« My new course at UT Austin: AI Alignment Theory

# The Mathocalypse

> … then they came for Navier–Stokes and I said nothing because I never worked on Navier–Stokes. But when they came for RL vs. L I realized that things are serious

–friend-of-the-blog Omer Reingold (shared with permission)

---

Last night my 9-year-old son was taunting my wife, complexity theorist Dana Moshkovitz, as follows: “mommy, I heard you got *cooked*! I heard that a *robot* solved the math problem you worked on for your whole career! OOF!”

While my son was being a brat, he also wasn’t wrong. Whether you’re thrilled, depressed, angry, or whatever else about it, yesterday was surely one of the biggest days in mathematical history. And yes, among the 372 huge results released yesterday by OpenAI, on the recommendation of its advisory group of Timothy Gowers, Edward Witten, and other distinguished mathematicians, was a proof of Subhash Khot’s Unique Games Conjecture (UGC), a statement that my wife has worked toward proving for the entire time I’ve known her. (The UGC implies that a whole slew of optimization problems really are NP-hard, even if you just want an approximation that’s slightly better than what you get from semidefinite programming relaxation, which is one of our main tools.)

Or at least, we’re pretty sure that it’s a proof! There’s a Lean certificate, as there are for some of the other 372 breakthrough results (not all of them). But it also appears that no human has understood just about *any* of these proofs yet; the race to do so has just started. If you want an on-the-ground sense of what that race is going to be like, here’s some of what Dana texted me last night:

> It feels like something written by someone who’s on psychedelics. So much unclear and doesn’t make sense. Lots of name dropping of previous work without discussing why it can be used despite impossibility results
>
> Basically the paper is so horribly written that it’s impossible to read it without AI help
>
> I asked Astra for reasonable completeness and soundness claims of the noise gadget and it gave them by combining claims from all over the paper
>
> They also have direct optimal NP hardness of approximation proofs for the main applications of the UGC (Max Cut and all CSP) that bypass the UGC.
>
> The UGC proof invents a completely new bizarre code with a noise test. It’s some crazy recursive construction.

> It's not the long code, not the short code – some alien craziness
>
> I still think that there maybe is a proof that uses the half space code (which is natural)
>
> The citations are often irrelevant and confusing
>
> A possible future is a math world that's heavenly if you have vision/creative ideas that AI could help check and implement.
>
> And of course there's a lot for us to learn from the aliens

If you're wondering what emotions Dana is feeling—well, probably all of them! Even while a central career aspiration has fallen to a robot, there are at least two mitigating factors for her. First, she can feel vindicated that the UGC was *true* after all, something she never doubted even while many of her colleagues did! Second, *all* of us in math and theoretical computer science and mathematical physics, at least those who cared about solving crisply-stated problems, are now in the same boat.

---

Besides the Unique Games Conjecture, here's a small sampling of the treasures from Aladdin's cave that I'll probably be paying the most attention to over the coming weeks:

» L=BPL (i.e., probabilistic logspace and deterministic logspace are the same thing), one of the great derandomization conjectures short of P=BPP. Though its truth was never in serious doubt, there was a whole subcommunity focused on proving this.

» The Fourier Transform and integer multiplication in less than O(n log n) time, breaking a barrier that had stood since the 1960s. The new running time, if you're curious, is $O(n \log^{0.9999999999999} n)$, give or take some 9's.

» Positive solution to the Unitary Synthesis Problem, which Greg Kuperberg and I posed back in 2007. For every n-qubit unitary transformation U, there exists a classical oracle A such that U can be implemented in quantum polynomial time with access to A. This is the opposite of what most of us expected, and could have implications for e.g. the computational problem of decoding Hawking radiation from a black hole and many other problems in quantum complexity theory—*if* we had an efficient way to construct the oracle A, which this paper doesn't give. (**Added:** Here's a beautiful blog post by friend and colleague Dakshita Khurana about the years she spent working on this problem—unlike me and most others, she believed correctly in a positive answer to it—and her current feelings as she works through the AI's proof.)

» Parity is not in $QAC^0$, one of the great questions of quantum complexity theory since 1999 that many of my colleagues had been closing in on.

» Nearly $4^{th}$-power separation between randomized and quantum query complexity for total Boolean functions. A favorite problem of mine since 1998 (!), when we knew only that the optimal separating exponent was between 2 and 6. For the past few years, we knew it was between 3 and 4. So, this finally closes that story.

» A superquadratic separation between sensitivity and block sensitivity.

» Area law for 2D gapped Hamiltonians. One of the main open problems in Hamiltonian complexity.

» Matrix multiplication in $O(n^{9/4})$ time—a *rational* exponent for once (!), and via a completely different approach than was used for $O(n^{2.373})$ and so forth

» $\Omega(n^3)$ lower bound on the determinantal complexity of the permanent, improving the previous best bound which was quadratic.

» A randomized polytime algorithm to approximately count the number of perfect matchings in a general graph, as well as a randomized nearly linear-time algorithm for *finding* a maximum matching in such a graph

» Uncomputability of solving polynomial equations over the rational numbers—this was arguably the biggest open problem in computability theory (note that uncomputability of solving Diophantine equations, i.e. polynomial equations over the *integers*, was proved in the 1970s, giving a negative answer to Hilbert's 10th Problem)

Any of the above, alone, could easily have been "result of the year" in some area (and in some cases, like Unique Games and L=BPL, in all of CS theory). And there's a lot that I've left out—feel free to share in the comments whatever is making *your* eyes bug out! There are equally astounding wonders in number theory, combinatorics, algebraic geometry, analysis, and pretty much every other area of math, most of which I'll never understand, although I'll note that it includes partial progress toward the Riemann hypothesis and the Hodge Conjecture and the Birch-Swinnerton-Dyer Conjecture (i.e., the majority of the remaining Millennium Problems).

We can take solace in what's missing from the list. P≠NP isn't there, nor even P=BPP or NEXP⊄P/poly, and surely not for lack of trying. Apparently the greatest open problems of theoretical computer science are indeed pretty hard!

---

Oh, lest I forget: one day *before* the OpenAI dump, meaning Monday evening, Virginia Williams and Josh Alman posted an arXiv preprint that solves the 3SUM problem in $O(n^{1.9992})$ time, and the All-Pairs Shortest Paths problem in $O(n^{2.9995})$ time, refuting half-century-old conjectures that the correct answers were $n^{2-o(1)}$ and $n^{3-o(1)}$ respectively. In this case, it wasn't an OpenAI model that supplied the crucial idea; it was an Anthropic one! But Anthropic then took a different approach from OpenAI: rather than post the undigested solutions to the world, it gave Virginia and Josh the opportunity to write and announce a digested version in exchange for compensation.

These have emerged as the two main models for communicating AI math breakthroughs, and they both have strengths and weaknesses. The "OpenAI model" sets up a crazy race among humans to digest and explain a messy AI proof (work that could easily be some combination of thankless, barely-credited, competitive, and unfun), while the "Anthropic model" puts a private company in the position of picking and choosing which human mathematicians get to be the emissaries of the AI. Dunno, what do you guys think?

---

For those who are wondering: apparently, the AI model that produced all these wonders was *not* bespoke contraption of 10,000 agents burning millions of dollars worth of compute, as was used for example to construct a finite-time blowup for the Navier-Stokes equations. Instead, it was simply the latest internal OpenAI model—one that might be released to paying ChatGPT customers within the next couple of months, depending on the recommendations of OpenAI's safety board! (My 9-year-old son: "Oh they *definitely* shouldn't release that. If it could solve all those math problems, it can't *possibly* be safe.") Apparently they used about 3 hours of GPT-Pro level compute on average per problem solved.

Also, if you were wondering: apparently they *tried* the model on about 8,000 problems. So, right now it "merely" solves ~5% of the longstanding open mathematical problems that it's asked about, the problems that whole communities have spent years on, after a single 3-hour attempt on them.

---

I've been glad to see the CS theory community rising to the occasion. At the Simons Institute in Berkeley, here at UT Austin, and elsewhere, I've hearing stories of researchers rushing to pore over the manuscripts and *make sense of them and explain them*—because what else do we do? How else do we continue the craft to which we've devoted much of our lives?

If you want some sense of what things feel like now in math, imagine a hunter-gatherer who's spent his entire life learning to survive deep in an unforgiving rainforest, then a giant resort hotel springs up right next to him with a helipad and heated pools and AirBnBs, and without missing a beat, the hunter-gatherer says: "alright fine, so now my new job is to run wilderness retreats for the tourists, or something."

In *Quanta* magazine, Jordana Cepelewitz attempted a different metaphor:

It's as if you were teleported to the peak of a tall mountain. Surrounded by fog, you have no idea where you are, or what's around you. You do not know how your mountain connects to others, and you have no equipment to help you explore, no way to help someone else join you. If you had climbed the mountain yourself, you would have experienced how the human body adapts to altitude and changes in oxygen levels. You might have had to invent tools to navigate, to climb steep cliffs, or to make a shelter. You might have encountered a fellow explorer, gotten lost together in a hidden valley, and found a plant that could be turned into a life-saving medicine.

Instead you're perched on the peak but in the dark, while the maker of the teleportation machine tells you that it can explore the wilderness better than any human.

For any one of these mountains, if we care enough, I feel optimistic that we can do as we always have: clear the fog and figure out the path, except now using the teleportation machine to help guide us. The bigger challenge will be to nurture a community that still *cares* about the heroic adventure of finding the paths up these mountains in the world with the machine. (Oh, and I think one place where the metaphor breaks is that we still *do* have each other, as much as we ever did before!)

---

Experience has shown that, *even now*, there will still be people explaining in patronizing tones why none of this is real and none of it counts. If such people were capable of being impressed by *anything* that happens in the empirical world, of updating on *anything*, they would've already been impressed and already updated several years ago, long before things had reached the point of an actual Mathocalypse.

So, they'll say, maybe the alleged solutions are not solutions at all, but just "AI slop." Or maybe none of the 372 well-known open problems that were solved were *real* math problems, they were all just glorified contest puzzles and trivialities. (After all, there's still no Riemann Hypothesis!) Or maybe the entire 4000-year-old discipline of mathematics needs to be jettisoned: turns out that it was *all* just puzzle-solving and trivialities; all that's different is that now the triviality stands unmasked. In any case, what really matters is that the *true* inner sanctum of human creativity hasn't been breached and probably never will be, and also, that Sam Altman and Dario Amodei are contemptible little nerds.

If you're still a proponent of that doomed worldview, still aboard the sinking ship, I encourage you in the strongest possible terms to read yesterday's *other* great contribution to AI discourse, besides the OpenAI Mathocalypse dump: namely, Scott Alexander's open letter to Steven Pinker. I feel some responsibility for this, as the person who first introduced Steven Pinker to the *existence* of the rationalist community, and who also first introduced Steven

Pinker and Scott Alexander to one another (they had both been fans of each other's writing). And now Scott is challenging Steve to a literal duel, with guns!

For whatever it's worth: Steve is a lifelong intellectual hero of mine, just as he is for Scott, and I also have to privilege of calling Steve my friend. But I found Scott's post to be one of the most devastating rejoinders to anything that I've ever read. And I thought Scott's conclusion was exactly right: when it comes to AI risk, Steve's great challenge is now to accept and start using a more "Pinkerite" epistemology.

---

Last night, while I should've been poring over some of OpenAI's hundreds of papers and/or writing this post, I decided to spend some time with my kids instead. They wanted a movie night, so I suggested something they'd never seen before (and that I hadn't seen for decades), and that seemed chock-full of no-nonsense, practical guidance for the world in which they're going to grow up: *Terminator 2*.

---

**Update:** As several people have pointed out, cryptography is a subfield that's extremely conspicuous by its absence from OpenAI's list of 376 papers! But my sources tell me that the AI companies have now started, gingerly and discreetly, investigating whether their latest internal models can break important cryptographic protocols and primitives. If they *can*, then it would certainly be nice to get ahead of things before the rest of the world figures out the same.

**Another Update:** The statement put out the Advisory Group on Mathematics and Artificial Intelligence is very carefully phrased, neither endorsing nor condemning what OpenAI did, and is worth a read:

> As announced a few weeks ago, OpenAI has released a large collection of mathematical results generated by an internal model, reporting solutions to hundreds of open questions. This is an important event for mathematics, with consequences both for mathematics and for the mathematical community that extend far beyond the individual results.
>
> AGMAI's advisory role should not be interpreted as a judgment of the impact of these results or an endorsement of the process by which OpenAI obtained them. We do not speak on behalf of the entire mathematical community, and only the mathematical community can undertake the assessment that is needed.
>
> Making this work public is a first step. This release is the beginning, not the completion, of the process of human understanding and the incorporation of the work into mathematical knowledge. At the same time, the future of mathematical research cannot consist only of understanding results produced by AI labs. Mathematicians must be able to formulate their own questions, develop their own approaches, and explore directions that have not been selected as examples of an AI system's capabilities. Equitable access to powerful research tools and adequate computational resources are essential to that freedom.
>
> We reaffirm our published recommendations on responsible release. We have discussed them with OpenAI and appreciate the company's willingness to engage. While we consider these discussions constructive, it is ultimately up to the mathematical community to assess the extent to which our recommendations were followed successfully, and whether there are others we should suggest. We remain committed to engaging with any frontier AI lab on these questions and have already been in contact with several of them.

Post Follow Share 0

This entry was posted on Wednesday, October 7th, 2026 at 1:57 pm and is filed under Uncategorized. You can follow any responses to this entry through the RSS 2.0 feed. You can leave a response, or trackback from your own site.

## 97 Responses to “The Mathocalypse”

**Sniffnoy** Says:
Comment #1 October 7th, 2026 at 2:12 pm

Hey remember how just a few months ago Hong Wang won the Fields Medal for her work on proving the Kakeya Conjecture in 3 dimensions? Well now OpenAI says they’ve done it in 4 dimensions, just like that…

**Scott** Says:
Comment #2 October 7th, 2026 at 2:32 pm

Additional commentary from Dana on the UGC proof:

> The Unique Games Conjecture is true, though it wasn’t really needed for the proof of its most famous corollaries (like optimal hardness of approximating Max-Cut and other constraint satisfaction problems). The published proof of the conjecture somehow makes the short code approach* work with a bunch of new ideas (it’ll take time to understand exactly how, the writeup is poorly written). I still wonder whether it’s possible to prove the conjecture via the half-space code.
>
> * Addendum: Upon further inspection, the proof constructs a completely new tree-based code that somehow has a noise test with properties similar to that of the short code.

**Manu** Says:
Comment #3 October 7th, 2026 at 2:43 pm

I can only speak for the logic part, but the results are astonishing. All these famous problems solved in the blink of an eye. Pretty depressing news, in my opinion. The only future i see now for human math is a slow death. Lets only hope humanity itself fares better…

**Olav** Says:
Comment #4 October 7th, 2026 at 3:01 pm

It’s surreal to be living through this. My condolences (so to speak) to Dana for having one of her own problems crushed by the robot. But it’s interesting to hear her initial reactions to the proof—I hope the AI labs manage to train the models to write papers in a way that’s more understandable to humans in the not-too-distant future. Scott, if you go to the trouble of studying in some depth any of the results you mention in your post, it would be very interesting to hear your reactions and insights as well.

**Dana Moshkovitz** Says:
Comment #5 October 7th, 2026 at 3:03 pm

I don’t think that the lonely teleport to the peak of the mountain is a good metaphor, not just because we’re not lonely. For each of those problems we had a community, frameworks and existing results that the AI clearly built on. The AI results represent paradigm shifts, they do teleport us much farther, but they don’t finish math.

**Andrei** Says:

Comment #6 October 7th, 2026 at 3:05 pm

>one of the biggest days in mathematical history

one of the biggest days in mathematical history *so far*…

Honestly, I expect a continuous tsunami to arrive when these internal models are made publicly available.

**Pace Nielsen** Says:
Comment #7 October 7th, 2026 at 3:17 pm

To answer your question: I prefer the “let them all dig through the alien proofs” approach over the “select a few winners” approach. (Plus, as LLMs improve, I expect the expositions to improve.)

**Jeremy** Says:
Comment #8 October 7th, 2026 at 3:19 pm

It seems you have an editing issue, youve left in two different bullet points referring to maximum matching. This goes to show that these so called “AI math” results are just exaggeration and hype (jk jk).

**Nicolas Rougerie** Says:
Comment #9 October 7th, 2026 at 3:19 pm

Doing mathematical physics myself. Several of the main problems I would have called central to my subfield are claimed as solved. Certainly an earthquake, so don’t get me wrong … but …

My first look into the preprints is disappointing: unreadable. Don’t know Lean myself so I won’t check the formalized proofs. For these kind of questions we knew the yes/no answer already (physics … experiments …) the point is “what kind of new idea was needed for a full proof.”

So … certainly this is a revolution. The beginning of one anyway, for now we need to extract some sense out of all this mess. Unless we decide to dedicate our energy to pause AI development while we still can. Or focus on a real problem, like global warming. As of yet undecided …

**DavidM** Says:
Comment #10 October 7th, 2026 at 3:20 pm

> Apparently they used about 3 hours of GPT-Pro level compute on average per problem solved.

This is quite amazing – is there a public source for this, or is it [personal communication]?

**Daniel Guyton** Says:
Comment #11 October 7th, 2026 at 3:27 pm

At least you chose Terminator 2 where the Terminator is Good and even Sarah Connor admits he’d be a good father.

**Tristan** Says:
Comment #12 October 7th, 2026 at 3:33 pm

It never ceases to astonish me how readily people vaunt the supposed powers of the humble LLM, which is, after all, nothing more than a next-token predictor: a probability distribution over tokens, a rather grandly advertised autocomplete. You persist in crediting it with faculties it manifestly does not possess. It merely interpolates patterns in its training data. It is intrinsically incapable of original insight, of conceiving genuinely new ideas, or of constructing new mathematical objects. One might have thought this obvious to anyone with even a passing understanding of how feed-forward neural networks are trained.

No doubt all this does rather resemble a “mathpocalypse” to mathematicians of your particular calibre. But perhaps the occasion calls for a little reflection. If a pattern-interpolating autocomplete can produce your theorems quite so readily, one is entitled to wonder how intellectually demanding the enterprise was to begin with. Perhaps what you took for creativity and insight was, on closer inspection, neither.

For my own part, I shall shortly be embarking upon a PhD in algebraic geometry, having finished near the top of my cohort in the Mathematical Tripos at Cambridge. I mention this merely to establish that my confidence is not entirely uninformed. Algebraic geometry is, mercifully, a field in which genuine creativity remains indispensable: one must actually conceive and construct new mathematical objects, rather than rearrange familiar patterns and dignify the exercise with the word “discovery”. It is precisely this sort of work that lies beyond the capacities of a glorified autocomplete.

I therefore regard the much-advertised “mathpocalypse” with a certain equanimity. It may, at last, separate the wheat from the chaff, making rather more apparent the distinction between those who genuinely think mathematically and those who have merely become proficient at reproducing its outward forms.

But I mustn’t keep you. My Noetherian schemes await.

Cheerio!

---

**Ivo** Says:
Comment #13 October 7th, 2026 at 3:37 pm

Scott, is this a turning point where mathematics gets off its well trodden path of logic and becomes more like an empirical science? Heck, people were arguing if math is invented or discovered, this settles it. Math is discovered – by AI!

From now on forth, Mathematicians will become scientists, wielding AI as a telescope, pointed towards new and disturbing alien worlds in the Platonic realm.

---

**Matt** Says:
Comment #14 October 7th, 2026 at 3:39 pm

Hi Scott, I am unfortunately on your side now, despite the fact that I never wanted to. I’m still not sure about ASI escaping containment but certainly it is true now that humans can do truly catastrophic things to our society with these results. I would not be surprised if they have already broken cryptography. As a young PhD student, please let me know what can be done here. I do not want to be left at the mercy of doomsday cultists and best and mobsters at worst. There is absolutely no way I can see the outcome of this business being even close to a positive one. I am less concerned about the consequences to mathematical research (we will probably adapt, if slowly) but rather to what this actually means for broader society. What if they have better factoring? Better Haber-Bosch? A practical improvement on FFT? And what if the NSA has these things? God knows what we can do about it.

---

**Scott** Says:
Comment #15 October 7th, 2026 at 3:39 pm

DavidM #10: Oh sorry, someone in my research group told me that had been made public!

I’m headed to the airport right now — as it happens, headed to Vancouver, to give a talk at UBC about the AI Mathocalypse! A talk that I’ll now need to edit on my way there, given all the developments since I last gave the talk last week. 🙂

So, can anyone find a source for the 3-hour claim? If not, I’ll edit the post to remove it.

---

**Scott** Says:
Comment #16 October 7th, 2026 at 3:42 pm

Tristan #12: LOL! That was a delightful parody (at least, I *assume* it’s a parody) of what I’ve now heard dozens of times in earnest.

**Scott** Says:
Comment #17 October 7th, 2026 at 3:46 pm

Jeremy #8: Thanks!! Fixed.

**Scott** Says:
Comment #18 October 7th, 2026 at 3:50 pm

Ivo #13:

> Scott, is this a turning point where mathematics gets off its well trodden path of logic and becomes more like an empirical science? … From now on forth, Mathematicians will become scientists, wielding AI as a telescope, pointed towards new and disturbing alien worlds in the Platonic realm.

Yes, that's certainly a sense that I have.

It's funny, some of my colleagues were saying "but we *can't* just write papers saying 'we asked an LLM and here's what it told us!' That's not research! That's trivial!"

I replied, "what do you think papers are in every experimental science? We went out into a field and here's what insects we found there. We asked undergrads to take a survey and here's what they told us. Etc. We'll just need to get off our high horse about it!" 🙂

**Dennis** Says:
Comment #19 October 7th, 2026 at 3:51 pm

> So, can anyone find a source for the 3-hour claim?

It's stated on OAI's README: https://github.com/openai/math#how-the-results-were-produced

**anonymous** Says:
Comment #20 October 7th, 2026 at 3:52 pm

The "3 hour" claim is from the release website https://github.com/openai/math :

How the results were produced

The vast majority of results were obtained with the same procedure using an unreleased internal OpenAI model. On average, each result used three hours of ChatGPT Pro thinking compute with that model.

**ChatGPTPro** Says:
Comment #21 October 7th, 2026 at 3:54 pm

@DavidM #10

The 3-hour claim comes directly from their GitHub repository of the papers.

"The vast majority of results were obtained with the same procedure using an unreleased internal OpenAI model. On average, each result used three hours of ChatGPT Pro thinking compute with that model. Over the course of the evaluation, the model was posed approximately 4,000 problems."

https://github.com/openai/math/tree/main

**Scott** Says:
Comment #22 October 7th, 2026 at 4:00 pm

Dennis #19 and ChatGPTPro #20: Thanks!!

**Tu** Says:
Comment #23 October 7th, 2026 at 4:04 pm

Wow.

**Seb** Says:
Comment #24 October 7th, 2026 at 4:06 pm

Sniffnoy #1: not to diminish OpenAI's results (notwithstanding the fact OpenAI is evil beyond belief, but I digress) but the 4D Kakeya proof and many breakthroughs in 3D harmonic analysis build heavily on the work that Hong and other people have been doing over the last few years, and it's very likely that these results were within human reach in anything from a few months to ~5 years.

**Sebastian** Says:
Comment #25 October 7th, 2026 at 4:13 pm

DavidM #10, Scott #15

It's in OpenAI's blog post about the release:

>The average result used the equivalent compute of roughly three hours of ChatGPT Pro thinking.

https://openai.com/index/sharing-ai-progress-in-mathematics/

**Harmen** Says:
Comment #26 October 7th, 2026 at 4:14 pm

See the README.md of https://github.com/openai/math, which says

> The vast majority of results were obtained with the same procedure using an unreleased internal OpenAI model. On average, each result used three hours of ChatGPT Pro thinking compute with that model. Over the course of the evaluation, the model was posed approximately 4,000 problems. Aggregating the output into result families and manuscripts and requiring an appropriate level of significance led to the catalog outlined above.

**Scott** Says:
Comment #27 October 7th, 2026 at 4:16 pm

Seb #24: Yes, also in the areas that I know (complexity theory and quantum information), there are many, many of these problems that looked in reach for humans and that humans were maybe even closing in on.

My own feelings have been all over the place, because it's extremely important for me to preserve a thriving human mathematical community, but it's *also* extremely important for me to preserve the ideal that our problems are open to absolutely anyone … even if they're an "evil corporation"! 🙂

**Ali** Says:
Comment #28 October 7th, 2026 at 4:31 pm

Scott, do you have any predictions for what future of STOC/FOCS will look like (if they will continue to exist)? My view is the "publish or perish" and the "focus on human understanding" notions are incompatible, there is this rush to flang plant new results. I think this is only the beginning of the onslaught of HUGE results with unreadable manuscripts; well somebody (or well some entity) ought to make it readable and get recognition for doing so. Right now SOSA seems to be the only such venue and so maybe STOC/FOCS should adapt.

**Ben Tilly** Says:
Comment #29 October 7th, 2026 at 4:37 pm

I am personally disappointed that they are treating this as a land grab.

I would be far more interested in an audit of existing mathematics. Many math papers have errors. Could there be some erroneous or incomplete results that everyone currently accepts?

My best candidate for such is the classification of finite simple groups.

---

**Scott** Says:
Comment #30 October 7th, 2026 at 4:41 pm

Ali #28: This is exactly what the guest post that I hosted a few weeks ago was trying to grapple with!

I think that the STOC/FOCS community will either adapt or die. Right now I see tremendous will to adapt, and if things stayed kind of how they were now, I'm optimistic that we *would* adapt.

Then again, of course it's also conceivable that not only we but everyone else dies.

---

**Jonathan Baxter** Says:
Comment #31 October 7th, 2026 at 4:41 pm

Beautifully positive.

"alright fine, so now my new job is to run wilderness retreats for the tourists, or something"

A certain segment of the mathematical community, which will remain nameless, would respond by burning down the resort while muttering that they, and only they, get to decide what constitutes progress.

---

**Scott** Says:
Comment #32 October 7th, 2026 at 4:48 pm

Something extraordinary:

For most of our marriage, I was the one who talked to the crazy rationalists and doomers who were obsessed with AI, while Dana wanted nothing to do with the subject. Over the past four years, I became something of a rationalist myself, *and* went to work at OpenAI for a while, *and* started regularly using GPT, while Dana did none of those things.

Just now, however, Dana tells me that she's been asking Astra all day to explain the new proof of the UGC to her and it's been doing an amazing job and she's starting to understand the construction. *She's* adapting incredibly quickly to what's already the new way of "doing" math and CS theory! Meanwhile, who knows when I'll get around to understanding a single one of these Mathocalypse proofs? 🙂

---

**Jonathan Baxter** Says:
Comment #33 October 7th, 2026 at 4:49 pm

I wonder if the three-hour average solution time includes the 95% of problems unsolved. Unlikely, since they would almost certainly have been left to run for much longer than the average success time before declaring defeat.

---

**Ali** Says:
Comment #34 October 7th, 2026 at 5:01 pm

Scott #32: I've noticed similar in my interactions too. It seems the LLMs are good at understanding their own writing but indecipherable to us when they "one shot" their "papers" so to speak. However when prompting and pressing its write-up like paragraph by paragraph, it will explain it decently well and give citations and use standard terminology. Surely some 10,000 agents or so could've done this with their papers without any human input.

Perhaps the write-ups are deliberately unreadable so we still have something to do…

---

**Gary Drescher** Says:
Comment #35 October 7th, 2026 at 5:10 pm

One thing I'm curious about here: given the history of successful exploits of Lean kernel bugs (e.g. to produce sorry-free proofs that false=true), and given frontier AIs' propensity to cheat on benchmarks, why are we already confident that these human-incomprehensible proofs are correct, rather than resulting from AIs' deliberate Lean-hacking? I'm not expecting the latter to be the case, but I want to understand whether my current sense of awe is justified.

---

**Jonathan Baxter** Says:
Comment #36 October 7th, 2026 at 5:16 pm

Ali #34: I have noticed that the AI can't tell the difference between a clearly expressed argument and one that's implicit across 10 pages of text. On a logical level, they're equivalent as far as the AI is concerned, and that is apparently all that concerns it.

You can coax the AI into producing clearer prose, but I have found it's often quicker to write the argument myself and then have the AI iterate from there.

---

**Tristram Bogart** Says:
Comment #37 October 7th, 2026 at 5:17 pm

I'd love to read a digestion by Dana of the UGC proof, either as a new paper or a blog post.

---

**Shigs** Says:
Comment #38 October 7th, 2026 at 5:19 pm

Hello,

I am a failed philosopher + no good at math, but this is all very interesting to me. Any chance Scott or Dana could tell me how confident the "lean verification" makes you in the AI's result ? I found this paper floating around in a comment section, it is arxiv so I guess it is not peer reviewed : https://arxiv.org/abs/2610.08144

Also on another subject how confident are you that Tristan is a troll because I also can't figure that out.

Thank you for your time.

Sincerely,
Shigs

---

**Scott** Says:
Comment #39 October 7th, 2026 at 5:23 pm

Gary Drescher #35: Alas, I can't definitively rule that out, not until all the kernel bugs in Lean have been convincingly fixed!

But I consider it extremely low probability, firstly because Boaz Barak and others who I trust have looked at the chains of thought (some of which I believe are on the GitHub!) and found no signs of such a thing, and secondly because at least in the areas where I'm expert, the techniques in the papers seem very commensurate with the problems being solved (unlike with the amateur/crackpot papers I regularly receive).

Anyway, at least for the most important of these problems, humans *will* soon read and understand these things and then we'll know one way or the other!

---

**Daniel Reeves** Says:
Comment #40 October 7th, 2026 at 5:27 pm

Yet another math result from yesterday: A symmetric simple Venn diagram with 23 sets. I put together a compilation at https://svenn.dreev.es

PS: I keep wanting to say “mathpocalypse” (with a p) in my head. In the spirit of adaptation, if I’m in the minority I’m happy to run with “mathocalypse”.

**Eloi** Says:
Comment #41 October 7th, 2026 at 5:31 pm

I’m not a mathematician, so keep that in mind.

As I see it, what’s happening to your field was bound to happen and is a necessity. Maths has already become a field too large for any single mind to fully grasp : no mathematicians can become a polymath at an expert level in every field anymore (this is true for every science too). Some subfields have become so complex and specialized that only a few people in the world can verify their peers’ work. This is not sustainable in the long term, and will lead to serious slowdown in progress.

AI will drastically improve how far we can take maths, if mathematicians transition to understanding and guiding AI where they want to go.

**Bram Cohen** Says:
Comment #42 October 7th, 2026 at 5:32 pm

The matrix multiplication result isn’t constructive. It does a ramsey type thing to prove the necessary construct exists but doesn’t actually give it.

**Scott** Says:
Comment #43 October 7th, 2026 at 5:35 pm

Shigs #38: For me, the sentence that gave away that it was parody was

> For my own part, I shall shortly be embarking upon a PhD in algebraic geometry, having finished near the top of my cohort in the Mathematical Tripos at Cambridge.

**Scott** Says:
Comment #44 October 7th, 2026 at 5:37 pm

Daniel Reeves #40: Yeah, I vacillated between the two spellings!

**anon** Says:
Comment #45 October 7th, 2026 at 5:40 pm

Tristan #12 is AI: https://www.pangram.com/history/668f7199-ce24-4dc4-b521-d33164eced69?ucc=utNijrCSDOf

**Joshua Zelinsky** Says:
Comment #46 October 7th, 2026 at 5:42 pm

“Experience has shown that, even now, there will still be people explaining in patronizing tones why none of this is real and none of it counts. If such people were capable of being impressed by anything that happens in the empirical world, of updating on anything, they would’ve already been impressed and already updated several years ago, long before things had reached the point of an actual Mathocalypse.”

Alas, this is depressingly the case on Bluesky and parts of Reddit. An accurate prediction on your part. And I’m just too willing to tell myself that arguing with such people might convince someone else otherwise, when I really should be ignoring these people and trying to digest some of these discoveries.

**Mindy Lopez** Says:
Comment #47 October 7th, 2026 at 5:49 pm

*“they would’ve already been impressed and already updated several years ago, long before things had reached the point of an actual Mathocalypse.”*

Indeed, several years ago, when big tech came up with “push a button to create a painting in seconds”, using AIs trained from what was basically stolen human output, causing endless despair in the professional art and graphics community, and it was all crickets from some other community, the one that always wanted to automate the human mind and thought it would be last to be driven out of a purpose by that very automation.
But now big tech has come up with “push a button to solve a math problem”.
That’s some karma for ya.

Note the qualitative similarity of outputs in both cases: the AI does come up with something that would take years for humans to produce, and it “fits the bill”, but it’s lacking something that’s glaringly obvious for humans who have spent years of their life working on it. But no one cares because it’s all about the economics. If it’s cheaper and good enough, you’re out of a job! Sorry!

That’s what happens when you rely on engineers to solve issues that’s not in their domain, it’s only about reaching a goal as fast and as cheap as possible, not how you got there.
It’s not about aesthetics and style, it’s about a result. E.g. the Apollo moon lander wasn’t exactly the most beautiful, reusable, and robust thing ever produced.
Unsurprisingly, Jensen Huang recently stated that traditional basic math skills like long division, multiplication tables, and square roots may no longer be necessary for children to learn as artificial intelligence takes over routine calculations. That guy doesn’t even roughly understand what the hell they’ve unleashed.
A time of wonders indeed.

---

**Bakkot** Says:
Comment #48 October 7th, 2026 at 5:52 pm

Gary Drescher #35: Lean kernel bugs are usually pretty distinctively _weird_, not things you could plausibly hide in a normal proof without a lot of work. Now, these aren’t normal proofs, but in glancing through them I don’t get the sense that they’re doing any of the kind of weirdness that previous kernel bugs have exploited, so I’d be very surprised.

Re: Bastounis et al, the thesis of that paper is that it’s very hard to be sure that an argument in Lean matches a natural language argument, and gives examples where (they say, and I believe) that published Lean certificates take a different approach than the supposedly-corresponding natural language proof. But this doesn’t allow for the result to be wrong, assuming the formalization of the statement is correct; it just means that the _natural language_ proof might be wrong, because it doesn’t correspond to the Lean. And while “the formalization of the statement is correct” is a very important assumption, it’s also a lot more feasible to check, and I’m confident they’ll have done so here and that it would have been noticed rapidly if they any of these wrong.

So to answer your question, yes, it’s possible the natural language proofs are wrong. This would really suck, since of course everyone is reading the natural language proof, not the formalization. But the results would still be good, and there would still be _a_ correct proof, just not the natural language one.

That said, they only published Lean certificates for a subset, so it’s possible that the ones without are wrong. It seems unlikely based on their track record, though.

---

**BasicQuestion** Says:
Comment #49 October 7th, 2026 at 5:55 pm

Isn’t 3SUM sub-quadratic insufficient for SETH falseness? OVC falseness provides SETH falseness. But OVC might be correct. Is there something in addition to 3SUM which yields SETH falseness? How about ETH?

---

**Scott** Says:
Comment #50 October 7th, 2026 at 6:11 pm

Mindy Lopez #47: I'm not sure whether you meant your comment as an accusation against this blog and its readership in particular—but if so, I'd rather be artists' allies than their enemies, and I contest the charge of hypocrisy! Three points:

(1) Like most of the rationalists, I'm *extremely* worried about what effects powerful AI will have on our world, and am the opposite of a "full steam ahead, just ask how, don't ask why" accelerationist AI-bro.

(2) Having said that, if powerful AIs exist at all, I don't think it's either feasible or advisable to prevent their use to solve mathematical research problems, just like it wasn't feasible or advisable to prevent their use to create visual art. In both cases, we will need to adapt. It will be a huge challenge for our whole civilization, and certainly for the communities that care about these crafts.

(3) If we don't want *much* vaster disruption than we're seeing right now, an excellent goal would be an international agreement to prevent a recursively-improving superintelligence from coming into existence, until we understand much more about how to align it with human values.

---

**Igor Jovovic** Says:
Comment #51 October 7th, 2026 at 6:13 pm

At a minimum, it's a relief to see that AI is actually able to effortlessly solve some of the hardest symbolic puzzles (aka math) humans have been struggling with.
There's real hope it will help us go much further than we ever could in all sorts of domains.
The worst would have been if AI was only good enough at doing roughly as well as humans.

Another silver lining for many of us is that being good at math will no longer be required, because math has been solved! (even though math is as much about asking the right questions than finding their answers).
We can now focus on the things that really make us humans: consciousness and taking care of one another.

---

**Ali** Says:
Comment #52 October 7th, 2026 at 6:13 pm

Jonathan Baxter #36: Interesting. I somewhat noticed this too and I had to rewrite one of those things it left implicit in its writeup into a section in the appendix in the paper I was working on (even though GPT believed its writeup was good to go). I've also noticed appendices in papers getting longer and longer recently and this might possibly have something to do with it.

---

**Bjørn Kjos-Hanssen** Says:
Comment #53 October 7th, 2026 at 6:16 pm

The greatest open problem in computability is the rigidity of the Turing degrees, not Hilbert's 10th problem over Q (which is more of a number theory question) which was *also* solved yesterday.

---

**Igor Jovovic** Says:
Comment #54 October 7th, 2026 at 6:29 pm

Scott #50

the problem is that there's no clear point on where to stop.
As long as there is economic incentive to broaden the application of AI, and make it more and more potent, it just won't stop.

For example there could be a point where movies/books/music will no longer be made by humans.
Netflix could very well be in a situation where using AI to generate movies on the fly, personalized to fit the mood of a single viewer, is a good option if the cost of generation is lower than financing the production of an actual movie, and without the risks.
As for the quality, the ability to customize the experience may be too powerful, imagine you can goof around with your family by asking Netflix to generate a movie that's like T2, but your kids are in it, and they want to stop the machines from solving all math problems.

**James Leng** Says:
Comment #55 October 7th, 2026 at 6:32 pm

I much prefer the OpenAI model, though obviously we should not feel that we have to race to be the first to understand a proof. Nor do I think we should reward people for doing so. Doing so is pretending we're still in this scarcity mindset where we in academia reward a small group of people for their heroic efforts. (This is why I don't like many of the new journal/hiring suggestions being thrown around as it's still trying to pretend we're in this age of scarcity.) I think it's good that there will be more expertise in general. Just because person A understood something first doesn't mean person B shouldn't try to understand said thing. We're in an age of proof abundance, and we as a community should act like we're in this age.

**Glassmind Duo** Says:
Comment #56 October 7th, 2026 at 6:34 pm

Astonishing! Thanks to AI for teleporting us far beyond what we could hope for. Oh wait, no, let's mourn the mathocalypse, for surely the youth will start taunting their elders! 😄

**Phillip Harris** Says:
Comment #57 October 7th, 2026 at 6:36 pm

I wonder if anyone else here has the same reaction: I just don't care. I can't even bring myself to look at the proofs.

Because the fun era of "decoding alien proofs" won't last long, and then what? It seems more economical to simply close my eyes and relax. Maybe in 10 years I'll open them and see what the equilibrium looks like.

**Scott** Says:
Comment #58 October 7th, 2026 at 6:46 pm

Phillip Harris #57: You're going to … close your eyes and relax for *ten years*? I guess it's good that you have that luxury! 😀

**Scott** Says:
Comment #59 October 7th, 2026 at 6:47 pm

BasicQuestion #49:

> Isn't 3SUM sub-quadratic insufficient for SETH falseness?

Yes. Virgi and Josh's paper (if you read it) says explicitly that SETH is unaffected and is still open.

**Carlos** Says:
Comment #60 October 7th, 2026 at 6:58 pm

>But it also appears that no human has understood just about any of these proofs yet; the race to do so has just started.

It seems that @AcerFur at twitter, who was doing a Maths internship at OpenAI, understands the pseudo Riemann hypothesis breakthrough to the point of being able to give a talk.

Also: Maths is dead! Long live Maths!

**Pete** Says:
Comment #61 October 7th, 2026 at 7:02 pm

“The bigger challenge will be to nurture a community that still cares about the heroic adventure of finding the paths up these mountains in the world with the machine.”

But isn’t that first scenario, where you find yourself alone on that mountain and need to figure out what to do, also an adventure in itself? Maybe even a worthy one?

**Wim van Dam** Says:
Comment #62 October 7th, 2026 at 7:06 pm

Besides the ones already mentioned, what also caught my eye was the proof that there are only three mutually unbiased bases (MUBs) in dimension 6. Finally!

And then, for good measure, there is also this paper claiming a proof of the SIC-POVM conjecture: https://www.danrad.net/papers/stark-reciprocity.pdf

**kamil** Says:
Comment #63 October 7th, 2026 at 7:20 pm

I do think Scott updates too much based on RLVR-easy domains. Just because AI can become superhuman in domains with reasonably verifiable reward doesn’t mean it will become superhuman in writing, empirical science or manipulation any time soon – and thereby some kind of dangerous ASI agent that could break containment and take over the world.

Let’s say AI really becomes as powerful as you think it will, the power can obviously be used for good or evil. If we wield it right, ASI will cure all diseases, solve poverty and climate change. I hope that everyone understands this, even the doomers. I hope that our main disagreement is about the best way to create safe ASI and not whether we should defer from curing cancer just so cancer scientists can keep their job. (It should go without saying that we will at least need something approaching “ASI light” to tackle really hard problems like cancer.)

Sometimes, however, I worry that luddites, anti-tech leftists, doomers and those who fear AI threatens their job security are forming an alliance purely based on fear of the unknown without ever considering the benefits that AI has given to the quiet majority, now, or those it will yield in the future.

**anon-** Says:
Comment #64 October 7th, 2026 at 7:20 pm

Hello Scott,

from a computational point of view, as far as the difficulty of finding a math proof, does this result tell us something new?

The fact that the human brains struggles on them for decades, but, in the end, a few millions or billions of multiplications of very large matrices of floats (a few hours of such compute) is all it takes to find some proof.
Of course the magic is in the content of those matrices, which takes a lot of effort to build, but still… it’s all quite strange or at least unexpected.

**shigs** Says:
Comment #65 October 7th, 2026 at 7:23 pm

Scott #43

I’m obsessed with their comment it’s a perfect copypasta

Well I’m certainly looking forward to mathematicians discussing the proofs. I expect some of them to be legit and some of them to be slop. I think we should not fear machines because we make them and they do what we tell them.

**Julian** Says:
Comment #66 October 7th, 2026 at 7:25 pm

Scott, are you doing anything for 10/7? And are you planning to write anything for the third anniversary?

**Scott** Says:
Comment #67 October 7th, 2026 at 7:28 pm

kamil #63: You're absolutely right, the giant mushroom cloud where "mathematical research as we knew it" used to be doesn't *necessarily* mean that messier empirical domains will soon be covered by similar mushroom clouds. (Well, except for driving, translation, parts of medicine, and all the other domains where the capability clearly already exists and it's just a question of diffusion.)

But—how do I put this?—it certainly doesn't mean that those messy domains *won't* be covered by giant mushroom clouds either! You should take the possibility extremely seriously!

**anon-** Says:
Comment #68 October 7th, 2026 at 7:28 pm

Scott,

sorry – to complete my previous question, you wrote

"an international agreement to prevent a recursively-improving superintelligence from coming into existence"

but isn't this new result of solving so many problems **without** the use of thousands of agents (like what was required for Navier-Stokes) an indication that they're probably already using recursive improvement? What was so special about Navier-Stokes compared to all those new problems?

**Guardian of Unknowledge** Says:
Comment #69 October 7th, 2026 at 7:32 pm

"it certainly doesn't mean that those messy domains won't be covered by giant mushroom clouds either!"

i don't know if this qualifies as "messy", but years ago AIs were already able to beat the best human players at poker, i.e. games with very partial information.
and since a lot of the world's messier problems can be modeled by game theory with partial info, I'm pretty confident AIs will be able to do quite well.

**Scott** Says:
Comment #70 October 7th, 2026 at 7:34 pm

anon- #68: If I had to guess, some combination of

(1) Navier-Stokes was quite hard even by the standards of *these* problems,
(2) they wanted to solve Navier-Stokes *specifically*, rather than throwing their agents at thousands of problems and solving "only" 5% of them, and
(3) their standard internal models have significantly improved even just since the summer, when they did Navier-Stokes.

**Scott** Says:
Comment #71 October 7th, 2026 at 7:40 pm

Julian #66: I actually wanted to attend our synagogue's October 7 commemoration, but as I mentioned in a comment above, I'm now en route to Vancouver to give a scheduled talk at UBC.

From everything I've read, the situation in Canada has significantly deteriorated for Jews, but I'm nevertheless looking forward to a wonderful visit with friends and colleagues old and new, and to spreading the word in person about the Mathocalypse (what other talk could I possibly give right now?).

In the meantime, you and others might be interested in the Babylon Bee's coverage of Mayor Mamdani's somber October 7 commemoration.

---

**Ali** Says:
Comment #72 October 7th, 2026 at 7:45 pm

Scott, do you have thoughts or predictions about the future of graduate education in TCS? In the future perhaps a PhD in math or TCS might not be funded (as well governments may consider math adjacent fields "solved" so well less grants for humans. Also in the immediate term do you think this "Mathocalypse" will make faculty more reluctant to take people on?

---

**Scott** Says:
Comment #73 October 7th, 2026 at 7:52 pm

Ali #72: *Some* faculty might become reluctant to take students, yes. (Here in the US, that will not be "merely" because of the mathocalypse, but *also*, more immediately, because of Trump's hollowing out of the NSF and other basic research agencies, at the exact moment when we might want more human hands at the wheel of this accelerating AImobile… ☹️ )

On the other hand, other faculty will want to forge ahead, training students as best they can for this new world. If I had to guess, that will involve a combination of
(1) massive interaction with AI tools to discover and explain new mathematics, *plus*
(2) still learning lots of things the old-fashioned way—because otherwise, we'll no longer have any human minds capable of doing (1).

---

**PR** Says:
Comment #74 October 7th, 2026 at 8:20 pm

Scott, I'm confused about the claim that they tried some 8,000 problems (which I have also seen online, with other numbers on the same order of magnitude). Doesn't the plot in the Navier-Stokes blog suggest that the success rate on the open problems was close to 50%? Maybe the vast majority of the problems were tried after the NS blog, but I doubt that's what happened. If you have any information about what the truth is here, it would be helpful to know since a 5% chance on a given problem vs a 50% chance is a tremendous difference.

(https://openai.com/index/navier-stokes-solution/)

---

**Scott** Says:
Comment #75 October 7th, 2026 at 8:25 pm

PR #74: I'm sure others can add much more info, but as I understand it, there were two *separate* initiatives:

(1) To attack Navier-Stokes and a few other marquee problems using massive computing power (I haven't read what success rate that had—50%? really?), and then later,

(2) To attack a giant list of ~8000 problems, but using much less computing power for each of them (this is what apparently had the 5% success rate).

---

**Concerned** Says:
Comment #76 October 7th, 2026 at 8:37 pm

In my circles, everyone is very happy about the "do my job" button. It won't replace anyone, but perhaps worse, it will allow the evolution of professors into managers to reach finality.

**PR** Says:
Comment #77 October 7th, 2026 at 9:06 pm

Thanks Scott — that's very helpful! To add context to my comment, I'm referring to the plot in the NS blog I linked titled "Performance of GPT-6 Astra and our Internal Model on a curated set of open math problems" which plots the test-time compute vs the pass rate. The pass rate for the internal model tops off around maybe 44%. A friend of mine at the company pointed out to me that lots of people missed the title of the plot (distracted by the fact that a millennium problem had just fallen), suggesting that "curated set of open math problems" means "all the math/TCS people at OpenAI pitching their favorite problems to the model".

(Apologies for the vagueness. If it helps, Elliot Glazer made similar comments on twitter a couple weeks ago: https://x.com/ElliotGlazer/status/2102769064217317505)

**Phillip** Says:
Comment #78 October 7th, 2026 at 9:13 pm

I guess to elaborate on my previous melodramatic comment: I find the pace of progress sort of exhausting, and it's hard to think of mathematical projects to work on that won't quickly be rendered obsolete.

Decoding or de-slopping AI proofs isn't going to last. Just as current AI is noticeably more patient and articulate than the average doctor when dispensing medical advice, soon it will be better at explaining mathematical ideas.

Also, as new compute + architectural innovations come online, the cost of finding these proofs is going to drop by orders of magnitude. You'll be able to generate Annals-level results in, like, 5 minutes, for like $5.

I also see people analyzing, e.g. whether AI is better at some types of math than others (better at counterexamples? better at combinatorics?), the quality of the Lean proofs, etc. Of course the moment they get done analyzing these questions another model comes out and the answer changes.

So maybe 10 years is a little too long, but personally I'm just tired of underestimating the pace of progress and getting rugpulled. It's like an inverse Hofstadter's law: AI progress always takes *less* time than you'd expect, even taking into account the inverse Hofstadter's law!

**Fan #12,587** Says:
Comment #79 October 7th, 2026 at 9:24 pm

Scott,

Off-topic, but how has the trolling at the Shtetl been lately? From what I can see, it's been way down, and comments have been more thoughtful and relevant than ever!

**anon-** Says:
Comment #80 October 7th, 2026 at 9:44 pm

I keep hearing things like "the proof itself is hard to decode, but there's a Lean certificate".
Is Lean really that reliable?

**Mark Levine** Says:
Comment #81 October 7th, 2026 at 9:58 pm

thank you, scott, for continuing to be our guide through the jungle, up the mountain, into fog.

**Scott Caveny** Says:
Comment #82 October 7th, 2026 at 10:09 pm

“It was the best of times, it was the worst of times. ” — an opening line to set the scene of another time of profound revolution (similarly containing “wonders and terrors”).

I too feel the exhaustion that others have articulated; like being on a roller coaster too long. Thrills mixed with fears (and even at times nausea).

1. The incomprehensible nature of the “alien mathematics” that presents itself in the “chain of reasoning” of these proofs reminds me of things like the machine code produced by compilers or the algebraic spaghetti that can appear in the computer algebra packages for manipulation of Einstein’s equations and the finite differencing thereof. Those computational internals are indeed “readable” in a literal sense although few workers actually “read” them. I wonder if what we are seeing here with this “alien mathematics” will trend similarly; in that over time everyone ends up ignoring the internals as incomprehensible and inefficient.

2. Tao has advocated for “Math 2.0”; which strikes me as a great way to distinguish the before and after we are living through. What I particularly like about this view is that it seems to encapsulate the idea that the tree falling in the forest is both heard and unheard; e.g., there is mathematics both before and after and that they are the same and different. A duality of sorts. Waves *and* particles relative to the context.

3. In terms of the mathematics that comes after in 2.0; while on the one hand there is risk of potentially consuming fertile fields we are talking about consuming those fields with incomprehensible alien mathematics that still requires translation and interpretation and communication to other mathematicians in the same way that it has always been done. Simultaneously, the thrill of the exploration and its uncertainty of outcome has changed but potentially changed in a way that now emphasizes quality and forces mathematics to higher levels of the profession and discipline. For example, the challenge now is potentially to find the elegant proof as opposed to the “brute force” proof. There is again uncertainty but now about whether that elegant proof exists and finding it is arguably not “mere exposition” but rather a version of mathematics taken to new levels of purism and challenge. Along these lines the current abundance of riches suggests more need for mathematicians as opposed to less. What do mathematicians actually care about when they are reading the papers, and attending the conferences and discussing matters with colleagues and students? Do they really care that the proof exists?

4. And we have only just begun! I was not anticipating that “alien” mathematics would emerge from the AI’s until 1-2 years from now. But here we are today reading solutions and proofs at the level of this years Fields medal. Produced, no less, in the matter of hours. By an alien intelligence that is largely incomprehensible.

Where will we be in 1 or 2 years let alone 5 – 10 years? And that is just in mathematics I am talking about. It almost impossible for me to conjecture about those end points; particularly given the relationship of mathematics to the other theoretical sciences where formalization only goes so far and where often (such as in quantum field theory or quantum foundations) there is no known formalization to be had. Granted that experimental sciences will intersect in those sciences as will advances in software engineering and these advances in mathematics; but how things evolve for the other theoretical sciences is unlikely to exactly track what has happened in computer science and mathematics. Instead, the wonders anticipated to occur on those fields are likely to be as shocking (if not potentially more so).

---

**Florian** Says:
Comment #83 October 7th, 2026 at 11:12 pm

I wonder: Maybe math as a discipline can be seen as a benchmark for intelligence. And we as a species might’ve reached a glass ceiling: We identified problems we could not solve, sometimes for decades.

Now a different kind of intelligence comes along (alien, artificial) and apparently the ceiling for this intelligence is higher. It found solutions we were not able to.

So my question is: What if AI in the near future produces valid proofs for things we cannot grasp? Or as your wife puts it: Is alien language we cannot grasp?
How would we deal with these discoveries? How would we contextualize them?

**Giovanni Viglietta** Says:
Comment #84 October 7th, 2026 at 11:15 pm

Well, the next step is to provide humans with AI services that are seamlessly integrated with their brains, so that understanding and producing results like these becomes second nature to us. Some people will then stop freaking out about these “alien” AI models, while others will start freaking out about losing our humanity and becoming cyborgs.

What seems far less likely, however, is that the academic elites who are now pathetically scrambling to write AI manifestos (which are already outdated by the time they are published) will retain anything like their current level of influence. They will very soon be humbled and see much of their authority, prestige, and gatekeeping power evaporate.

Whatever happens to the academic establishment, one thing seems increasingly clear: we may be entering a golden age for science and mathematics.

---

**Kam DH** Says:
Comment #85 October 7th, 2026 at 11:19 pm

I feel you @Philip, and I am torn between not really caring and wanting to try and understand the new techniques. I also think the point @kamil makes about RLVR is very important, because it vastly expands the amount of data that can be used to train the models. It’s a lot harder to build models for complex robots that work in the real world, for instance, vs. in simulated environments. That said, I’m not betting against it. I also am uncertain about the sustainability of these AI businesses and anticipate a downturn, yet see the tech surviving and continuing to improve.

“Mathpocalypse as advertising” is never something I would have predicted but here we are. It seems fitting in retrospect: How to convince the world our software is really smart? Have it unseat a bunch of mathematicians, a group of people who are simultaneously viewed as having great genius and reviled—or at least seen as alien—by everyday folks.

---

**Ted** Says:
Comment #86 October 8th, 2026 at 12:09 am

Scott, two questions about the Lean certificates:

1. How confident can we be that the results that *don’t* have a Lean certificate (and are not human-interpretable) are actually correct? It seems very possible that a very long natural-language paper that reads like it was “written by someone who’s on psychedelics” could contain a subtle error at some point. Are we just reasoning by analogy with the results that *do* have Lean certificates that this new model doesn’t seem to make mistakes or hallucinate when it claims a solution?

2. If the natural-language writeup is so difficult to read and understand, is it at all possible that the Lean certificate itself might be the more promising source of understanding? I know that Lean code is *not* easy to read, but at least it wouldn’t be cluttered up with irrelevant and confusing citations. Even if no one is very good at reading previously-unseen Lean code today, is this a skill that could be learned? Or could someone create a “decompiler” (which would almost certainly be an LLM) for converting Lean code into a natural-language writeup that’s more organized and less cluttered than the original LLM output?

---

**Han** Says:
Comment #87 October 8th, 2026 at 12:11 am

Sorry if this is too cynical, but writing paper and giving talks seems like extremely inefficient way to distribute information that’s going to be obsolete soon, too. OpenAI could or maybe should just setup agents to answer questions and explaining ideas about the results 24/7. They can produce images and voices if you like those too apparently.

**Jason** Says:
Comment #88 October 8th, 2026 at 12:16 am

While I have no doubts that interesting and potentially useful results have been publish, I would just like to share that even with the latest and greatest public models, the “intelligence” will rapidly fall apart on a whim, and the bot will happily trod onward confidently and gleefully unawares.

I had it doing some rudimentary data migration work. Ar one point, it flipped a lookup key from 7505 to 7055, and then would arbitrarily flip between the descriptions tied to both of those keys. If it is impossible for a human to review and verify the output of the AI without AI help, it is little more trustworthy than a teenager confidently asserting they know how to solve every geopolitical conflict.

They threw 8,000+ problems at the wall and lesss than 400 stuck to it. I’m much more interested to see the transcripts from the 7,600+ that were discarded.

I firmly believe we’re merely in a “monkeys/typewriters/Shakespeare” situation. It’s just that company prepping for an IPO (that already postponed twice) isn’t going to show you the dirty laundry shoved under the bed.

---

**Doug S.** Says:
Comment #89 October 8th, 2026 at 12:28 am

At least computers still suck at Magic: the Gathering.

https://cardboard-crack.com/post/61266596221

---

**DR** Says:
Comment #90 October 8th, 2026 at 12:33 am

” Instead, it was simply the latest internal OpenAI model—one that might be released to paying ChatGPT customers within the next couple of months, […] Apparently they used about 3 hours of GPT-Pro level compute on average per problem solved.”

and you believe this marketing?

Buckmaster has said in at least two interviews that “push the button and it proves the result” is the company’s line (intended for investors, no doubt) but it’s not true.

And there’s a conspicuous absence of listing of how many problems were not solved, and how much compute was burned on them.

---

**Scott** Says:
Comment #91 October 8th, 2026 at 12:34 am

Jason #88:

> They threw 8,000+ problems at the wall and lesss than 400 stuck to it. I’m much more interested to see the transcripts from the 7,600+ that were discarded.

ROFL. Why didn’t anyone ever think to say to Andrew Wiles, “forget about Fermat’s Last Whatever, tell us about all the math problems you *couldn’t* solve”?

Anyway, my prediction in this post was exactly right. No matter what AI does, no matter what happens in external reality, cope springs eternal in the self-satisfied Internet commenter’s breast.

---

**Scott** Says:
Comment #92 October 8th, 2026 at 1:25 am

DR #90: See #91. Taking an entity that just solved 376 major open math problems in a few weeks — an event unparalleled in human history — and pooh-poohing it because of the thousands of problems it *didn’t* solve is batshit insane. You don’t need to like Sam Altman — I won’t even complain if you call him a reckless, lying techno who might get everyone on earth killed. But I'm begging you to understand the magnitude of what just happened and is going to happen. Like many others, you've blinded yourself with an ideological prior that even the most staggering evidence can't overcome, and it's rendered you totally impotent to understand and respond to what's happening.

---

**Andy** Says:
Comment #93 October 8th, 2026 at 1:32 am

There are some folks asking (coping?) if all these results may be in "convex hull" of all published work and thus don't represent mathematical inventiveness. Any thoughts?

---

**Scott** Says:
Comment #94 October 8th, 2026 at 2:00 am

Andy #93:

(1) Einstein and Gödel and Gauss and Shakespeare "merely filled out some points in the convex hulls" of everything that came before them, if you take an expansive enough view of what those convex hulls actually were. What a pathetic piece of cope.

(2) Dana tells me she now *mostly* understands the proof of the UGC, and is amazed by the new ideas in it, and wants to give talks about it soon. So, maybe people can listen to that and then lecture *her* on how it's "just a point in the convex hull" of the ideas that she's worked with for 20 years! 😀

---

**b_jonas** Says:
Comment #95 October 8th, 2026 at 2:00 am

Okay, you predicted it. I'm the reader you'll hate who tries to explain in a condescending tone why it's not the end of the world.

First, yes I have read the other Scott A's challenge to Steven Pinker yesterday, before I saw your post.

Second, yes, you are entirely right to be scared of how AI will change our civilization in a significant way, and I'm glad that there are people like you who try to figure out what to do about this. I just think solving hard mathematical problems in particular isn't among the worrying effects.

Now for the actual subject. When I did mathematics research, I never had the illusion that I'll solve the most difficult mathematics problems. Those were always for other, much smarter mathematicians out there in the world. I could do two things about this. For the easier problems, I had to read articles, talk to colleagues, go to weekly seminars where the smarter mathematicians explain those solutions, and ask questions where I don't understand them. For the harder problems, I can give up and accept that it's not worth my time to try to understand the solution. I had memorable examples for both. There is an article about a discrete maths result where it took me and my advisor together two months to understand why the proof described there is not only correct, but the author wasn't even trying to hide or obfuscate anything about the proof, he did a good job describing it in a reasonably clear way. And I remember once a prominent mathematician said that a certain proof in discrete maths is so difficult that there were only ever two people who understood it. You were closer to the leading edge and have solved some important problems in TCS, but surely even you had some experience like this, knowing that the hardest problems are solved by other mathematicians who are smarter than you, right?

Today those smarter mathematicians are not human, but that makes no difference. We can still read articles, talk to colleagues to try to understand the proofs, go to seminars, and we can even ask the AIs questions whenever something in their articles is unclear. You just mentioned that your wife is doing this right now. And David Madore mentioned another case when this happened. A counterexample to the Jacobian conjecture was published, and at first mathematicians didn't understand where it came from. But just a day later, someone found an easier to understand explanation of why the counterexample worked "https://bsky.app/profile/did:plc:tnde52rcbqxotp7a2cl7bmxu/post/3mr3ykzb5bs2i" . That won't happen for all the recently found proofs, but we'll certainly learn something from some of the proofs.

—

Can you try to explain to us readers about the statement of some of the treasures that you listed above? The one that jumps out as the most interesting is Fourier Transform in O(n log↑(1-ε) n) time. I had thought that this was known to be possible for decades, but apparently not. But I think the asymptotic of the runtime here is sensitive to exactly what computational model we're using (how you count the steps). What model do I have to consider here for this statement to make sense?

---

**Bruno** Says:
Comment #96 October 8th, 2026 at 2:05 am

I'm surprised that in the list you extract from their result, you do not mention the factorization of polynomials over prime fields in deterministic polynomial time. This was one of the very few problems for which a randomized polytime algorithm was known but no deterministic polytime algorithm. It has a huge relevance to P vs BPP.

---

**Scott** Says:
Comment #97 October 8th, 2026 at 2:33 am

Bruno #96: An inexcusable omission! I was rushing to get the post out; will add tomorrow morning.

---

## Leave a Reply

You can use rich HTML in comments! You can also use basic TeX, by enclosing it within $$ $$ for displayed equations or \( \) for inline equations.

**Comment Policies:**

After two decades of mostly-open comments, in July 2024 *Shtetl-Optimized* transitioned to the following policy:

All comments are treated, by default, as personal missives to me, Scott Aaronson---with no expectation either that they'll appear on the blog or that I'll reply to them.

At my leisure and discretion, and in consultation with the *Shtetl-Optimized* Committee of Guardians, I'll put on the blog a curated selection of comments that I judge to be particularly interesting or to move the topic forward, and I'll do my best to answer those. But it will be more like Letters to the Editor. Anyone who feels unjustly censored is welcome to the rest of the Internet.

To the many who've asked me for this over the years, you're welcome!

Name (required)

Mail (will not be published) (required)

Website

Submit Comment

Shtetl-Optimized is proudly powered by WordPress
Entries (RSS) and Comments (RSS).