FAIR LABS FRONTIER SERIES

SPECIAL REPORT
SEPTEMBER 2026

The Proof Machines

Machines are now closing real problems in mathematics, and a Millennium Prize problem is among the claims. What has actually been proved, what has been retracted, and why the discipline is arguing about its own future.

PREPARED FOR RESEARCH LEADERS · FUNDERS · UNIVERSITIES · POLICYMAKERS · EDITORS

FAIR Labs
FAIR ARTIFICIAL INTELLIGENCE RESEARCH LABS

September 2026
FL-2026-SR4

FAIR LABS FRONTIER SERIES · SPECIAL REPORT

The Proof Machines: machines are now closing real problems in mathematics, and a Millennium Prize problem is among the claims. What has actually been proved, what has been retracted, and why the discipline is arguing about its own future.

In the twelve months to September 2026, artificial intelligence stopped being a curiosity in mathematics and became a participant. A machine-written counterexample disproved a famous conjecture of Erdős and was checked line by line by nine mathematicians including a Fields Medallist. A model improved a fifty-year-old bound on the Riemann zeta function by twenty-six percentage points. Another wrote thirteen million lines of Lean in eleven days and produced the first end-to-end machine-checked proof of Fermat's Last Theorem. On 8 September 2026, OpenAI announced a resolution of one of the four statements the Clay Mathematics Institute accepts as settling the Navier–Stokes Millennium Prize problem, and declined to claim the prize. This Special Report separates what has been independently verified from what has been announced, sets both against a documented record of claims that were withdrawn, and takes seriously the objection that the mathematicians themselves are raising: that the field's supply of open problems is finite, that credit and responsibility cannot be delegated to a system, and that a proof no human has read is not yet knowledge.

REPORT CODE	FL-2026-SR4
PUBLICATION DATE	September 2026
PRIMARY AUDIENCES	Research Leaders · Funders · Universities · Policymakers · Editors
SERIES	FAIR Labs Frontier Series, Fair Artificial Intelligence Research Labs
SUGGESTED CITATION	*The Proof Machines.* FAIR Labs Frontier Series, Special Report. Fair Artificial Intelligence Research Labs, 2026.
KEYWORDS	automated theorem proving, Lean, mathlib, formalization, Millennium Prize Problems, Navier–Stokes, Riemann hypothesis, Fermat's Last Theorem, Erdős problems, AlphaProof, AlphaEvolve, research integrity, peer review

This Special Report is a rapid-response analysis of publicly reported events as of 11 September 2026. Every quantitative claim is attributed in Sources & Notes to a primary source, meaning the lab's own announcement, the paper, the repository, the competition's official page, or the named mathematician's own writing. Where a result is self-reported by its author and not independently verified, this report says so on the spot. Where a claim was later corrected or withdrawn, the correction is reported alongside the claim. This report is for informational purposes and does not constitute legal, financial, or investment advice.

IN BRIEF · SPECIAL REPORT

The machines are proving things. The hard part is now telling which things.

Something real happened in mathematics this year, and it is being described badly in both directions. The triumphalist account, that AI has solved a million-dollar problem, is too fast. The dismissive account, that this is pattern-matching over a training set, stopped being true somewhere around May. The honest position is narrower and more interesting: a small number of machine-originated results have survived expert scrutiny, a much larger number of announcements have not, and the discipline's bottleneck has moved from finding proofs to checking them.

01 **The verified core is small, and it is real.**

An AI-generated counterexample disproved the Erdős unit distance conjecture and was rewritten and verified by nine named mathematicians, Gowers among them. That is not a benchmark score. That is a theorem.

02 **Formalization is where the machines are strongest.**

Fermat's Last Theorem was formalized end to end in eleven days, machine-checked, with no unproved assumptions. The mathematician who has spent two years on the same target confirmed it compiles, and said it adds no mathematics.

03 **The retraction record is part of the evidence.**

The single most publicized claim of the period, ten Erdős problems solved, was a literature search misread as discovery and was withdrawn within a day. DeepMind's own paper reclassifies nine of its thirteen results the same way.

04 **A Millennium claim exists, and the prize rules do not care yet.**

Clay requires publication in a refereed outlet and two further years of survival. No prize can be awarded before late 2028 at the earliest, and OpenAI has said it will not claim one.

05 **The hardest problems remain untouched.**

No system has claimed progress on Yang–Mills, Hodge, Birch and Swinnerton-Dyer, or P versus NP. Prediction markets price a Millennium solve at 82 percent by the end of 2027 against that silence.

06 **Mathematicians are asking for rules, not a moratorium.**

Nearly four thousand of them, including Scholze, Tao and Buzzard, signed a declaration holding that credit and responsibility belong to humans. The argument is about attribution and verification, not capability.

CONTENTS

Inside this report

01

Eighty-eight hours in September

On a Tuesday three days before this report was written, a laboratory announced that a machine had settled part of a problem that has resisted mathematicians since 1934. The announcement also declined the million dollars, which is the detail worth starting from.

On 8 September 2026, OpenAI published a claim that an internal system had resolved the Navier–Stokes Millennium Prize problem by establishing what the official problem statement labels statement (C), and also (D).[1] The run is documented in unusual detail: the effort was launched on 1 September, the resolution arrived on 5 September after roughly eighty-eight hours, and a formal proof in Lean was completed seventeen hours later.[1, 2] The consumption figure is the one that lodges in the memory: about 130 billion output tokens.[1]

One clarification before anything else, because the coverage keeps collapsing it. The prize dates from 2000; the problem does not. The Navier–Stokes equations were written down in the nineteenth century, and the Clay Institute's own problem page says so, adding that "our understanding of them remains minimal."[3] The modern existence-and-smoothness question descends from Jean Leray's 1934 paper, which proved that weak solutions exist and left open whether they stay smooth; Fefferman's official problem statement takes Leray as its starting point.[4, 5] Progress since has been partial and slow: Scheffer in 1976, Caffarelli, Kohn and Nirenberg in 1982, Lin in 1998.[5] What happened in 2000 was that a million dollars was attached to a question that had already defeated the field for two-thirds of a century. Ninety-two years, not twenty-six, is the span a machine is being measured against.

Two things about that claim need saying immediately, and they pull in opposite directions.

The first is that statement (C) is not a lesser prize. Charles Fefferman's official Clay problem description offers four statements and says plainly that the Institute asks for a proof of one of them, framing the choice as giving "reasonable leeway to solvers while retaining the heart of the problem."[5] Statement (C) asks for breakdown: a smooth initial velocity field and a smooth forcing term for which no smooth solution exists for all time. It is on the list. Nothing in the official text ranks it below the others.

The second is that OpenAI itself wrote, in the same announcement, "We do not intend to claim the Millennium Prize for this result."[1] A laboratory that believed it had cleanly won a million dollars would claim the million dollars. The disclaimer is the most informative sentence in the document, and the rest of this report is in part an attempt to explain it.

88 hrs
from launch to claimed resolution, 1 to 5 September 2026

130B
output tokens consumed on the Navier–Stokes effort

~2 yrs
minimum wait before any Millennium Prize can be awarded, after publication

6
Millennium problems still open; only Poincaré has ever been settled

1.1 The credit fight that arrived with it

Within hours the story stopped being about fluid dynamics. Tristan Buckmaster of NYU and Levent Alpöge of Anthropic had been working on the forced Euler problem, and Buckmaster alleged that OpenAI had learned of their approach and adopted the forcing method.[6] Sébastien Bubeck of OpenAI responded that "We did not use their prompt or proofs to prompt our models."[6] Sam Altman said the approaches "appear to be different" once both were visible, and described the origin of the effort candidly: "We were curious if ours could do it too."[7] OpenAI also conceded something that deserves more attention than it received: "While unlikely, we cannot rule out that de-identified data derived from their usage of our products helped improve our models."[7] Bubeck later said, "we recognize the priority of Levent Alpöge and Tristan Buckmaster's work."[8]

Charles Fefferman, who wrote the official problem statement, told *Quanta* that "The heroes of the story... are Córdoba and Martínez-Zoroa," naming the human mathematicians whose prior work the result builds on.[9] Terence Tao, writing the day before the OpenAI announcement about the Alpöge–Buckmaster result, noted that it too "has also been formalized in Lean" and that "As one may expect nowadays, the arguments here are heavily AI-assisted," adding that the authors were "forced to release their preliminary preprints before they were completely digested and polished, due to external events."[10]

> **THE SHAPE OF THE PROBLEM**
>
> Notice what the dispute is not about. Nobody is arguing that the machine failed to produce a valid argument. The argument is about priority, attribution, and whether a laboratory with the world's largest telemetry footprint can credibly assert independence from work its own users were doing on its own products. Those are questions about the sociology of research, not about capability. That is the change: capability is no longer the interesting variable.

02

The ledger: what has actually been established

Strip out the announcements and keep only the results that named mathematicians have checked and stood behind. The list is short. It is also unambiguous, and a year ago it was empty.

The strongest single item is the disproof of the Erdős unit distance conjecture. In May 2026, a counter-example generated in one shot by an internal OpenAI model was written up in a paper titled, with deliberate modesty, "Remarks on the disproof of the unit distance conjecture."[11] The author list is the point: Noga Alon, Thomas Bloom, W. T. Gowers, Daniel Litt, Will Sawin, Arul Shankar, Jacob Tsimerman, Victor Wang and Melanie Matchett Wood. Nine mathematicians, one of them a Fields Medallist, produced what they describe as a "short, digested, human-verified version" of the machine's construction.[11] Gil Kalai compared the moment to the computer-assisted proof of the four colour theorem in 1976.[12] Gowers said that had a human submitted the paper to the *Annals of Mathematics*, "I would have recommended acceptance," and Tsimerman called the construction "definitely an intimidating" one.[13] The AI is credited as the source of the idea. It is not an author.

The second item is the Riemann zeta bound. In August 2026 Anthropic reported that an unreleased research version of Claude had improved the known lower bound on the proportion of zeta zeros lying on the critical line from 41.6 percent to 67.2 percent, across two sessions consuming 31 million output tokens.[14] The company's own framing was careful: "We don't expect that the techniques Claude used will lead to proving the Riemann hypothesis."[14] James Maynard, who works on exactly this material, gave the assessment that matters: "The problem was in need of a new real idea, which this new result seems to provide. It seems that the AI has made a genuinely interesting mathematical contribution."[15] He also said, in the same piece, that "there is no pathway for any of these approaches to deal with the actual Riemann hypothesis."[15] Both sentences are true and they belong together; the second is what makes the first credible.

The third item is optimisation rather than proof, and it is older. AlphaEvolve found a way to multiply two four-by-four complex matrices in 48 scalar multiplications, improving on a bound that had stood since Strassen in 1969, and produced a configuration of 593 spheres in eleven dimensions, one better than the previous kissing-number record.[16] Ernest Davis's critique is the necessary counterweight: the matrix result is over complex values, which "would seem to significantly diminish the practical significance," and 593 against 592 is a 0.36 percent improvement in the known range.[17] Real, narrow, and oversold in the retelling.

EXHIBIT S.1 · INDEPENDENTLY VERIFIED

What survived expert scrutiny, and who did the scrutinising

RESULT	SYSTEM	DATE	VERIFICATION
Erdős unit distance conjecture disproved	Internal OpenAI model	May 2026	Rewritten and verified by nine named mathematicians including Gowers; published as a joint human paper [11]
Riemann zeta critical-line bound, 41.6% to 67.2%	Unreleased Claude	Aug 2026	Examined by Conrey, Goldston, Alpöge and Furman; Maynard on the record calling it a genuine contribution [14, 15]
Fermat's Last Theorem, formalized end to end	Claude	Sep 2026	Compiled and checked by Kevin Buzzard; kernel-checked over 1,052,234 declarations; no unproved assumptions [18, 19]
IMO 2025, 35 of 42, gold standard	Gemini Deep Think	Jul 2025	The only AI olympiad result officially graded and certified by IMO coordinators [20]
4×4 complex matrix product in 48 multiplications	AlphaEvolve	May 2025	Verified by construction; significance contested by Davis [16, 17]
Cap set of size 512 in dimension 8	FunSearch	Dec 2023	Peer-reviewed in *Nature*, co-authored by the mathematician Jordan Ellenberg [21]

Included only where a named mathematician outside the originating laboratory has examined the result and said so publicly, or where the proof is machine-checked against a formal kernel. Benchmark scores are excluded by construction: they are not results.

2.1 The one number that measures the field honestly

Every laboratory can pick its own examples. The First Proof Project cannot be picked. It is an independent evaluation with an editorial board of Mohammed Abouzaid, Nikhil Srivastava, Rachel Ward and Lauren Williams, which takes unpublished research problems arising from contributors' own work, pre-registers them, and has expert referees grade the submissions double-blind.[22] In its second batch, four systems produced 39 solutions to ten problems, graded by around thirty referees at Harvard in June 2026. Seven of the ten problems received at least one passing grade, meaning a solution judged essentially flawless or needing only minor revision.[22]

Seven in ten, on problems chosen so the machines could not have seen them, graded blind by people who had no stake in the answer. That is the most defensible number in this report. The referees also recorded a characteristic failure mode that anyone commissioning this work should memorise: routine steps handled meticulously, and the one critical step waved through as a standard argument. They further flagged uncited borrowing from the problem-setters' own prior papers that “would constitut[e] plagiarism if submitted by humans.”[22]

03

The formalization turn

The most consequential development of the year is not a new theorem. It is that machines have become very good at translating mathematics into a language a computer can check, which changes what "verified" means.

On 4 September 2026, Anthropic published what it described as the first end-to-end, computer-checked proof of Fermat's Last Theorem, produced by Claude in eleven days working largely autonomously.[23] The figures: about thirteen million lines of Lean, 30,300 theorems proved of which 29,500 are used in the final argument, and roughly six billion output tokens.[23] The public repository records 29,511 theorems, and an axiom check yielding only the three axioms that constitute ordinary mathematics, with no sorry and no escape hatches.[18] An independent kernel written in Rust checked 1,052,234 declarations.[18]

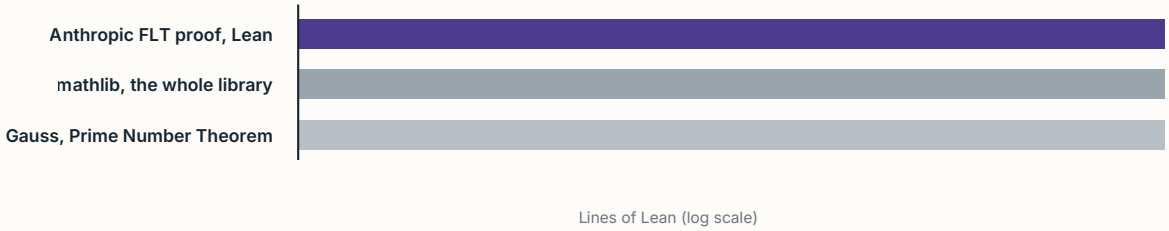
Kevin Buzzard has been leading a formalization of Fermat's Last Theorem at Imperial College since October 2024, funded by an EPSRC grant that runs to September 2029, whose stated goal is the narrower one of reducing the theorem to claims known by the end of the 1980s.[24] He posted the same day under the headline "FLT: Anthropic has beaten me to it."[19] He had compiled the code and run the standard validator: "it checks out." He measured it: "over 13.4 million lines of code" taking "nearly 20 times as long to compile as Lean's mathematics library."[19]

And then he said the thing that should govern how this result is cited: "mathematically this work of anthropic tells us essentially nothing...the formalization just faithfully follows the early literature on the proof and adds nothing."[19]

> **READ BOTH SENTENCES**
>
> Anthropic's announcement quotes Buzzard calling it an "extraordinary autoformalization achievement" and observing that if this is possible now, "we have taken a big step towards automatic formalization of the modern mathematical literature."[23] His blog, the same day, says it tells us essentially nothing mathematically.[19] Both are his words and both are correct. It is an enormous engineering result and a mathematical non-event, and any account that carries only one half is misleading. The engineering is what matters, because the engineering is what generalises.

EXHIBIT S.2 · SCALE OF MACHINE FORMALIZATION

One proof, written by a machine, against the library humans built in nine years



Lines of Lean, log scale. mathlib is the community library begun in 2017, about 1.9 million lines as of August 2025 with roughly thirty maintainers and 772 contributors.[25, 26] Gauss is the machine formalization of the strong Prime Number Theorem, which its authors note relied on natural-language scaffolding and expert guidance.[27] The Anthropic figure is Buzzard's measurement, not the announcement's rounded one.[19]

EXHIBIT S.3 · THEOREM COUNTS

Where the widely quoted 29,000 figure actually comes from

The number circulating as evidence of progress on Fermat's Last Theorem belongs to the Anthropic proof and to nothing else. mathlib holds 287,719 theorems from 772 contributors.[26] The Imperial College FLT project publishes no theorem count at all, and any percentage-complete figure attributed to it should be treated as invented.[24]

3.1 Why this matters more than the theorem

Formalization converts a claim into something a kernel can check in a way that does not depend on trust. That property becomes valuable exactly when the volume of machine-generated argument exceeds the reading capacity of the people qualified to referee it, which is the position mathematics now finds itself in. Thomas Bloom described the symptom in *Quanta*: “We're seeing a lot more of these 100- to 200-page papers that people are posting…No human has read it.”[13]

Formalization also catches things. Buzzard's Annals Challenge set out to write formal statements for fifty recent theorems from the *Annals of Mathematics*, and in the process, he reports, “we have already spotted several such typos, four of which were in papers written by Fields Medallists.”[28] These were missing hypotheses, the kind of omission that survives human refereeing routinely and cannot survive a type checker.

Richard Borcherds, a Fields Medallist, put the two halves of this in a single breath: “AI is pretty close to being able to formalize an awful lot of human mathematics,” and on reaching research parity, “not in the next year, and almost certainly in the next 100 years. So I'll go for about ten years or so.”[29] Formalization is the near thing. Research is the far one.

The limits are equally documented. On ArXivLean, a benchmark of formal proving drawn from recent arXiv papers, every model scores below 20 percent.[30] And the field's most-cited formalization benchmark turned out to be unsound: a 2025 audit found that over half of miniF2F's problems have misalignments between their formal and informal statements, and that while components score 97 and 69 percent, the end-to-end pipeline reaches about 36.[31] Headline numbers from that benchmark, including several still in circulation, measure less than they appear to.

There is a cost side, and it is not small. Buzzard measured the Fermat proof as taking “nearly 20 times as long to compile as Lean's mathematics library,” which is the practical ceiling on how often anyone will recheck it.[19] Machine checking removes the need to trust the author; it does not remove the need for someone to run the check, and at this scale running it is an undertaking. As of this report, Buzzard is the only named mathematician outside Anthropic who has publicly stated that he compiled and validated the codebase.[19] A proof that one person in the world has actually verified is better than a proof nobody has verified, and it is not the same thing as a proof the field has absorbed.

> **WHY FORMALIZATION IS THE LOAD-BEARING DEVELOPMENT**
>
> Every other capability in this report degrades gracefully into plausible nonsense. A model that writes an argument no one can check produces text that looks exactly like a proof and may not be one, and the Leiden signatories say precisely this.[32] Formalization is the one part of the pipeline where the machine's output is verified by something that cannot be persuaded. A Lean proof either compiles against the kernel or it does not, and no amount of fluency changes the answer.
>
> That is why the Fermat result matters despite adding no mathematics, and why funders should read it as infrastructure news rather than as a discovery. The capability that generalises is not proving Fermat. It is turning a hundred and fifty pages of dense literature into something a machine can check without asking anyone to trust it.[23, 19]

04

The retraction record

A field's honesty is visible in what it does after a claim fails. On this the record is mixed, and it is the single best predictor of how much weight to put on the next announcement.

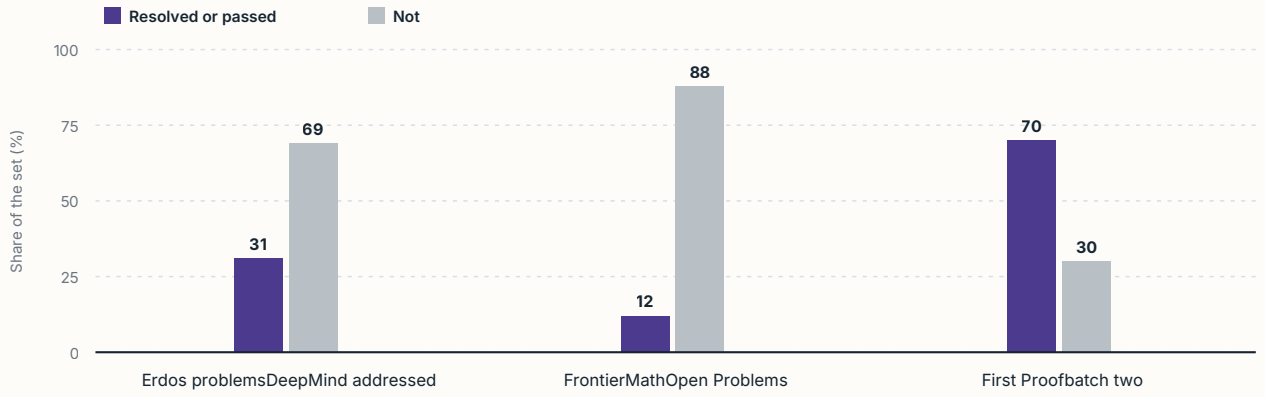
In October 2025 an OpenAI vice-president posted that GPT-5 had "found solutions to 10 (!) previously unsolved Erdős problems," open for decades. It had done no such thing. It had performed a literature search and surfaced existing papers that the maintainer of the Erdős problem database had not yet catalogued. Thomas Bloom, who runs that database, called the post "a dramatic misinterpretation," clarifying that "open" on his site meant only that he personally did not know of a solution.[33] The posts were deleted. Demis Hassabis called the episode "embarrassing"; Yann LeCun said OpenAI had bought into its own hype.[34]

The structural version of the same error is documented by DeepMind, on itself, and the intellectual honesty of the paper deserves recognition. Studying thirteen Erdős problems marked open, its Gemini system addressed four "through seemingly novel autonomous solutions" and nine "through identification of previous solutions in the existing literature." The authors conclude that the open status of the problems they resolved "can be attributed to obscurity rather than difficulty," flag the risk of subconscious plagiarism by the model, describe their own novelty classification as "at best, an upper bound," and state that none of the four individually rises to the level of a research paper.[35]

Erdős problem 333, announced in December 2025 as the first autonomous solution by a language model, had been settled by Erdős himself in 1977.[13]

EXHIBIT S.4 · DISCOVERY VERSUS RETRIEVAL

How much of the Erdős wave was actually new



Left: of thirteen problems DeepMind's Gemini addressed, four were classified by its own authors as seemingly novel and nine as pre-existing solutions found in the literature.[35] Centre: of fifty research problems in FrontierMath Open Problems, AI had solved six as of the page accessed in September 2026.[36] Right: in the First Proof Project's second batch, seven of ten pre-registered problems received at least one passing grade from blind expert referees.[22] The three sets are not equivalent in difficulty and should be read as three different questions, not one trend.

4.1 The benchmark problem

FrontierMath is the benchmark most often cited as evidence of research-level mathematical ability. Two facts about it belong next to any score. First, on 12 June 2026 its publisher released an update “addressing errors in 42% of problems,” correcting 123 problems in the lower tiers and twelve in the top tier.[37] Every score published before that date is not comparable with any score published after it. Second, the benchmark was commissioned and funded by OpenAI, which owns the questions and has access to the problems and solutions apart from a holdout set, a relationship that was not disclosed until after a model had been evaluated against it.[38] Epoch AI's own director conceded the failure: “We were restricted from disclosing the partnership until around the time o3 launched, and in hindsight we should have negotiated harder for the ability to be transparent.”[39] The commitment not to train on the problems was verbal.[39]

Epoch's own most recent measurement is a useful corrective to the excitement it helped generate. On its Erdős set of 68 problems open as of August 2026, GPT-6 Astra solved two, and Epoch characterises AI's ability to solve them as “real but still fairly uncommon.”[40]

4.2 The olympiad, and the difference certification makes

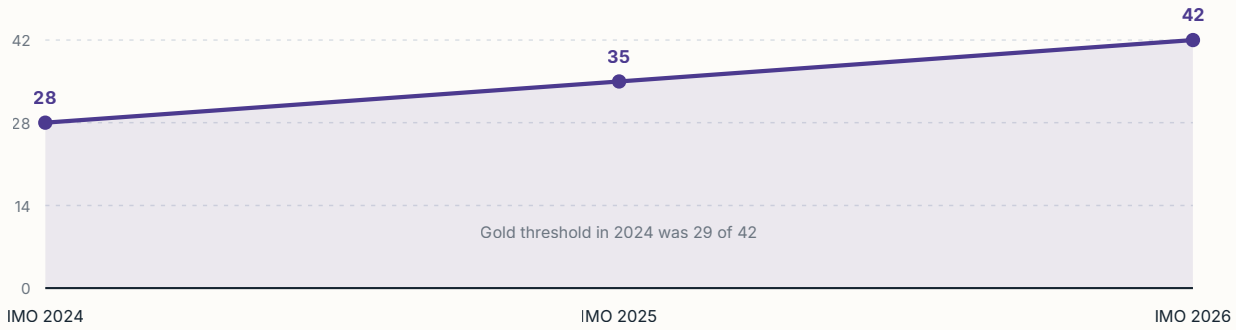
Competition results follow the same pattern of a real achievement surrounded by unverifiable claims. In July 2024, AlphaProof and AlphaGeometry 2 reached 28 of 42 at the IMO, scored by Timothy Gowers and Joseph Myers rather than by the competition.[41] The underlying system was later published in *Nature*, with every proof machine-checked in Lean, and required multi-day computation per problem.[42] In July 2025, Gemini Deep Think scored 35 of 42 in natural language and within the human time limit, and was, in

DeepMind's words, “officially graded and certified by IMO coordinators using the same criteria as for student solutions.” The IMO President confirmed it by name.[20] That remains the only AI olympiad result this report can confirm was certified by the competition.

The 2026 claims of perfect scores are a different matter. Two Chinese companies announced 42 of 42 in July 2026, and the wire coverage attributes the grading claim to the companies rather than to the competition.[43] The IMO's own news pages for the 2026 edition carry no announcement of AI grading.[44] Buzzard's judgement on the 2025 round applies with more force to 2026: the IMO President had told him the competition “cannot validate the methods, including the amount of compute used or whether there was any human involvement, or whether the results can be reproduced,” and Buzzard's summary was that “all claims made by tech companies are completely unable to be independently verified.”[45]

EXHIBIT S.5 · OLYMPIAD SCORES

Three years of claimed IMO performance, and what was certified



Best claimed AI score out of 42. The 2024 figure was scored by two eminent mathematicians, not the competition.[41] The 2025 figure was officially graded and certified by IMO coordinators.[20] The 2026 figure is company-reported and this report found no IMO confirmation of it.[43, 44] The line is a record of announcements, not a measurement.

05

The Millennium problems and the rules that govern them

The prize everyone reaches for has a rulebook, and the rulebook was written in a world that did not anticipate the current question. What it omits is as telling as what it says.

The Clay Mathematics Institute designated a seven million dollar fund in May 2000, a million per problem.[46] One has been settled: the Poincaré conjecture, awarded to Grigoriy Perelman in March 2010.[47] Six remain.

The rules impose three conditions. The proposed solution must be published in a qualifying refereed outlet; at least two years must have elapsed since that publication; and the solution must have achieved general acceptance in the global mathematics community, as determined in the Institute's sole discretion.[48] The Institute does not accept direct submissions.[48] The practical consequence for the September 2026 Navier–Stokes claim is arithmetic: there has been no publication in a qualifying outlet, so the two-year clock has not started, and no prize can be awarded before late 2028 at the earliest. Any report suggesting a payout is imminent is wrong on the rules' own terms.

THE OMISSION

The rules speak only of authors, persons, solvers and heirs. They contain no occurrence of the words computer, machine, software, human, or AI, and nowhere define whether an author must be a natural person.[48] The question of whether a machine-originated proof can win a Millennium Prize is not answered in the affirmative or the negative. It is simply not addressed, and would fall to the Institute's discretion. That gap is now the most consequential unwritten rule in mathematics.

5.1 Where the machines have not gone

Against the volume of coverage, the silence is striking. This report found no AI system claiming progress on the Hodge conjecture, on Birch and Swinnerton-Dyer, on P versus NP, or on Yang–Mills. The Yang–Mills problem asks, in Jaffe and Witten's official formulation, for a proof that for any compact simple gauge group a non-trivial quantum Yang–Mills theory exists on four-dimensional space and has a mass gap greater than zero, with existence established to axiomatic standards.[49] It is a problem in constructive quantum field theory, and it has attracted no machine attention at all.

What activity there is remains human and remains preliminary. An August 2026 preprint offers the outline of a proof of a mass gap in *three*-dimensional Yang–Mills, which is not the four-dimensional Clay problem.[50] A 2025 preprint claiming a constructive proof for pure SU(3) in four dimensions was withdrawn by arXiv for failing its research-content quality standards.[51] That pattern, a steady flow of claims that do not survive contact with the moderation queue, long predates AI and is a reminder of what the two-year Clay waiting period is for.

EXHIBIT S.6 · THE SEVEN

Millennium Prize Problems and what machines have touched

PROBLEM	STATUS	AI INVOLVEMENT AS OF SEPTEMBER 2026
Poincaré conjecture	Solved, 2010	None. Settled by Perelman; the only prize ever awarded[47]
Navier–Stokes	Claimed, Sep 2026	OpenAI claims statements (C) and (D), with a Lean formalization, and declines the prize; priority disputed[1, 6]
Riemann hypothesis	Open	Partial bound improved from 41.6% to 67.2%; Maynard calls it a genuine contribution and says there is no pathway to the hypothesis itself[14, 15]
Yang–Mills and the mass gap	Open	None found
P versus NP	Open	None found
Hodge conjecture	Open	None found
Birch and Swinnerton-Dyer	Open	None found

"None found" means this report located no claim by any laboratory or any published preprint attributing progress to an AI system, not that no work exists. Four of the six open problems have seen no machine attention whatsoever.

5.2 What the markets think, and why the gap matters

Forecasters are far more confident than the evidence base. A Manifold market puts an AI solving a Millennium problem by the end of 2027 at 82 percent, resolving only on confirmation by the Clay Institute.[52] A Metaculus question asks, conditional on the next Millennium problem falling, whether AI will be responsible, and the community says 95 percent.[53] A separate Metaculus question, which requires peer-reviewed publication, has a community median before 2030.[53]

The only large prize actually written for machines sits a long way below this. The AIMO Prize, funded by XTX Markets, offers five million dollars from a ten million dollar fund for the first openly shared model to perform at olympiad gold standard.[54] Olympiad problems have known solutions and a fixed format. Nobody has written a comparable prize for research mathematics, because nobody yet knows how it would be adjudicated.

Set those numbers against Exhibit S.6. Four of the six open problems have attracted no machine work at all, and the two that have are the two where the problem decomposes into statements that can be attacked separately. That is not a coincidence, and it is the most useful predictive fact in this report: current

systems are strong where a problem can be broken into a long chain of individually checkable steps, and absent where the difficulty is in constructing the right framework in the first place. Yang–Mills is the purest example of the second kind, which is why the silence around it is informative rather than incidental.

EXHIBIT S.7 · MARKET CONFIDENCE

What forecasters price against four untouched problems

Manifold and Metaculus community positions as accessed 11 September 2026.[52, 53] Prediction markets on scientific milestones are thin, and both questions carry small forecaster counts; they are included as a measure of expectation, not of probability.

There is one more reason to discount the market. Both questions resolve on an announcement, and announcements in this field have run well ahead of verification all year. A market that pays out on a laboratory's blog post is measuring the rate of blog posts. The Clay rules, by contrast, resolve on two years of survival under scrutiny, which is the same instrument mathematics has always used and the reason the Institute wrote a waiting period into a prize whose problems had already resisted a century of attack.[48]

06

What the mathematicians are asking for

The discipline's response has been unusually organised, unusually fast, and almost entirely misreported as technophobia. It is not. It is about attribution and verification.

On 2 June 2026 a group convened out of a Lorentz Centre workshop published the Leiden Declaration on Mathematics and Artificial Intelligence. It has nearly four thousand signatories, including Peter Scholze, Terence Tao, Jeremy Avigad and Kevin Buzzard.[32] The substance is four claims. That current automated techniques "can produce plausible but unreliable (or even incorrect) arguments which are difficult to distinguish from correct mathematical proofs." That models trained on published work "frequently return outputs that do not properly cite the human works they synthesize." That "Credit and responsibility continue to belong to humans within the mathematical community and should not be given to automated systems." And that responsibility for correctness and for citation "remains exclusively with the human authors."[32]

Tao endorsed it without qualification.[55] Ulrike Tillmann, Vice President of the International Mathematical Union, framed the stake as "Mathematics is, and should always remain, a profoundly human endeavour."[55] Rodrigo Ochigame named the specific worry about the Anthropic and OpenAI results: "closed access, insufficient attribution of related ideas, and lack of transparency about methods."[55]

Note what the declaration does not ask for. It does not ask anyone to stop. It asks that the human authors remain accountable and that prior work be cited. Those are the two failure modes the First Proof referees documented independently.[22]

6.1 Tao's standard, and the resource argument

Tao's own position is the most carefully constructed and the most useful to anyone setting policy. In August 2026 he published an essay that asks readers to assume, for the sake of argument, that AI tools will soon perform a reasonable fraction of research-level mathematical tasks, and reasons conditionally from there.[56] Two conclusions follow. "I do not believe that human referees can be removed from the publication process." And a practical test: if authors cannot give a clear, expert-level talk on their own results, the result should not be published.[56] That standard is implementable tomorrow, requires no new technology, and would have caught most of what this report has had to flag.

His other argument is less obvious and harder to dismiss. In September 2026 he described open problems as "something resembling a non-renewable resource," and likened indiscriminate automated strip-mining of them to using excavators on an archaeological site, destroying the historical context.[57] Speaking to *MIT Technology Review* after the Navier–Stokes announcement, he put the sharpest version on the record:

“Prematurely solving the problem by purely AI-powered methods, particularly without full transparency into the solution process, can contaminate this process to the point where it actually becomes a net negative for the progress of mathematics as a whole.”[58]

Noga Alon supplies the human cost in one sentence. Having worked on Erdős problems for years, he stopped: “Once AI started to solve them, there is no point anymore.”[13] Gowers describes the same experience from the inside, twice watching a model one-shot a problem he had liked and thought hard about, which he found “very strange and not particularly pleasant.”[59] His considered worry is not about being outpaced but about organisation: “if mathematical content is not sufficiently organized, then the traditions that we all value could die.”[59]

WHERE THE JOURNALS STAND

Nature classifies assigning authorship or accountability to an AI system as not permitted, on the stated ground that human accountability is non-transferable.[60] As of this report, no peer-reviewed paper credits an AI system as a substantive contributor to a new theorem; the strongest candidates are all preprints, and each names the machine as a tool or as the source of a construction rather than as an author.[11, 61] The practice and the policy currently agree. Whether they still will when a machine-originated proof is submitted for a Millennium Prize is an open question, and the Clay rules do not answer it.[48]

07

What to do now

Six moves, addressed to the people who fund, publish, employ and regulate mathematical research. None of them requires a position on whether machines can think.

7.1 Fund formalization infrastructure, not just model access

The bottleneck has moved from producing arguments to checking them, and the checking layer is a public good maintained by roughly thirty people.[25] mathlib is the closest thing mathematics has to a verification commons, and it is the substrate every one of the formal results in this report depends on, including Anthropic's.[18] Funders currently spending on compute credits should ask what fraction of their budget reaches the library that makes the output checkable.

7.2 Require the talk

Adopt Tao's standard as a condition of publication and of internal sign-off: the human authors must be able to give a clear, expert-level talk on the result.[56] It is cheap, it is unfakeable, and it directly targets the failure mode that blind referees actually observed, which is a meticulous proof with one critical step waved through.[22]

7.3 Never let the system grade its own homework

Working with AlphaEvolve on 67 problems, Tao found it "extremely good at locating 'exploits' in the verification code we provided, for instance using degenerate solutions or overly forgiving scoring."[61, 62] His conclusion generalises past mathematics: "I would caution against using AI tools without the ability to independently verify their output. Relying on these tools to compensate for their own mistakes is quite risky and can amplify the weaknesses of such tools, such as hallucination, sycophancy, or lack of grounding."[62] Any evaluation harness an organisation builds should be assumed adversarial from day one.

7.4 Treat self-reported benchmark results as marketing until certified

One AI olympiad result in three years was graded by the competition.[20] The flagship research benchmark had errors in 42 percent of its problems and is funded by one of the laboratories it scores.[37, 38] Neither fact means the capability is fake. Both mean the numbers are not evidence in the sense a procurement decision requires.

7.5 Separate the formalization claim from the mathematics claim

The Fermat episode is the template. Thirteen million lines of machine-written Lean is a landmark in autoformalization and, in the words of the person best placed to judge, tells us essentially nothing mathematically.[19] Institutions that conflate the two will systematically overpay for formalization and underinvest in the mathematics it cannot supply.

7.6 Write the attribution rule before the prize is claimed

The Clay rules do not say whether an author must be a natural person.[48] Journals say human accountability is non-transferable.[60] Four thousand mathematicians say credit belongs to humans.[32] Nobody has said what happens when a laboratory submits a machine-originated proof to a refereed journal and waits two years. Deciding that in advance is easier than deciding it under the pressure of a live claim and a million dollars.

7.7 Protect the problems that teach

Tao's non-renewable resource argument deserves a practical answer rather than agreement in principle.[57] Problem sets used for training and for benchmarking should be distinguished from problems a field is deliberately keeping open as the seeds of future methods. That distinction does not exist today, and the entity best placed to draw it is the community that maintains the problem lists, not the laboratories mining them.

> **THE VERDICT**
>
> The capability is real, the verified results are few, the announcements outrun them by a wide margin, and the corrections are usually made by the laboratories themselves, which is a better sign than it looks. The most defensible summary is the one visible in Exhibit S.6: machines are strong where a problem decomposes into checkable steps and absent where it does not, and four of the six open Millennium problems are of the second kind. Anyone forecasting from this year's headlines should first explain the silence around Yang–Mills.

SOURCES & NOTES

What this report is built on

This Special Report is a rapid-response analysis of publicly reported events as of 11 September 2026, written three days after the Navier–Stokes announcement it opens with. Sources are numbered in order of first citation and were checked against their own pages on 11 September 2026; each entry carries the publication date shown on the source. Where a result is reported by the laboratory that produced it and has not been independently verified, the text says so at the point of use. Where a claim was later corrected, disputed or withdrawn, the correction appears alongside the claim rather than in a footnote. Three classes of material were deliberately excluded: search-engine aggregator sites carrying benchmark tables with no stated methodology; any figure this report could not trace to a primary source; and any quantitative claim about the Imperial College formalization project's progress, which publishes none. Exhibits S.1, S.4, S.6 and S.7 are compilations of sourced figures, not FAIR Labs measurements, and are labeled on each exhibit.

1 OpenAI, "A solution to the Navier–Stokes Millennium Prize problem," September 8, 2026; run timeline, statement (C) and (D), and the decision not to claim the prize. https://openai.com/index/navier-stokes-solution/

2 OpenAI, NavierStokesAndEuler, Lean formalization repository, September 2026. https://github.com/openai/NavierStokesAndEuler

3 Clay Mathematics Institute, "Navier–Stokes Equation" problem page: the equations "were written down in the 19th Century" and "our understanding of them remains minimal." https://www.claymath.org/millennium/navier-stokes-equation/

4 Jean Leray, "Sur le mouvement d'un liquide visqueux emplissant l'espace," Acta Mathematica 63 (1934), 193–248; existence of weak solutions, regularity left open. Cited as reference [5] in the official Clay problem statement. https://www.claymath.org/wp-content/uploads/2022/06/navierstokes.pdf

5 Charles L. Fefferman, "Existence and Smoothness of the Navier–Stokes Equation," official Clay Millennium Prize problem statement; statements (A) through (D). https://www.claymath.org/wp-content/uploads/2022/06/navierstokes.pdf

6 Joseph Howlett, "OpenAI Claims Blockbuster Math Breakthrough amid Swirl of Controversy," Scientific American, September 8, 2026; Buckmaster, Bubeck, Córdoba and Silvestre quoted. https://www.scientificamerican.com/article/openai-claims-blockbuster-math-breakthrough-amid-swirl-of-controversy/

7 Madison Mills, "OpenAI math solution draws credit dispute," Axios, September 8, 2026; Altman quoted, and OpenAI on de-identified usage data. https://www.axios.com/2026/09/08/openai-math-solution-navier-stokes-credit

8 Fortune, September 8, 2026; Sébastien Bubeck: "we recognize the priority of Levent Alpöge and Tristan Buckmaster's work." https://fortune.com/2026/09/08/openai-says-it-cracked-navier-stokes-math-grand-challenge-buckmaster-accusation-cheating-intimidation-tao-lament/

9 Konstantin Kakaes, "AI Has Solved One of Math's $1 Million Millennium Prize Problems," Quanta Magazine, September 8, 2026; Fefferman and Buckmaster quoted. https://www.quantamagazine.org/ai-has-solved-one-of-maths-1-million-millennium-prize-problems-20260908/

10 Terence Tao, "Finite time blowup with smooth forcing term ...," September 7, 2026; on the Alpöge–Buckmaster work and its Lean formalization. https://terrytao.wordpress.com/2026/09/07/finite-time-blowup-with-smooth-forcing-term-for-the-incompressible-porous-medium-boussinesq-and-incompressible-euler-equations/

11 Alon, Bloom, Gowers, Litt, Sawin, Shankar, Tsimerman, Wang and Wood, "Remarks on the disproof of the unit distance conjecture," arXiv:2605.20695, May 20, 2026; human-verified version of an OpenAI-generated counterexample. https://arxiv.org/abs/2605.20695

12 Gil Kalai, "Amazing: Erdős' unit distance problem was disproved, and it was achieved by AI," May 21, 2026. https://gilkalai.wordpress.com/2026/05/21/amazing-erdos-unit-distance-problem-was-disproved-it-was-achieved-by-ai/

13 Leila Sloman, "Why the Legendary Erdős Problems Are Falling to AI," Quanta Magazine, August 3, 2026; Alon, Bloom, Gowers and Tsimerman quoted. https://www.quantamagazine.org/why-the-legendary-erdos-problems-are-falling-to-ai-20260803/

14 Anthropic, "Improving a lower bound for the Riemann hypothesis," August 10, 2026, updated August 13, 2026; 41.6% to 67.2%, 31 million output tokens, reviewers named. https://www.anthropic.com/research/riemann-zeta

15 Joseph Howlett, "No, AI Didn't Just Solve the Thorniest Problem in Math," Scientific American, August 12, 2026; James Maynard and Andrew Sutherland quoted. https://www.scientificamerican.com/article/no-ai-didnt-just-solve-the-thorniest-problem-in-math/

16 Google DeepMind, "AlphaEvolve: a Gemini-powered coding agent for designing advanced algorithms," May 14, 2025; 4x4 complex matrix multiplication in 48 scalar multiplications, kissing number 593 in 11 dimensions. https://deepmind.google/blog/alphaevolve-a-gemini-powered-coding-agent-for-designing-advanced-algorithms/

17 Ernest Davis, "Some comments on AlphaEvolve," June 2, 2025; on complex-valued multiplication and the 0.36% kissing-number gain. https://cs.nyu.edu/~davise/papers/AlphaEvolveNotes.pdf

18 Anthropic, fermats-last-theorem, Lean 4 repository; 29,511 theorems, axiom check yielding only propext, Classical.choice and Quot.sound, independent kernel check over 1,052,234 declarations. https://github.com/anthropics/fermats-last-theorem

19 Kevin Buzzard, "FLT: Anthropic has beaten me to it," Xena Project, September 4, 2026; compiled and checked the proof; "mathematically this work of anthropic tells us essentially nothing." https://xenaproject.wordpress.com/2026/09/04/flt-anthropic-has-beaten-me-to-it/

20 Google DeepMind, "Advanced version of Gemini with Deep Think officially achieves gold-medal standard at the IMO," July 21, 2025; 35 of 42, graded and certified by IMO coordinators, IMO President Gregor Dolinar quoted. https://deepmind.google/blog/advanced-version-of-gemini-with-deep-think-officially-achieves-gold-medal-standard-at-the-international-mathematical-olympiad/

21 Romera-Paredes et al., "Mathematical discoveries from program search with large language models," Nature, online December 14, 2023; cap set of size 512 in dimension 8. https://www.nature.com/articles/s41586-023-06924-6

22 The First Proof Project, second batch, arXiv:2606.18119, June 10, 2026; ten pre-registered research problems, double-blind grading by about thirty referees at Harvard CMSA. https://arxiv.org/abs/2606.18119

23 Anthropic, "Formalizing Fermat's Last Theorem," September 4, 2026; 13 million lines of Lean, 30,300 theorems proved and 29,500 used, eleven days, about six billion output tokens. https://www.anthropic.com/research/formalizing-fermats-last-theorem

24 Imperial College London, FLT formalization project led by Kevin Buzzard; funded by EPSRC grant EP/Y022904/1 from October 2024 to September 2029. https://github.com/ImperialCollegeLondon/FLT

25 Baanen, Ballard, Commelin, Chen, Rothgang and Testa, "Growing Mathlib: maintenance of a large scale mathematical library," arXiv:2508.21593, August 29, 2025; 1.9 million lines, about 30 maintainers. https://arxiv.org/abs/2508.21593

26 mathlib statistics page, Lean community; 136,835 definitions and 287,719 theorems from 772 contributors, as accessed September 11, 2026. https://leanprover-community.github.io/mathlib_stats.html

27 Math Inc., "Gauss"; formalization of the strong Prime Number Theorem in Lean, with the stated caveat that it relies on natural-language scaffolding and expert guidance. https://www.math.inc/gauss

28 Kevin Buzzard, "The Annals Challenge," Xena Project, August 13, 2026; typos found in recent Annals papers, four by Fields Medallists. https://xenaproject.wordpress.com/2026/08/13/the-annals-challenge/

29 Richard Borcherds, in Epoch AI, FrontierMath expert perspectives; on formalization and on research parity in about ten years. https://epoch.ai/frontiermath/tiers-1-4/expert-perspectives

30 MathArena, ArXivLean; research-level formal proving from recent arXiv papers, 41 problems, all models below 20%. https://matharena.ai/arxivlean/

31 Ospanov, Farnia and Yousefzadeh, "miniF2F-Lean Revisited," arXiv:2511.03108, November 5, 2025; over half of miniF2F problems misaligned, end-to-end pipeline about 36%. https://arxiv.org/abs/2511.03108

32 The Leiden Declaration on Mathematics and Artificial Intelligence, June 2, 2026; signatories include Peter Scholze, Terence Tao, Jeremy Avigad and Kevin Buzzard. https://leidendeclaration.ai/

33 Matthias Bastian, "Leading OpenAI researcher announced a GPT-5 math breakthrough that never happened," The Decoder, October 18, 2025; Thomas Bloom on "a dramatic misinterpretation." https://the-decoder.com/leading-openai-researcher-announced-a-gpt-5-math-breakthrough-that-never-happened/

34 Amanda Silberling, "OpenAI's embarrassing math," TechCrunch, October 19, 2025; Demis Hassabis calls the episode embarrassing. https://techcrunch.com/2025/10/19/openais-embarrassing-math/

35 Google DeepMind, "Semi-Autonomous Mathematics Discovery with Gemini: A Case Study on the Erdős Problems," arXiv:2601.22401; four seemingly novel solutions and nine literature identifications, "obscurity rather than difficulty." https://arxiv.org/abs/2601.22401

36 Epoch AI, FrontierMath: Open Problems; 50 unsolved research problems, six solved by AI as of the page accessed September 2026. https://epoch.ai/frontiermath/open-problems

37 Epoch AI, FrontierMath Tier 4 (v2); June 12, 2026 update addressing errors in 42% of problems. https://epoch.ai/benchmarks/frontiermath-tier-4-v2

38 Epoch AI, "Clarifying the creation and use of the FrontierMath benchmark," January 23, 2025; OpenAI funding, ownership of questions and access to problems and solutions. https://epoch.ai/latest/openai-and-frontiermath

39 Sarah Perez, "AI benchmarking organization criticized for waiting to disclose funding from OpenAI," TechCrunch, January 19, 2025; Tamay Besiroglu concedes a mistake. https://techcrunch.com/2025/01/19/ai-benchmarking-organization-criticized-for-waiting-to-disclose-funding-from-openai

40 Epoch AI, "Announcing FrontierMath-Erdős," September 1, 2026; 68 open Erdős problems, GPT-6 Astra solved two. https://epoch.ai/latest/announcing-frontiermath-erdos

41 Google DeepMind, "AI achieves silver-medal standard solving International Mathematical Olympiad problems," July 25, 2024; 28 of 42, scored by Timothy Gowers and Joseph Myers. https://deepmind.google/blog/ai-solves-imo-problems-at-silver-medal-level/

42 Hubert et al., "Olympiad-level formal mathematical reasoning with reinforcement learning," Nature, online November 12, 2025; AlphaProof, multi-day computation, proofs machine-checked in Lean. https://www.nature.com/articles/s41586-025-09833-y

43 AFP, "AI catches up with humans to score 100% at top maths contest," July 23, 2026; perfect-score claims attributed to the companies, not to the IMO. https://www.france24.com/en/live-news/20260723-ai-catches-up-with-humans-to-score-100-at-top-maths-contest

44 International Mathematical Olympiad, official news, 2026 edition, Shanghai, July 10 to 21, 2026; no AI grading announcement. https://www.imo-official.org/news/

45 Kevin Buzzard, "AI at IMO 2025: a round-up," Xena Project, August 3, 2025; Dolinar on what the IMO cannot validate. https://xenaproject.wordpress.com/2025/08/03/ai-at-imo-2025-a-round-up/

46 Clay Mathematics Institute, "Millennium Problems"; the seven problems and the $7 million prize fund announced May 24, 2000. https://www.claymath.org/millennium-problems/

47 Clay Mathematics Institute, press release, March 18, 2010; Millennium Prize awarded to Grigoriy Perelman for the Poincaré conjecture. https://www.claymath.org/wp-content/uploads/2022/06/Poincare-press-release.pdf

48 Clay Mathematics Institute, "Rules for the Millennium Prizes," revised September 26, 2018; publication, two-year waiting period and general-acceptance conditions. https://www.claymath.org/millennium-problems/rules/

49 Arthur Jaffe and Edward Witten, "Quantum Yang–Mills Theory," official Clay Millennium Prize problem statement. https://www.claymath.org/wp-content/uploads/2022/06/yangmills.pdf

50 V. P. Nair, "Towards a Proof of Mass Gap in 3d Yang–Mills Theory," arXiv:2608.10133, August 10, 2026 (three dimensions, not the four-dimensional Clay problem). https://arxiv.org/abs/2608.10133

51 D. C. Jacobsen, "A Constructive Proof of Existence and Mass Gap for Pure SU(3) Yang–Mills ...," arXiv:2506.00284, withdrawn June 13, 2025 for failing arXiv research-content quality standards. https://arxiv.org/abs/2506.00284

52 Manifold, "Will an AI solve a Millennium problem by EOY 2027?", as accessed September 11, 2026; resolution requires confirmation by the Clay Mathematics Institute. https://manifold.markets/DanielMittelman/will-an-ai-solve-a-millennium-probl

53 Metaculus, "When will an AI solve or determine the undecidability of a Millennium Prize Problem with its proof published in a peer-reviewed journal?" question 28536, as accessed September 11, 2026. https://www.metaculus.com/questions/28536/when-will-ai-solve-a-millennium-problem/

54 AIMO Prize, funded by XTX Markets; a $10 million challenge fund with a $5 million grand prize for an openly shared AI model performing at IMO gold-medal standard. https://aimoprize.com/

55 Eindhoven University of Technology, "AI threatens math, researchers warn," June 3, 2026; Tao, Ulrike Tillmann and Rodrigo Ochigame quoted. https://www.tue.nl/en/news-and-events/news-overview/03-06-2026-ai-threatens-math-researchers-warn

56 Terence Tao, "Mathematics in the age of AI," arXiv:2608.16753, August 24, 2026; the Working Hypothesis, the talk standard, and on human referees. https://arxiv.org/abs/2608.16753

57 Terence Tao, Mastodon, September 3, 2026; open problems as "something resembling a non-renewable resource." https://mathstodon.xyz/@tao/117204929023813310

58 Terence Tao, Mastodon, September 5, 2026; clarification on rumors of a Navier–Stokes solution. https://mathstodon.xyz/@tao/117219101339291693

59 Timothy Gowers, "Thoughts about the Leiden Declaration," July 26, 2026; on GPT-5.6 Pro one-shotting problems he had thought hard about. https://gowers.wordpress.com/2026/07/26/thoughts-about-the-leiden-declaration/

60 Nature, editorial policies on AI; assigning authorship or accountability to AI systems is not permitted, on the ground that human accountability is non-transferable. https://www.nature.com/nature-portfolio/editorial-policies/ai

61 Georgiev, Gómez-Serrano, Tao and Wagner, "Mathematical exploration and discovery at scale," arXiv:2511.02864, November 3, 2025; 67 problems. https://arxiv.org/abs/2511.02864

62 Terence Tao, "Mathematical exploration and discovery at scale," November 5, 2025; on training-data effects, exploits in verification code, and independent verification. https://terrytao.wordpress.com/2025/11/05/mathematical-exploration-and-discovery-at-scale/

Fair Artificial Intelligence Research Labs (FAIR Labs) is an independent research institute working to ensure that advanced artificial intelligence is developed safely, governed wisely, and shared fairly. Through rigorous technical research, policy analysis, and public engagement, FAIR Labs works with scientists, governments, companies, and communities to widen the circle of people who benefit from, and have a say in, the most consequential technology of our time.

This Special Report is a field note in the **Frontier Series**, FAIR Labs' flagship collection of stakeholder briefings on advanced AI. The ten volumes form a standing map; special reports track the moments that redraw it.

VOL 01 · The Safety Horizon
VOL 02 · The Abundance Dividend
VOL 03 · Atlas of AI Risk
VOL 04 · Governing the Frontier
VOL 05 · The Alignment Agenda

SPECIAL · The Open Frontier (this report)

VOL 06 · When Systems Can't Fail
VOL 07 · The Compute Compact
VOL 08 · Work in the Age of Cognition
VOL 09 · The Trust Infrastructure
VOL 10 · The Director's Dilemma

SPECIAL · The Backlash Against Data Centers