

Democracy Dies in Darkness

For centuries, math has been a human endeavor. Does AI mean it's over?

Math illuminates how traffic flows, how our cells build proteins and even how to speed up medical imaging scans. Some worry the academic discipline now faces an existential threat.

Today at 5:00 a.m. EDT



By Carolyn Y. Johnson

This month, 30 mathematicians gathered at Harvard University to grade a consequential and difficult problem set. This wasn't just any test; its outcome would guide a field grappling with existential questions about its future.

The student? AI.

The leaders of the First Proof effort announced Wednesday that across four AI systems, seven of 10 problems got a passing grade.

It has been a head-spinning month for math, the latest area of human endeavor to find itself provoked, challenged and put on the defensive by large language models.

OpenAI announced in May that an internal model had disproved a conjecture that had languished on math's long list of unsolved problems for 80 years. Then, mathematicians unveiled the Leiden Declaration, an international manifesto with more than 2,300 signatories that confronts the potential and threat of AI — putting forth guidelines on how to use the technology ethically and transparently. The new First Proof results are the second round of a mathematician-led effort to ground truth the technology's strengths and weaknesses.

Mathematicians often work on arcane problems many of us don't comprehend. But math can illuminate how traffic flows, how our cells build proteins and even how to speed up medical imaging scans. Is AI an existential threat to math or an impressive tool? Something in between? Mathematicians are finding themselves arguing for the soul of their profession.

“There’s some misconception that ... there’s a box of conjectures out there and that all mathematicians do is pull a conjecture from a box and try to prove it,” said Lauren Williams, a math professor at Harvard and part of the team behind First Proof.

AI’s capabilities are genuine and startling to many experts. But, they also point out, math relies on human taste, judgment and intuition — choosing the problems to solve, framing those dilemmas and situating them in a larger context. Those tasks haven’t been mastered by AI — yet.

“What’s the interesting question to ask? You know, there’s a lot of stupid questions you could ask. A geologist could ask, ‘what’s the average color of a rock on earth,’ right?” Williams said. “That’s a question. It’s probably not an interesting question. So scientists have to figure out what are the fruitful directions to follow?”

An 80-year-old puzzle

In mid-May, OpenAI announced that its model had disproved a conjecture proposed in 1946 by famed mathematician Paul Erdős, a prolific problem-poser whose hundreds of still-unsolved problems have become a kind of proving ground for the AI industry. This problem was deceptively straightforward: If you place points in a single plane or flat surface, how many pairs can be separated by the same distance?

Erdős put forth a possible solution and offered a \$300 prize to whoever could prove or disprove it. He upped the ante to \$500 in 1995.

Eight decades after the puzzle was first proposed, it was a computer mind and not a human one that furnished what Princeton mathematician Noga Alon called a “spectacular solution.”

“AI is helping us to more fully explore the cathedral of mathematics,” wrote Thomas Bloom, a research mathematician at the University of Manchester.

Sébastien Bubeck, a mathematician and researcher at OpenAI, said the solution emerged from a general model that wasn’t fine-tuned to solve math problems, without human input. He marveled at the accomplishment when comparing it to his own effort to solve similar problems. (The Washington Post has a content partnership with OpenAI.)

“Sometimes, in some proofs I did in the past, it’s so taxing, mentally. You have to keep so many things in mind and make sure everything fits just right together,” Bubeck said. “AI definitely has an advantage to [do] this. It can make a more sophisticated construction.”

But he was quick to add that this was not the end of human math.

“The weakness I currently see in these models is understanding: Why it is that we’re doing what we’re doing?” Bubeck said. “I’m trying to not only solve this problem ... but it’s part of a broader program. And these models don’t have broader agendas.”

The job of being a mathematician may evolve, said Rodrigo Ochigame, an anthropologist and historian of computing at Leiden University and one of the authors of the declaration, but it won’t vanish.

“Perhaps there will be a greater emphasis on setting research directions, creating new techniques and definitions, cultivating understanding and insight, judging the depth and significance of ideas, and connecting particular problems to larger questions in mathematics and in the wider world,” Ochigame wrote in an email.

Four AI teams, 10 problems

As AI companies let their models loose on math's unsolved problems, many human mathematicians have found themselves intrigued but also skeptical. They see a commercial technology that often doesn't give credit to where its ideas come from. Companies trumpet successes, but in their own hands, these tools perform unevenly and checking their work is not trivial. They want a more transparent accounting of how the models work and how they can contribute to the field.

"Of course, everybody's impressed, that's the easy question," said Martin Hairer, a mathematician at EPFL, or the Swiss Federal Technology Institute of Lausanne, and Imperial College London and a Fields medalist. "It doesn't always do it right, and it's actually quite a lot of effort to convince yourself it's right. It doesn't write in a way that we write, it somehow doesn't write in an honest way."

The First Proof project was launched to try to regain control of the narrative being spun by tech companies.

To benchmark AI's abilities, they took problems from across math that had been privately solved but never published. In a first batch earlier this year, AI solved somewhere around six to eight of the 10 problems, Williams said, though they didn't do any formal grading. The teams fine-tuned their experiment for a second batch of problems, with the results announced last week.

"I didn't know what the outcome would be — this is a very clean experiment," Nikhil Srivastava, a mathematician at the University of California at Berkeley, said at a webinar before the results were announced.

The results were illuminating — seven out of 10 problems had at least one correct solution from the four AI systems tested. A few of the solutions were "flawless," others required minor revisions and some were just failures. In one case, the model used a different strategy than humans had "and impressed the referees," the [First Proof authors wrote](#).

"The results show genuine capability," the First Proof editorial board wrote in a statement. "Some solutions were correct, complete, and novel, while others exhibited systematic weaknesses that are useful for the research community to understand."

Terry Tao, a Fields medalist at the University of California at Los Angeles, was on a team that created an "AI harness" — software built around the AI to allow it to use a tool and take the next steps in reasoning. Tao said the models keep improving, but he drew a contrast between climbing and jumping.