into the "solved" column](#) since October. The bulk of this assistance has been a kind of souped-up literature search, as it was with Sawhney's initial success. But in many cases, LLMs have pieced together extant theorems—often in dialogue with their mathematician prompters—to form new or improved solutions to these niche problems. In at least two cases, an LLM was even able to construct an original and valid proof to one that had never been solved, with little input from a human.

The story of the Erdős problems is just part of a sea change that has taken place over the past few months. LLMs have become unrivaled in their ability to scour and synthesize the literature on any mathematical topic, however esoteric. They can also guide working mathematicians, helping them sketch a path to proving a larger result and proving small chunks of it to save time. This assistance is often misguided and riddled with holes that require expert eyes to suss out. But mathematicians can see its potential.

"They are now useful research assistants," says Andrew Sutherland, a mathematician at the Massachusetts Institute of Technology. "Mathematicians whose only experience with LLMs is with earlier models don't yet fully appreciate this."

AI is still nowhere near being able to solve major open problems in math, let alone replace mathematicians. Despite widespread anxieties voiced by graduate students during conference coffee breaks and in online message boards, no major mathematics journal has published a peer-reviewed proof citing the use of LLMs. But that, at least, could change this year.

Assessing the State of Things

Erdős problems are a useful LLM "benchmark" because there are so many of them. And they've proved a distinctive showcase for the technology's burgeoning strength as a mathematical search engine.

"Erdős problems sort of fit in a category of their own," Sutherland says. "For the most part, they're individual problems whose solution is not necessarily going to have any broader implications." As a result, solving a more obscure Erdős problem is a feat that often goes unnoticed. It's rarely worth submitting to a journal and rarely cited in subsequent work.

None of that matters to an LLM. It can easily unearth preprint papers unknown even to experts—proofs that sometimes don't reference Erdős at all. Google's Gemini found an offhand remark deep in a paper from 1981 that unknowingly solved Erdős problem #1089. But more surprising is LLMs' ability to make meaningful mathematical suggestions.

"I think it's a mistake to say it's 'just a search engine,'" Sutherland says. "I've had one or two interactions where it actually pointed me to a result that let me prove something I was stuck on."

Similar experiences motivated the team behind [First Proof](#), a fresh attempt to test AI's math skills. Eleven top mathematicians picked discrete chunks of proofs they have completed but not yet published and posed them as a challenge to AI last Thursday. The problems cover a wide range of areas and vary in complexity. "A system that could resolve all of them would be very useful for a professional mathematician," says Daniel Litt, a mathematician at the University of Toronto.

The team is giving LLMs until Friday to produce proofs of the 10 problems. The one-week time limit was chosen carefully, according to Lauren Williams, a Harvard University mathematician on the First Proof team. It's less time than her own problem took her and a coauthor to prove, so likely not long enough for human mathematicians without AI assistance.

By Monday the e-mails and social media pages of Williams and her collaborators were inundated with claimed solutions. "There's a lot of excitement, which is really great to see," she says. A Discord server hosting discussions on the challenge has quickly garnered hundreds of members, many carrying purported proofs from ChatGPT and other LLMs.

Familiar troubles have already arisen. First Proof was meant to be more than a literature search—the team tested its questions on LLMs to be sure no answers existed in their training data. But pretty quickly an online solution surfaced to a problem from Martin Hairer, winner of a 2014 Fields Medal, math's highest honor—and one of the First Proof team members. When he picked the problem, he had overlooked a partial proof in the bowels of his personal website that was archived by the Wayback Machine.

And contestants lacking the team's expertise in these particular mathematical niches aren't sure what to do with the deluge of confident claims their LLMs keep spitting out—it's up to the First Proof team to check every submission. "Verification is a problem because 90 percent of the time it will come up with a solution," Williams says. "It's going to write something and sound confident about it."

Litt has glanced over many of the "proofs" circulating this week and found them to be largely bogus—although he's seen a few that may be correct. "It's absolutely very impressive that the models are sometimes able to generate correct answers to some of the problems," he says. "But they're generating a huge amount of garbage." Even by Saturday, it may not be clear whether the LLMs have won or lost.

A Pivotal Year

Regardless of the First Proof outcome, the last month has brought many signs that LLMs will soon be part of many mathematicians' tool chests.

In January Ravi Vakil, current president of the American Mathematical Society, posted a preprint with two other mathematicians and two researchers from Google [in which they collaborated to solve a math problem](#) that bears on his research. The authors document how Google's LLM

helped them get to a proof. “It really did lead us to new ideas,” says Vakil, who wanted to “get a sense of how mathematicians should reasonably be doing math in five years.”

Still, LLMs have yet to contribute a proof that would create buzz if it came from a human. “Every individual result has been vastly overhyped by certain corners of the Internet,” Litt says. Carlo Pagano, who collaborated with Google’s DeepMind team to [work on several Erdős problems using Gemini](#) in research posted as a preprint, is also hoping for a more substantial benchmark. “The Erdős problems are not great in some sense,” he says. “It’s important to do this also on problems that we know are of broader interest.”

But several mathematicians predicted that 2026 will be the year where results of this type, in which AI is a stated contributor, first make it through peer review in major mathematics journals.

“I think it’s going to change the subject,” Sawhney says. “And that’s a really exciting thing.” Given that change, Sawhney has taken an academic leave from Columbia to work for OpenAI. This week Pagano started a joint position at Google DeepMind. “It’s clear that this will change how we do math,” he says, “so better to start early rather than later.”