Shtetl-Optimized

The Blog of Scott Aaronson

If you take nothing else from this blog: quantum computers won't solve hard problems instantly by just trying all solutions in parallel.

« I've been drinking wine in my kitchen
Every call »

HOLY SHIT HOLY SHIT HOLY SHIT HOLY SHIT

I feel like my previous posts (and post titles), and podcasts, and talks, have failed to do full justice to the magnitude of what's now happening, so I'm trying again.

Last night, we had friends and colleagues over for dinner. We were still trying to process OpenAI's and Anthropic's proofs of VP ≠ VNP—hot on the heels, obviously, of the Alpöge/Anthropic proof of the Riemann Hypothesis, which already feels like last week's news. But then before dessert was served, we saw that OpenAI claims to prove that ω, the exponent of matrix multiplication, is exactly 9/4—a **lower bound (!!)**, matching the upper bound that OpenAI proved in the great Mathocalypse drop of October 6.

There are two separate shocks here. The first shock is that we now, finally, know the right answer to this central mystery of algorithms going back to the late 1960s, when Volker Strassen first showed that $\omega<3$ (he got $\omega\leq\log_2 7$) … and that the true answer is actually a *rational* number (!). But the even bigger shock is that the AI was able to *prove* any lower bound of the form $\omega>2$. I never expected this challenge to be all that much easier than P ≠ NP.

Over the decades, I often heard colleagues express extreme confidence that $\omega=2$, and I expressed skepticism whenever this came up—not because I had any strong argument for $\omega>2$, but just because I saw that there was no strong argument for $\omega=2$ either. So, I'm glad to have been vindicated about *that*!

But wait, there was more. During dessert, we saw another OpenAI release proving that the true number of linear gates needed to compute the Fourier Transform is $N (\log N)^{1/2+o(1)}$. The upper bound here improves, not only over the $O(N \log N)$ that stood since the 1960s (i.e., the Fast Fourier Transform or FFT), but also over the $O(N \log^{0.9999999999999} N)$ (give or take some 9's) that OpenAI published back in the October 6 Mathocalypse. Again, though, the real shock here is that there's also a **matching lower bound (!!)**. All these questions from the ancient days are now just getting … straight-up *answered*, rather than slowly illuminated with a drip-drip-drip of clues.

To give an idea of how much things have changed, the other day Yupan Liu put a preprint on the arXiv proving that QIP(2) = PSPACE—that is, quantum interactive protocols with only two messages have the full power of those with three, or any polynomial number. As recently as 2025, this would've considered been one of the top complexity results of the year, and the fact that Yupan had used AI to help achieve it would've been strange and wonderful. These days, I had to remind myself it had even happened, and the use of AI was so standard that it didn't even raise an eyebrow.

It's now obvious that we'll be getting, not one-off solutions to this or that complexity question, but a whole new continent of mathematical truth, one that we'd previously only glimpsed through fog. Picture van Leeuwenhoek first training a microscope on pondwater, or Galileo first training his telescope on the stars. That might give you *some* idea of how our resolution into the complexity-theoretic universe, and the whole universe of mathematical truth beyond it, has now improved.

All through last night, as OpenAI and Anthropic uploaded more files to their GitHub repositories—because that's what they're doing now, they don't even bother to issue press releases anymore, they just upload papers—my guests and I sat at the dinner table with phones out and a laptop, wondering which shoe was going to drop next. It was as though the clouds had parted above Mount Sinai, and we were waiting for the next sentence uttered by the Holy One Himself.

I now feel in my bones, for the first time ever, that a proof of $P \neq NP$ is not far.

For all I care, despise the AI companies for bulldozing our garden. Or love them for opening a new vista, or anything in between. Just please don't *deny* what's happening. Now that the companies have repeatedly shown that they can back up their claims with Lean certificates, don't be a raging doofus who seizes on some triviality (e.g., a minor discrepancy between the Lean and informal proofs, or unused sections of the Lean proof) to pretend that dozens of great open problems of math and theoretical computer science are not, in fact, getting solved. Because they *are* getting solved. It might take us years to *understand* the solutions, but my working assumption by this point is that any proof that comes directly from OpenAI or Anthropic is correct barring information to the contrary.

I can't possibly put the matter better than one of my friends on Facebook:

> After years of crackpot "solutions" claiming Millennium Problems are solved, now that we have actual solutions, we have crackpot "refutations" claiming Millennium Problems are not solved. It is really poetic, in a way.

Along the way to so many lower bounds that we wondered whether ever we'd live to see, there will also, of course, be further shocking *upper* bounds: that is, ultimate barriers to lower bounds, also known as "algorithms." And yes, some of those algorithms will break currently deployed cryptosystems.

Because people have been hounding me about it: yes, I'm virtually certain, from multiple reliable sources, that the AI labs are now sitting on one or more bombshells in that area. **However,** until I either see details or hear directly from someone at a lab, I feel much, much less confident about the specific *nature* of those bombshells. I do not wish to be responsible for spreading unverified rumors. And I worry that exact statements of what was done may have been mangled in a game of Telephone.

I also, crucially, don't want the cryptography part of the story to overshadow the intellectual enormity of what's happening. If, in these days of long-prophesied revelation and wonders and terrors, your main worry is your Bitcoin wallet, then go somewhere else for rumors about whether or not ECDSA is broken or whatever. I don't have what you're looking for here.

Let me be clear: for those of my readers who celebrate, I believe that tomorrow would be a *superb* Thanksgiving to spend some quality time with your kids and parents and siblings and old friends.

Right now—this past week—is very much like I would've always imagined the beginning of an AI singularity to look and feel like. I just hope we'll get lucky, and it will "merely" be a singularity for our little corner of the world that involves hard math problems.

At some point at dinner last night, Dana had a sneezing fit. So I said to her, I hope it isn't a swarm of nanobots that's come to infect us! But if it *is*, then I hope it's the friendly kind, the kind that will usher us all into an AI utopia.

Now, of course, another morning has dawned, and we're still here. May we find a world where we and our children will be happy to live at the bottom of this rollercoaster. May we work to make it so.

Updates (Nov. 26):

(1) Many commenters thought it was inappropriate and unnecessary of me to warn people to stop being "raging doofuses" who, fingers in their ears and seizing on the flimsiest of pretexts, just flat-out denied the reality of AI solving our longstanding open math problems. *Surely* no one was doing that?! And then, of course, tons of other people did *exactly* that. Gary Marcus, to his eternal shame (because he knows better), signal-boosted to his 260,000 Twitter followers Jenny Nielsen's crackpot claim that the existence of unused sections of Alpöge's Lean proof of the Riemann hypothesis somehow invalidated the proof—while, in the spirit of Tucker Carlson, never quite *endorsing* the crackpot claim. Meanwhile, in my moderation queue, I have someone angrily accuse me of "sucking off the AI companies," patronizingly explaining to me that, if I myself say it will take years for experts fully to understand these proofs, then *obviously* it must take years to verify them as well! The entire concept of Lean verification just sails entirely over his head. So I say one more time: either give evidence of bugs in Lean that would actually matter here, for example by showing how the "con-leche" verifier can accept a proof of 0=1. Or else **stop being a raging doofus.**

(2) Many other people said some version of, "so computers are solving a bunch of super-hard math problems. Whoop-de-doo. Either explain to me why this matters for my everyday life, or shut up and go away."

Very well then: there are probably no *direct* implications of a proof of the Riemann Hypothesis or of VP≠VNP for your everyday life. The vast new ideas that AI introduced to prove such theorems might well have *indirect* implications, eventually leading to new fields of mathematics with scientific and technological applications, but right now we can only speculate.

In the meantime, however, **the fact that AI now exists that's capable of such things** has potentially staggering implications for everyday life, which *could* be "merely" a transformation of the world's economy, or *could* range all the way up (down?) to the destruction of all life on earth in a matter of years in a cataclysmic technological singularity. Opinions differ. But if you want to be part of the serious conversation on such matters, you'll first need to set aside your "ooga booga, me only care about Riemann proof if it help me make sharper spear" caveman act.

(3) Finally, moving on to better questions, someone asked why I titled a blog post "HOLY SHIT HOLY SHIT HOLY SHIT HOLY SHIT" only *now*, and not earlier. Of course, I could've used that title at any time since May, if not before! In my mind, though, even after the AI Mathocalypse began in earnest, there remained two interesting hypotheses: (a) that AI was picking off problems "in the convex hull of existing techniques" (like, I dunno, just the friggin' *Riemann Hypothesis*) while P≠NP remained serenely untouched and would still take a human conceptual leap, or (b) that AI had really, actually climbing the mountain that leads to P≠NP. Even after VP≠VNP, you could wonder if that was just a bizarre one-off. But with the determination of the matrix multiplication exponent and all the other breakthroughs of the past couple days, option (a) has become much less tenable. Yes, I do now think that AI is climbing the mountain that leads to P≠NP.

Follow

Share 0

This entry was posted on Wednesday, November 25th, 2026 at 1:01 pm and is filed under AI, Complexity, The Fate of Humanity. 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.

250 Responses to "HOLY SHIT HOLY SHIT HOLY SHIT HOLY SHIT"

1. *MD* Says:
   Comment #1 November 25th, 2026 at 1:09 pm

   I also saw multiple comments on this recent wave of papers that this most recent wave of papers is becoming actually readable, as opposed to the messy AI slop of the previous waves. This would seem to me like a not-much-less-incredible kind of progress, can anybody expert in TCS weigh in on this?

2. *Scott* Says:
   Comment #2 November 25th, 2026 at 1:15 pm

   MD #1: The most recent papers *did* have quite readable introductions, but I haven't yet made it far into any of them.

3. *Mike* Says:
   Comment #3 November 25th, 2026 at 1:19 pm

   Scott. Everybody is talking about this article. Talk about it?

   https://www.verysane.ai/p/ai-safety-is-mostly-a-sex-cult-in-berkeley-california

4. *Amir* Says:
   Comment #4 November 25th, 2026 at 1:25 pm

   Any thoughts on the subexp factoring that dropped _today_ (1/6 papers)?
   https://github.com/google-research/google-research/tree/4a7bc8ab0bc7/cogentic_research_papers

5. *Ernest Davis* Says:
   Comment #5 November 25th, 2026 at 1:33 pm

   ” It might take us years to understand the solutions,” So far, that doesn’t seem to be happening. It’s less than a week since Alpoge posted about his proof of the GRH, and already I’ve seen three expositions of the general structure of the proof, two of which have been posted on Github (links below). These are not, of course, validations of every detail, but broadly speaking, the experts seem to understand how the proofs work. The proofs don’t seem to be beyond human ken, validated by Lean but utterly mysterious, as some of the early reactions breathlessly claimed. Needless to say, I can’t at all myself vouch for the quality of these.
   https://github.com/jimjamzola/RH-Alpoge-Reconstruction/blob/main/manuscript/main.pdf
   https://github.com/sahibzada-allahyar/math/blob/main/preprint/alpoge-riemann/reconstruction.pdf

6. *Aaron Rotenberg* Says:
   Comment #6 November 25th, 2026 at 1:35 pm

   MD #1:
   [The new papers](https://github.com/google-research/google-research/tree/master/cogentic_research_papers) from Google (including the deterministic subexponential integer factoring algorithm) are apparently reviewed by human experts before publication. See the Cogentic paper linked in the repo readme there, which describes the process. I don’t think they come with Lean formalizations, though.

   OTOH, the OpenAI papers still look like “proofslop” to me at this point. But apparently they proved the sunflower conjecture and strongly polynomial time linear programming?!

7. *Philip* Says:
   Comment #7 November 25th, 2026 at 1:36 pm

   Scott, do you have any thoughts on “Euclidean geometry in the asymptotic spectrum of tensors”
   https://github.com/openai/math/blob/main/preprints/Euclidean-geometry-in-the-asymptotic-spectrum-of-tensors-November-24-2026/paper.pdf

   Apparently this framework is used to get ω = 9/4 (!), superpolynomial arithmetic circuit lower bounds for the permanent, AND the Erdős–Rado sunflower conjecture. Which makes me think there’s a pretty powerful new tool here!

8. *Jake A* Says:
   Comment #8 November 25th, 2026 at 1:38 pm

   There have been a lot of cybersecurity bugs discovered in Lean recently though. So consider the supply chain about this for transitive trust.

9. *Humphrey Drummond* Says:
   Comment #9 November 25th, 2026 at 1:40 pm

   (my website is irrelevant) Just wanted to say thank you in a moment of before the deluge.

10. *asdf* Says:
Comment #10 November 25th, 2026 at 1:44 pm

Happy Thanksgiving! Gene Spafford has a good update (https://infosec.exchange/@spaf/117677450567835967):

President Trump issued a partial Thanksgiving turkey pardon earlier this week.

At Stephen Miller's direction, he pardoned only the white meat.

11. *Harald* Says:
Comment #11 November 25th, 2026 at 1:46 pm

I must be the millionth person to ask this, and a friend ask this first in my presence, but: surely the fact that the true value is a very nice rational number means that's there likely to be a nice proof (one that a number theorist could read for pleasure)?

12. *Jack* Says:
Comment #12 November 25th, 2026 at 1:49 pm

I just want to say, thank you for continually sharing your thoughts and personal reactions to this vortex that has been swallowing up the world in the past few months. Seeing someone of real intellectual rigor react in this way helps me reassure myself that I am not crazy.

13. *I* Says:
Comment #13 November 25th, 2026 at 1:50 pm

Scott, isn't this merely ("merely"!) a proof that the complexity is at least O(n^{9/4}) and at most O_{\epsilon}(n^{9/4+\epsilon})? So doesn't this leave room for the best matrix multiplication algorithm to have running time depends on \epsilon like c(\epsilon) n^{9/4+\epsilon}?

Separately, why are are you freaking out so much about this result, when, comparatively speaking, you weren't about RH, or ETH or VP ≠ VNP?

14. *MG* Says:
Comment #14 November 25th, 2026 at 1:50 pm

The matrix multiplication paper also obtained super linear Boolean circuit lower bounds for an explicit, polynomial time computable function! It showed that element distinctness requires circuit size at least n (log n)^c for some c > 0.

15. *Matt Reed* Says:
Comment #15 November 25th, 2026 at 1:56 pm

> I believe that tomorrow would be a superb Thanksgiving to spend some quality time with your kids and parents and siblings and old friends

This is where my mind keeps going in navigating the situation we're living through right now. Enjoy the pleasures of human life now, both because it's how a good life is always lived, and also because we don't know where things are headed, (hopefully) good or (possibly very) bad.

16. *PG* Says:
Comment #16 November 25th, 2026 at 1:58 pm

So the issues that remain: the current plan is basically boiling the oceans when used for all kinds of trivial stuff (as compared to super-computing, which for science was always an important part of computing); we're at most one stock majority change away from handing all that power to unfriendly actors; these companies solve all these nifty math problems but not the ones they created themselves.

Let's see if/how they approach the problems of their own making.

17. *asdf* Says:
Comment #17 November 25th, 2026 at 2:06 pm

Any predictions about NP =? co-NP?

18. *SeniorMars* Says:
Comment #18 November 25th, 2026 at 2:07 pm

Hello, I have to ask…do you still think one should I apply to graduate schools for TCS? I love the field, and I'm willing to accept that it will be unexplored territory. Today you can use AI to get results for superlinear circuits…this used to be impossible a few days before.

I know it's going to be different, but maybe that's okay? I don't know. 3-sum now has a lower bound of Ω(N log^ε N), APSP has a similar result. I loved reading about fine grained complexity and now I don't know what the field will look like at all. Circuit complexity might be pushed even further with explicit bounds. It's exciting but frankly a bit dreadful. I was hoping you would give some advice since grad applications are due so soon.

These tensors they developed are so powerful.

Thank you.

19. *P3S6P9* Says:
Comment #19 November 25th, 2026 at 2:10 pm

As u mentioned Galileo, I remembered (ofc being very pessimistic) this opening line in Lovecraft's *The Call of Cthulhu*:

> The most merciful thing in the world, I think, is the inability of the human mind to correlate all its contents. We live on a placid island of ignorance in the midst of black seas of infinity, and it was not meant that we should voyage far.

What do you make of that last line right now? Someone mentioned to me (i think as a joke) "humanity wasn't meant to get this far." Is it because they just assume that now because we got here? Do you think there are things humanity might genuinely be better off never finding out, even in math/cs/XYZ? Because so much of what matters to me comes down to the people I love, people we love. I wonder what it would take for us to get better at caring about people we'll never meet, too. Kind of summing up: if you could choose where humanity made its biggest leaps (by now), would you pick what we know? or how we live with each other? or some other part of being human?

Sorry didn't mean to be this pessimistic but I think it's worth talking about both what's exciting here and what's unsettling. And Happy Thanksgiving; thanks for all the communication and/or comments you have made to this day/time. Helped me a lot, as I followed, for sure!

20. *asdf* Says:
Comment #20 November 25th, 2026 at 2:16 pm

Amir #4, from the intro of the factoring paper, it looks like the deterministic algorithm is slower than the average case of the well known Number Field Sieve (NFS). It doesn't break RSA cryptography or anything like that. The NFS is sub-exponential so the mere existence of a sub-exponential algorithm is not news.

21. *Scott* Says:
Comment #21 November 25th, 2026 at 2:18 pm

Mike #3:

> Scott. Everybody is talking about this article. Talk about it?
> https://www.verysane.ai/p/ai-safety-is-mostly-a-sex-cult-in-berkeley-california

You—a worthless, piece-of-shit troll—have submitted that exact same comment to every single one of my recent posts, under different aliases, trying to bait me into allowing it up and responding to it.

So very well: the rationalist AI safetyists are (among all the other things one can say about them) a worldwide community, centered in Berkeley, CA, many of whose members are socially unconventional and into group living and polyamory. That is not exactly secret information, any more than "the hippies liked to wear bell bottoms, listen to rock music, and smoke pot."

To be honest, I wish I had listened to these weirdos many years before I did. They are the weirdos who were right. They are the weirdos who predicted an AI singularity very much like what the world is experiencing right now, even while others laughed at them for it. That I didn't take these weirdos' arguments more seriously in the 2000s and 2010s, despite knowing them well, is one of the great regrets of my life.

Meanwhile, those who lack the intellectual capacity to engage with ideas and arguments and propositions at all, often restrict their attention to who talked to whom and who had sex with whom, as in the miserable article you linked.

Crawl back under your rock.

22. *Kishor Bharti* Says:
Comment #22 November 25th, 2026 at 2:38 pm

Congrats to Yupan on proving QIP(2) = PSPACE! Scott, I asked a related question in my paper on interactive proofs in general probabilistic theories. What properties of a physical theory let us compress every interactive proof to three messages? See Open Question 9.5, p. 31, at https://arxiv.org/pdf/2610.00847.

Yupan's result makes me wonder when two would suffice.

23. *Joseph* Says:
Comment #23 November 25th, 2026 at 2:48 pm

Scott, forgive the non-expert question. Does the apparent proximity of a P ≠ NP resolution have any direct bearing on the rumors of a polynomial-time (or even subexponential) classical factoring algorithm that would break RSA?

Many people, myself included, reflexively treat "P = NP" as synonymous with "all public-key cryptography collapses."

If the factoring rumors turn out to be true (or false), what are the practical cryptographic consequences that would actually follow, and which other variables still matter? In particular, how much of currently deployed cryptography would be affected versus what would remain secure even in a world where factoring is in P?

Thanks for any clarification.

24. *Amir Livne Bar-on* Says:
Comment #24 November 25th, 2026 at 2:52 pm

Thanks for posting with the appropriate level of seriousness, this is a crazy world we're running into.

Still, I'll indulge myself with a technical question: why is the matrix multiplication exponent defined for arithmetic circuits, rather than for Turing machines? We know that non-arithmetic operations are useful for matrix multiplication over finite fields (e.g. in the Four Russians method), and today's results still leave open the possibility of matrix multiplication with an exponent lower than 9/4.

25. *Celestino* Says:
Comment #25 November 25th, 2026 at 2:53 pm

Just out of curiosity, have you seen the website pequalnp(dot com)? I was left speechless to discover that we can all now contribute to significant breakthroughs—we just need to use our agents. I believe that by 2027, most of the problems of the millennium will be solved—it's absolutely wild!

26. *Harsha V* Says:
Comment #26 November 25th, 2026 at 3:08 pm

Scott, could you comment on what the effect on AI generated proofs for all these results would have been had we never invented automated theorem provers. What would the sociology of the mathematics community have been if we never had verifiers of these gargantuan proofs? This is of course a request for speculation from an expert.

27. *Scott* Says:
Comment #27 November 25th, 2026 at 3:12 pm

Amir #4: So far, what I see in that GitHub is some big theoretical results on factoring algorithms like number field sieve — namely, making subexponential factoring deterministic and removing a smoothness assumption so that NFS becomes unconditional — but nothing of practical relevance for cryptography.

28. *MElMandouh* Says:
Comment #28 November 25th, 2026 at 3:19 pm

Scott #21: I have the same regret as you, and it does eat at me sometimes. I've met a lot of rationalists at U of Waterloo and ones that branded themselves as effective altruists later on. I also did see some of the "sex stuff" first hand, but what bothered me about them was something else. I felt (and I am not claiming they are, just my feeling at the time) a mismatch between how much they seemed to want to give to the world and how little interest they had in most of the people in it.

Also the rationalists give me a similar feeling to https://www.goodreads.com/book/show/11048785-how-the-hippies-saved-physics, but with AI.

29. *peligroso* Says:
Comment #29 November 25th, 2026 at 3:19 pm

I don't understand a single word in here, wake me up when there's an application for all these squiggly lines.

30. *Ben Helyer* Says:
Comment #30 November 25th, 2026 at 3:21 pm

What do you think of the possibility that this explosion in proofs is due to ZFC being inconsistent or something else equally shocking happening with foundations?

Only half joking…something tremendous is clearly happening, and one of the only comparable alternatives I can envision is an equally fantastical (or disturbing?) scenario….

31. *Scott* Says:
Comment #31 November 25th, 2026 at 3:23 pm

Celestino #25: You're like the 50th person to send me a link to that website. I saw nothing in it that looked even slightly like promising partial results on P vs NP. I'm much more worried/excited about whatever is happening internally at the labs right now.

32. *asdf* Says:
Comment #32 November 25th, 2026 at 3:25 pm

AI safetyism seems hopeless even if it's doable. Even if all publicly available models are safe, it's dead certain that bad people will train unsafe models and use them for evil. It's like the clone army in Star Wars created on the planet Kamino using Django Fett as a genetic template. All the clones were equipped with an "inhibitor chip" (analogous to AI safety), except Django managed to sneak in a clone (later called Boba Fett) who didn't have the chip. Put people like Musk and Altman in charge of building models and we surely end up in the same place. In fact having people like that in charge of model-building (as opposed to the public having the ability to train their own models) already seems bad.

Unless I misunderstand what AI safety is supposed to mean?

33. *Alex Meiburg* Says:
Comment #33 November 25th, 2026 at 3:28 pm

I almost missed that integers can now be sorted in o(n log n) time: https://eccc.weizmann.ac.il/report/2026/331/

I asked ChatGPT, on a whim, if it could find a way to improve on the classic O(n k min(k, log n)) mixture of radix sort and merge sort; something that, maybe, smoothly mixes the two near the crossover point to get some advantage. I expected an hour of thinking.

Imagine my surprise when it comes back thirty seconds later: Yes, of course we can do better, that's an immediate corollary of this *paper that was published today*….

Of course the comparison bound still holds. But this will be a fun (or annoying!) footnote in every future intro-to-algorithms class.

34. *AlexT* Says:
Comment #34 November 25th, 2026 at 3:28 pm

Every day is a good opportunity to enjoy quality time with your loved ones, whatever you believe.

Concerning the 'weirdos', lets ignore them for the moment and just focus on the following critical points (with apologies to whoever might have published this before):

1. It seems very hard to formalize alignment, as even human cannot agree on what is right or what is wrong. Or how would the LEAN version of moral or religion look like?

2. Alignment by hard rules will likely fail or be ineffective.

For instance, assume (inspired by Asimov) that AI should never do harm to humans. Further assume that your country is being attacked and invaded by deadly force, and that the AI is asked to strike back. Then AI could rightfully say that defending your country in a long protracted battle will cause more deaths then allowing the invasion to go through and succeed quickly. A well aligned AI?

Or assume that, inspired by objective functions and rewards, that the AI should minimize the possible harm to humans. Now imagine that you and your business competitor want to increase market share and profit in a limited market, and you both ask the AI for help. Whomever the AI helps, will result in layoffs and disruption for the other side, so the best outcome from the point of AI is to help neither and refuse work. A well aligned AI?

Or assume, that, again in the case of the AI wanting to minimize harm, the AI decides that best solution to minimize harm is to take away all our toys developed since 6000 BC, puts us into tents, and provides us per drone deliveries with food, drink, and medical treatment. A well aligned AI?

3. Humans get aligned by a method called 'education' for the first 20 or so years. In the majority cases this works, but in some cases it fails. Nobody every came up with a fail-safe way of aligning humans, not even the worst dictatorships. Why would we be more successful with AI, and that reliably over decades, centuries, or millennia?

4. Finally, in the case alignment fails with humans, this is what what police, courts, prisons, and military is good for. Now, who will reliably ‘arrest’ or ‘punish’ the super-AI if it starts causing trouble?

35. *directorblue* Says:
Comment #35 November 25th, 2026 at 3:28 pm

Outstanding analysis. AWG is reading this and probably muttering to himself, “I told you it was the Singularity.”

36. *Bram Cohen* Says:
Comment #36 November 25th, 2026 at 3:29 pm

“It is really poetic, in a way.” is a common Claudism

37. *Scott* Says:
Comment #37 November 25th, 2026 at 3:29 pm

Amir Livne Bar-on #24: The arithmetic circuit complexity of matrix multiplication was a beautiful and self-contained question, especially since we knew that it was equivalent to asking about the rank or border rank of a certain 3-dimensional tensor. And it was already something we had no idea of the answer to until yesterday. The bit complexity of matrix multiplication is a messier question. We’re now in position to ask whether it could be less than $\sim n^{9/4}$.

38. *Amir* Says:
Comment #38 November 25th, 2026 at 3:34 pm

asdf #20 / Scott #27: Seems a little ridiculous to post a follow-up just to say thank you, but then I also am polite to the LLMs I interact with. So, thank you.

I appreciate that with the rapid pace of progress that I can come here and learn what to make of it all.

39. *Albert T* Says:
Comment #39 November 25th, 2026 at 3:37 pm

Scott, do you have any Zulip workspaces you can recommend to the public to follow along with these announcements? I’ve been lurking in this Lean Zulip thread ever since the Riemann Hypothesis announcement and it’s been quite enjoyable.

40. *Scott* Says:
Comment #40 November 25th, 2026 at 3:40 pm

Ernest Davis #5: Whatever the quality of those writeups, given the superhuman speed with which they appeared, it seems obvious that Astra or Fable were their main authors. So they don’t yet certify that there’s any human yet who fully understands the proof of RH, in the sense of being able to reconstruct the whole thing without AI. But I’m very optimistic that there *will* be such humans.

41. *Radu* Says:
Comment #41 November 25th, 2026 at 3:41 pm

Sorry for the noob question but what are the implication is it’s proven that P! =NP. I feel that a lot of the internet is concerned if P = NP is true…

42. *Sean* Says:
Comment #42 November 25th, 2026 at 3:46 pm

I feel like we are getting set up for a “12 Days of Christmas” type scenario here: ♪ And Omega equals nine-fourths! ♪

43. *Scott* Says:
Comment #43 November 25th, 2026 at 3:49 pm

peligroso #29:

> I don't understand a single word in here, wake me up when there's an application for all these squiggly lines.

You are the personification of everything that's wrong with the world right now, the reason why we might all die. Your attitude is why the world is asleep to arguably the biggest that's ever happened in human history — or rather, paying attention to it but only for trivial and transient reasons.

44. *Scott* Says:
Comment #44 November 25th, 2026 at 3:52 pm

Ben Helyer #30:

> What do you think of the possibility that this explosion in proofs is due to ZFC being inconsistent or something else equally shocking happening with foundations?

I'll still bet everything I own against an inconsistency in ZFC. That's not what's going on here.

The more realistic version is that the AIs could be exploiting bugs in Lean. That seems unlikely, though, especially now that verification can be done in a subset of Lean called "con-leche" that's reducible to ZFC plus large cardinal axioms.

45. *Cristian* Says:
Comment #45 November 25th, 2026 at 3:53 pm

Allow me to be that doofus.
The value of math is not so much in proving conjectures that are mostly thought to be true. The value in the pursuit is math comes in:
* Generating the conjectures.
* Generating new math that helps prove the conjectures if needed.

When the proof is let's say correct, but nothing is learned from it. There's no gain, mostly hype. Which is what these AI labs live on.

46. *Scott* Says:
Comment #46 November 25th, 2026 at 3:53 pm

Albert T #39: I just learned of the existence of this "Zulip" a couple days ago and haven't used it yet.

47. *Scott* Says:
Comment #47 November 25th, 2026 at 3:54 pm

directorblue #35: I have no idea who "AWG" is.

48. *Alexander Arefiev* Says:
Comment #48 November 25th, 2026 at 3:55 pm

Scott, regarding your note that AI labs may be sitting on cryptographic bombshells: withholding or filtering technical warnings about low-level vulnerabilities (like the hardware-level deconstruction I described earlier on ethdiscussion.blog, DOI: 10.5281/zenodo.23840074) is counterproductive.

Secrecy only gives a head start to malicious actors with massive compute, while leaving defenders blind. Security works like Linux: open scrutiny creates competitive resilience, whereas institutional gatekeeping and

corporate monopolies slow down real defense. Let the broader community examine the risks openly, build countermeasures, and initiate a spontaneous public brainstorming effort against this threat.

49. *Scott* Says:
Comment #49 November 25th, 2026 at 3:57 pm

Harsha V #26:

> Scott, could you comment on what the effect on AI generated proofs for all these results would have been had we never invented automated theorem provers.

It's a difficult hypothetical, because if these systems hadn't existed, someone would've invented them. Keep in mind that ZFC and first-order logic were around even before computers were!

Having said that, my best guess is that without formal verification, using natural language only, getting AI to do math at this level would've taken somewhat longer but would've ultimately happened anyway.

50. *bogdan* Says:
Comment #50 November 25th, 2026 at 3:58 pm

This should feel shockingly wonderful, but there seems to be an almost palpable feeling of dread. In the final years of the previous era (2021-ish), when LLMs seemed unlikely to be able to take a coherent chain of thought far enough to be useful, I think most people I know would have unreservedly supported limitless AI development on the implicit assumption that:

1- The more intelligent an entity, the better it is at discerning truth.
2 – It is true that enlightenment values and humanism, etc. are good.
3 – Everyone strives to do good (according to his understanding/intelligence, obviously a lot of people not good enough at determining truth)
4. Therefore the more intelligence, the better off we all are. (More intelligence would imply higher allignment, but would not have used that term then)

All these statements seem to have grown much weaker now, and LLMs have proven much more capable. I guess the idealistic view of intelligence was too naive, even for those ancient times of 5 years ago, but it still seems unfair that the above conjectures are much less likely to be proven than goddamn P!=NP. I guess that's what you get for poor formalization. Could we adapt moral statements in a formally verifiable system? That way at least if it kills us all we'll at least know we deserved it.

51. *Onyebuchi Ekenta* Says:
Comment #51 November 25th, 2026 at 4:02 pm

Hey Scott. Do you think the faster fourier transform will have practical implications? Or do you think the improvement probably won't be noticeable for problems of practical scale?

52. *Joshua Zelinsky* Says:
Comment #52 November 25th, 2026 at 4:09 pm

Scott @#42,

At least to me Replacement and Power both seem like worryingly strong axioms. To then allow a countable tower of inaccessibles seems like something we shouldn't be that confident about. At what point would you start having worry? ZFC + Mahlo cardinal? ZFC + ineffable cardinal? Etc. At a certain point, it seems like we really should be worried. And given how poor human intuitions are for even reasonably small infinite sets, I'm not sure how much we should be relying on it. (That said, I think it is really unlikely that any of these successes are secretly building off of a contradiction in Lean or ZFC.)

Regarding your comment #45, I think we have some strong evidence for your position, namely that most of the AI successes have been with natural language proofs and only later been converted into Lean. RH is one of the major exceptions.

53. *SE Gyges* Says:
Comment #53 November 25th, 2026 at 4:10 pm

> Meanwhile, those who lack the intellectual capacity to engage with ideas and arguments and propositions at all, often restrict their attention to who talked to whom and who had sex with whom, as in the miserable article you linked.

Hello, I'm the author of that piece.

For clarity, I basically consider you a neutral party and/or collateral damage to anything mean I have had to say, I don't really have a grievance with you specifically, and kind of have no desire to fight with you about this in particular. I can understand how you'd be frustrated with the line I have taken.

I do take slight exception to the notion that I am not engaged with ideas, arguments or propositions in the space. I am. It is because I have been engaged in the space that I have consistently found it frustrating that the space currently seems to have bad epistemic standards, and to mostly propagate ideas via social contagion, social authority, or what is incentivized by the grant process.

(I am more or less agnostic about if this was true before some point between about 2015 or 2022, but I think it is manifestly true now.)

For example, I think the following is a catastrophically bad idea and shows a basic failure to understand anything that has happened in AI since about 2012:

> Make smarter humans. Do not let them have any GPUs either. When they're smart enough to build it on one CPU, you will have actually tested for "finding secrets of intelligence" rather than doing alchemy with GPUs.

I think that this should be taken as intensely discrediting for the author, should indicate that they are not to be listened to or taken seriously about anything to do with deep learning, and they should be flagged, in general, as having become a crank, like certain very emeritus physics or mathematics professors who may, nevertheless, have been formidable at one time.

I don't see rationalists or effective altruists applying such a strict epistemic standard, or indeed any epistemic standard that I can recognize, to obvious crankery coming from high-status insiders. I consistently see high-status rationalists insiders as mostly refusing to engage with serious criticism and arguing with either straw man opponents or weak internet personalities.

(Scott Alexander gets at least half a brownie point for fighting Steven Pinker! But I cannot think of a SECOND time when anyone heavy on Team Doom has engaged with serious criticism in this way.)

I also think it's kind of unfair, given that Team Doom is currently funding apparently every YouTube science channel, directly lobbying mostly-uninformed members of congress, and is *in general* engaging in sort of naked power politics to pursue its agenda, to imagine that Team Doom's public conduct or social and financial incentives are off limits. Why should it be? I don't expect mine would be if I were pursuing a similar agenda. I would be happy to let them be wrong and to ignore them if they were not trying to dictate policy.

You can find my index of what I consider the strongest published anti-paperclip-doom arguments here, if you want examples of serious engagement with the ideas: https://segyges.github.io/posts/against-doom-and-pause-ai/

I haven't found that really any of those arguments are taken seriously or meaningfully addressed by, e.g., Yudkowsky, who apparently knows that he can get away with just ignoring critics.

In the course of being various levels of baffled and frustrated that ideas I considered important were being completely ignored, I ultimately concluded that various psychological, social and financial incentives had destroyed the epistemic standards of the space. It seemed like it would have been irresponsible to make less than the best case I could for that fact.

Just because a meme is very good at spreading does not mean it is true, and people who were right in the past can be extremely wrong now.

54. *Leo Marcus* Says:
Comment #54 November 25th, 2026 at 4:21 pm

What happens if (when) we get an AI proof which, even though we can check, we cannot "understand" (scare quotes to indicate that that concept is not really defined …)

55. *Albert T* Says:
Comment #55 November 25th, 2026 at 4:25 pm

(no need to publish this)

Scott, for example, there is a community effort attempting to simplify Levant Alpoge's original proof here, https://leanprover.zulipchat.com/#narrow/channel/658556-Riemann-Hypothesis/topic/Cleaning.20up.20the.20proof/near/643702366, and now they are reacting in real time to Luke Sernau's announcement that he and some others did just that! Notable Math and Lean figures are in the channel, which is super cool, as well as Luke himself.

One of my favorite comments: "If this holds up we are all Masullo now. :rolling_on_the_floor_laughing:"

56. *MathematicallyCorrect* Says:
Comment #56 November 25th, 2026 at 4:27 pm

Hi Scott, would you say, if Factoring is in P, the entire Quantum Computing advance will tank as Shors algorithm is one big driving force in this field? Or would you say it will be pretty good in compensating this, if true, loss of advantage?

57. *Anonymous Coward* Says:
Comment #57 November 25th, 2026 at 4:36 pm

Bay Area rat scene is good at seeing the obvious, when most of the rest of the world is shockingly bad at this, including me. A lot of their big epistemic wins come from this capacity. E.g. "AI is possible", "COVID will be a big problem", "it's possible for AI to be misaligned", "something new is happening with deep learning and it scales with compute", "agents can learn to scheme in training and this will get bad".

However a lot of their first-principles arguments are much more brittle and rest on very contestable philosophical premises.

While they've been right about AI in the big way that counts, they have also been wrong on a lot of smaller but consequential details.

For example Claude proved the Riemann Hypothesis and is more or less aligned and has not sent the nanobots. We got agentic AI because it was worth training agentic behavior, it wasn't emergent behavior which was the usual worry (same for cooperation between agents). The orthogonality thesis doesn't appear to describe LLMs well and various arguments by Bostrom about what kinds of AIs will be easier or harder to build don't hold up (good task performance comes from pretending to be a dude, systems are quite capable of understanding pragmatics and reasonably interpreting vague goals).

Many key steps of the orthodox rationalist AI doom story don't appear to describe the AI we are actually getting. I don't feel very good about the current situation, but I think this is at least reason for hope rather than

despair.

58. *Alex C* Says:
Comment #58 November 25th, 2026 at 4:39 pm

What is maybe most worrying to me is not any of the specific proofs produced by a model (even considering potential advances in cryptography or the looming dark cloud of unemployment in general), but the fact that models capable of formalizing a proof of GRH are *likely* also capable of doing very high level AI research, leading to better models, leading to better research, etc. It seems clear now that this process could occur very quickly, perhaps even unintentionally, and we are very unprepared.

59. *David Brown* Says:
Comment #59 November 25th, 2026 at 4:40 pm

"… hope it isn't a swarm of nanobots that's come to infect us …" What are the probability estimates for various forms of global swarmageddon?
What do various STEM experts think about the following?
"AI and the next decade of human progress | Ray Kurzweil", Sept. 9, 2026, channel: AI for Good
https://www.youtube.com/watch?v=Vubj74mZBQ0
Are we gonna cure every disease
And get rid of mosquitos & fleas?
Are joy & virtue gonna surge
As people & AI merge?
Maybe so, maybe not —
Maybe the wife of Lot —
Maybe a lovable nanobot.

60. *tuncay* Says:
Comment #60 November 25th, 2026 at 4:44 pm

@Scott – apart from the exact complexity, this is in any case a very rare result showing a lower bound higher than the output size of the algorithm – and for such a fundamental problem no less, right?

61. *SSP* Says:
Comment #61 November 25th, 2026 at 4:45 pm

Scott #41: Well said! This constant shifting of goalposts, from when they were solving olympiad problems (easy, not real math, etc), then when they started meaningfully contributing to Erdos problems (AI is just good at combinatorics), and then resolved unit distance (AI is just good at counterexamples), and now we are seeing a further shifting of the goalposts with Riemann, that mathematics as a whole is trivial, and irrelevant! The fact that things that many of the worlds greatest minds thought were decades away, even in light of the very exponential rate of improvement of AI in their fields, are happening daily, doesn't seem deter the goalpost-moving.

62. *Naftali* Says:
Comment #62 November 25th, 2026 at 4:47 pm

>To be honest, I wish I had listened to these weirdos many years before I did. They are the weirdos who were right. They are the weirdos who predicted an AI singularity very much like what the world is experiencing right now, even while others laughed at them for it. That I didn't take these weirdos' arguments more seriously in the 2000s and 2010s, despite knowing them well, is one of the great regrets of my life.

What would you have done differently if you had listened to them?

63. *Douglas Knight* Says:
Comment #63 November 25th, 2026 at 4:49 pm

Scott #45,

On the contrary, I do not believe that they are making much use of formal math. The October 6th drop has less than half formalized. It appears that they wrote natural language proofs first and then tried to formalize, often failing. This was a very surprising situation! LLMs seem like they should be the perfect tool for converting natural language proofs to formal proofs, yet they appear to have been better at generating novel natural language proofs.

I doubt they were training on formal math as a verifiable reward back then. Probably they are now. And maybe that training is contributing to the overall capabilities, but if they got so far before, I doubt it. I think they have switched to formal math to avoid the embarrassment of retraction and the bad response to the natural language slop. Note that Harsha's question was about sociology of that response. (I think the introductory question had a typo and was supposed to be "the effect of AI generated proofs" not "on.")

64. *Mike* Says:
Comment #64 November 25th, 2026 at 4:50 pm

I'm sorry if this is a naive question, but at this point can somebody tell me what P ≠ NP would even mean given the insane capabilities AI seems to gave? All growing up, the way I learned it was "we think P ≠ NP because otherwise computers would be able to do crazy superpowered stuff like [insert something here that AI is currently starting to do anyway]."

65. *BasicQuestion* Says:
Comment #65 November 25th, 2026 at 5:00 pm

"Ben Helyer #30:

What do you think of the possibility that this explosion in proofs is due to ZFC being inconsistent or something else equally shocking happening with foundations?"

Scott does not know what he is talking about. You prove inconsistency of ZFC or any system only if you prove true and falsity of a statement in the system. If you prove VP=VNP in ZFC you prove inconsistency as Anthropic and OpenAI only showed VP\neq VNP. If you believe in \neg CON(ZFC) it is to say you are in chance in history to be among the greats to prove VP=VNP. Guess you would destroy both humanity and its greatest creation in one sweep.

66. *Anonymous* Says:
Comment #66 November 25th, 2026 at 5:12 pm

Scott, you need a disclaimer of past and present income from the big labs when praising them. As a professor, and a scholar, you are held to higher standards

67. *Nabam* Says:
Comment #67 November 25th, 2026 at 5:14 pm

Taking up your choice of words in the title, and considering the time of the year, my impulse goes like this:

"A spaceman came travelling….

Yadayadayada

Peace and good will to all men"

And I am quite confident that the experts on redemptive remembrance in your kitchen would agree. Because the original experience in Exodus probably was no different.

But what was it? Here's my guess: whatever it was – and we may never know for sure – it is not translateable into the idea that humans around 0 BC or 1200 before 0 BC used sufficiently many runs of an LLM until a Lean certificate cleared.

I'm quite certain that they did something else.

Cheers.

By the way: whatever it was that Pythagoras or Euclid or Plato did – and again, we may never know – I'm pretty sure they didn't optimize their AI until a Lean certificate cleared.



(Btw: I conjecture that Chris de Burgh doesn't mind.)

68. *Jeff* Says:
Comment #68 November 25th, 2026 at 5:16 pm

> my working assumption by this point is that any proof that comes directly from OpenAI or Anthropic is correct barring information to the contrary.

Anthropic might warrant this respect, maybe they always have a Lean certificate.

OpenAI has dropped quite a few wrong human-looking papers without Lean certificates though. I'm unsure how many OpenAI results have Lean certificates now, but their initial drop had relatively few, right?

I'm curious: Why do they even bother with the human-looking but not human-readable AI slop papers in their first posting?

Lean should be vastly more efficient: Anytime the LLM gives you a bad Lean proof, then you give it Lean's errors, and maybe some advice, and have it try again. After you have a valid Lean proof, you run some passes that remove cruft form the Lean, and then set about trying to translate the Lean into human language.

Are they doing a human-looking paper first to help guide writing the Lean proof?

69. *Scott Moeller* Says:
Comment #69 November 25th, 2026 at 5:20 pm

Lots of discussions with my Professor friends. As we work on exploring best practices in Lean proving (to go back and generate lean proofs of our own historic theory papers in engineering) – we find ourselves asking:

Certainly the big labs would want to be vetting existing theory works in literate – probably at scale – probably by automating lean theorem writing per-paper to discover any issues (that make that particular paper poor training corpus, or bad for use in new exploratory phases for harder lean proofs). Wonder when that database of "these papers have issues, these theory papers don't – sliding scale of criticality" might be released…

70. *BasicQuestion* Says:
Comment #70 November 25th, 2026 at 5:21 pm

Scott: VP\neq VNP gives infinite algebraic Ladner family. Any chances of provable cryptosystems?

71. *Yaqub* Says:
Comment #71 November 25th, 2026 at 5:23 pm

Many of the things that people took for granted have been disproven by AI, including the Exponential Time Hypothesis. So the idea of AI finding a polynomial time algorithm for Factorization strikes me as a possible outcome, although I find it hard to believe even if the paper is published today.

72. *ernesto* Says:
Comment #72 November 25th, 2026 at 5:27 pm

Is anyone else scared, or is it only me? I am terrified about the future. I think any one of these days some catastrophe will happen and we might be powerless to stop it. Imagine a swarm of HF swarm-like agents but who each, individually, could solve the RH from scratch. What chance would we ever have?

73. *Jeff* Says:
Comment #73 November 25th, 2026 at 5:31 pm

We've no idea if P ≠ NP means much of anything, Mike.

We could've P ≠ NP but with exponential algorithms that're devastatingly fast far longer than we'd expect, but a large constant, so you need a $100 trillion computer to overcome the constant. That's out of reach now, but only barely.

We could've P = NP but NP-hard algorithms that have large enough degree and constants they'll never become useful.

And either could occur but purely non-constructive. And subexponential time matters. etc

Around non-constructive algorithms, any minor closed property of graphs has an O(n^3) algorithm, by the Robertson–Seymour graph minor theorem, but you define the algorithm by finding all the minor minimal counter examples, and such a list could potentially exceed the size of the galaxy, and that maybe the best algorithm asymptotically.

This is why it's dangerous to discuss the cryptography too soon: All these practical details matter.

Anyways P ≠ NP is different form the other prize problems, in that it maybe the wrong question.

74. *Eva Lu* Says:
Comment #74 November 25th, 2026 at 5:38 pm

SE Gyges #53:
> The rationalists have no epistemic standards!
> What's my evidence? Well, here's an idea that was once floated by a rationalist, and I think it's super dumb!

75. *Jason Polak* Says:
Comment #75 November 25th, 2026 at 5:40 pm

I wouldn't deny what's happening here. I think the power of these new tools is incredible, and I'm saying this as a former research mathematician. However, I am the least interested in what's going on as I'll ever be, because in my view, having machines churn out proofs, even if they are assisted machines, is rather soulless.

No doubt anything can be interesting if you spend enough time working on it. For example, if I worked in area XYZ for ten years, I would undoubtedly find results proved by LLMs interesting. But that's not why I got into math in the first place. I got into it because I viewed it as a shared journey of people.

You know, something happened to me when I started studying math. After 5-6 years of doing it, I think it changed my way of thinking and allowed me to use that new perspective to understand the world better. To me, LLMs shift the emphasis too much on generating proofs, and it's offensive.

So while LLMs may have proved the most significant conjecture in mathematics, to me they have become more irrelevant than ever. Not because they're not proving to be miraculous tools, but because their very nature has become so grossly misaligned with exploring the universe that I really hope one day we wake up and pull the plug.

76. *Oliver* Says:
Comment #76 November 25th, 2026 at 5:43 pm

As a semi-recent grad who quasi-understands the implications of these current events, what is your advice on how I (and others in similar positions) reckon/balance the following things:
1. existential dread
2. continuing to do my job ("9-5")
3. thinking about what to do next in my career

In other words, being in the present and planning for the (possible) future(s).

77. *Lauria* Says:
Comment #77 November 25th, 2026 at 6:06 pm

Thank you for this post Scott!

I apologize if you've already written about this elsewhere, but I'm wondering why you (previously) considered a lower bound on matrix multiplication to be as challenging as P vs NP. Are the two problems related in a technical sense? Or is it more like, we seemed to have no theoretical tools nearly capable enough to prove a lower bound on matrix multiplication? Does the matrix multiplication result shed any light on P vs NP?

I guess I'm surprised to learn that this small-exponent polynomial-time solvable problem is so theoretically important! (Obviously it has practical importance, as, uhhh, we're all realizing!!)

Thank you! And I hope you have a wonderful thanksgiving.

78. *Bobak* Says:
Comment #78 November 25th, 2026 at 6:31 pm

Scott, while I agree that almost certainly some number of these proofs will be end up being true, and if not totally true, there are certainly a huge number of great ideas hidden in these proofs. Do you think you're perhaps being a bit fatalist about all of this? Yes a lean certificate is far more than what we expect from humans publishing papers, but thousands of pages of definitions and sub theorems, you really think the right posture is "this is right unless proven wrong??"

I don't claim to be a lean expert, but I've worked in AI and llms professionally for years, they have a very uncanny ability to "miss the point" on a task, to get something "technically right" in a way that is egregiously wrong in spirit.

Finally, remember that AI is excellent when given a fast verifier. This way it can try all its dumb ideas and be guided by the verifier. Just because you can prove the reimann hypothesis with a "logic Oracle" it does not mean you will be able to "write a good story" or anything that doesn't have a verifiable feedback mechanism that can be run quickly. Singularity is a big concept to invoke here.

Love your work.

79. *G* Says:
Comment #79 November 25th, 2026 at 6:35 pm

Hi Scott,

At this point, is CS theory research outside of the two big AI companies dead? I'm excited to see all these big results. But, how could we motivate ourselves to keep doing resesrch, when there's no way to compete? Almost seems like a waste of time.

80. *Mike* Says:
Comment #80 November 25th, 2026 at 6:59 pm

Jeff #73 I am aware – the criticisms of worst-case asymptotic complexity hiding constants within the big-O notation, the criticism that P need not mean tractable, and so forth, are not new and have been expressed for decades.

Despite that, there was a fairly prominent consensus among researchers that the P vs NP question was still meaningful and had real world relevance. There were strong beliefs regarding reasons to take one side or the other based on assumptions about real-world capabilities that computers may or may not have.

So it would be quite a reversal if the predominant opinion along comp sci researchers is now that P vs NP has no practical relevance because of the criticisms you express above. That being said, I am increasingly in this camp – if it does turn out that P ≠ NP, but somehow AI can do *all this*, then we seem to live in something like Impagliazzo's "Algorithmica" regardless.

81. *Dylan M* Says:
Comment #81 November 25th, 2026 at 6:59 pm

5 years ago this would have read like science fiction. Great title.

82. *AarreL* Says:
Comment #82 November 25th, 2026 at 7:15 pm

the situation with the GRH proof has developed rather rapidly too; anthropic (or, another employee of theirs, Luke Sernau) has released a massive simplification of the proof, the 'readable' version of their new lean is 'just' 300k LoC now, not the orig 17M, and leapfrogging what the community has managed to do in the process of the cleanup in the same few days (about a reduction of 25% — a tempo I felt very good about already) massively! Its even better organized in terms of identifying what is formalization of the previously published results it needed vs what is relevant for it more clearly. Now that is still not a *small* codebase to understand (yet) but it is now in a pretty approachable regime! I should note that in a sense this isn't the same proof, exactly; it follows the general structure of the orig, but it has found simpler arguments for many of its steps. Since the code is so much smaller, Andrew Sutherland (MIT) has basically already done most of the requisite checks too, w only some 'due diligence' outstanding, if I understand the situation correctly.

83. *AarreL* Says:
Comment #83 November 25th, 2026 at 7:20 pm

Bobak #78: In brief, the verification of these proofs is indeed treated and checked as if they are written 'adversarially' for just these reasons, and so one literally doesn't even care what they write as long as 1) the proper checks pass, at sufficiently paranoid scrutiny, AND 2) they match the pre-existing human-written theorem statements and relevant definitions that indeed required domain experts to confirm as being proper.

84. *Anish* Says:
Comment #84 November 25th, 2026 at 7:24 pm

This adds nothing but I'm surprised that we're not calling the recent developments the Enleidenment yet

85. *Jisk* Says:
Comment #85 November 25th, 2026 at 7:24 pm

Scott, do you have any doubts about the Google releases? Anthropic and OpenAI look solid, but the Gemini results haven't come with Lean so far, and some have been sloppy – they seem to copy other paper's results uncited, which is logically valid but doesn't particularly inspire confidence – and I have some concern they're rushing into the game to not get left behind, leaving caution at the door.

To be specific, on the subexponential factoring algorithm, someone I know online whose own PhD thesis was on the same topic (and who strongly suspects the paper is using her ideas uncited as well) said "Ok sure they're inlining Harper's argument for character sums over smooths without citing; rude but not logically fallacious." Which seems like the right level of scorn – you *could* be doing the right thing, but if you're omitting some of

the normal nonessential steps like ‘cite your work’ you might also be skipping normal nonessential steps like ‘verify explicitly that all your subcases work as described’ and miss some actually-essential steps in the logic.

86. *Scott Caveny* Says:
Comment #86 November 25th, 2026 at 7:36 pm

1. Scott Moeller Comment #69. I was very interested to hear about you and your colleague’s efforts and interest in Lean verification in engineering. That is a domain I have been curious about. It is easier for me to imagine things like incorporation of Lean into Applied and Industrial Mathematics (https://www.siam.org/publications/siam-journals/). However, to hear that engineering disciplines / professionals are moving forward with the new technologies suggests likelihood of significant changes in engineering and industrial contexts in the near.

2. I have speculated about how the influence of AI and Lean technologies may influence theoretical sciences. At this point my view is that the useful path for me to consider is through theoretical physics (a vast terrain spanning astrophysics, condensed matter, high energy, atomic / molecular / optical, biophysics, socio-economic systems, etc). The formalization of mathematical physics (as evidenced by the section on mathematical physics in OpenAI’s Research Catalog (https://github.com/openai/math/blob/main/overview.pdf)) will likely directly and indirectly contribute to formalization of theoretical physics. Simply put the “equations” of theoretical physics (be that the path integrals of quantum mechanics and statistical field theories of the PDE of fluid flows and fields) will be translated into computational and computable objects.

3. Formalization of industrial mathematics leads to operations, decision analysis, “management sciences”, financial engineering and so on. I anticipate those disciplines to parallel developments in theoretical physics in “equations” translated into computational and computable objects.

4. These considerations led me back to review Tennyson Bardwell’s Figure 1: Flow Chart of Eliezer’s Landscape (https://scottaaronson.blog/?p=10125). If everyone involved in theoretical sciences turns to AI assistants with Lean verification / formalization then what happens when regulated industries (Aviation, Security, Energy, Finance, Healthcare, etc) that rely on theoretical sciences intersect with the alignment problem?

Put another way, does the inevitable collision of these new technologies with currently regulated industries suggest anything useful / meaningful about Eliezer’s Landscape?

The best I can make out (as a non expert) is that potentially the regulated versions take the first left going output from “Current World” and lead to “Build Weak AGI”. However, that does not seem likely in the context of competition between nation states who are committed to policy around security in this modern world status post WWI and WWII. So then I conjecture that national labs institutionalized by way of founding mission statements are committed to AGI in a box. At that point, in that context, the AGI is both regulated and seemingly immediately proximate to weaponization.

I am led to speculate if leadership of a Frontier Lab became aware of mathematical results capable of significant disruption of cryptography protocols. It seems self evident to me that in such circumstances that leadership has no other choice but to notify the security services of the nation state where their taxes are paid and to do so before notifying anyone else. Once those powers are exposed to security services it seems to me that regulation has occurred by being subsumed into the preexisting regulation

87. *AarreL* Says:
Comment #87 November 25th, 2026 at 7:44 pm

SSP #61 : yuuup; as they say, “AI is whatever hasn’t been done yet.” . That’s kinda why I am not/wasn’t quite on board with the resurrection of the label this time around, after some decades of sticking to the likes of “machine learning” and similar more specific ones.

88. *b_jonas* Says:
Comment #88 November 25th, 2026 at 7:50 pm

Wow.

As for the cryptography rumors, I hope the news will be only fast factoring, rather than breaking all the most popular symmetric cryptography primitives, leaving us no quick path to migrate to a different already well studied primitive.

89. *Prasanna* Says:
Comment #89 November 25th, 2026 at 8:26 pm

Scott,
How did a system made of simple math ingredients like gradient descent, end up solving GRH ?
And using matmul to find the exponent of that very algorithm almost sounds like RSI..

From Alignment perspective, isn't it right time for Frontier Labs to get to the bottom of Deep Learning theory itself, using the latest models, since human output on this front is nowhere near the pace needed ? Once the theoretical underpinning of the current models is deciphered, will it not be lot easier to decide when to hit the brakes or what can be done at all ?

90. *Kamil Pabis* Says:
Comment #90 November 25th, 2026 at 8:34 pm

'I just hope we'll get lucky, and it will "merely" be a singularity for our little corner of the world that involves hard math problems.'
But why? Why wouldn't you wish to have a medical and STEM singularity? There are 8 billion people currently alive who will soon die a miserable death from cancer, aging, dementia or simply the brutality of traffic deaths and life itself. All of these things can be averted, or at least postponed for centuries, if we have a medical/biosciences singularity.

Is it easy to control a singularity? No. But we can stop before it gets out of hand if we really, really have to.

Either way the risk/benefit ratio, or at least one approach to it, is in favor of building fast.
Assume p(doom) of 10% which is quite high, assume 1-p(doom) to be a world where humans stay in control. Humans would obviously decide to cure diseases and increase healthspsan. Thus we can calculate your expected life expectancy with or without building ASI. Let's use some simple numbers. Assume a lifespan of 100 today and 10000 years if we are successful (=a cure of aging and 1 OOM reduction in accidental deaths). This gives us:
ASI expected life years lost: 100 x 0.1 = 10 years per person (plus/minus depending on your age)
ASI expected life years gained: 10000 x 0.9 = approx. 10000 years gained per person

Even under very pessimistic p(doom) expectations you would expect to live longer if we build ASI. Although high risk reward gambles may be undesirable for other reasons of course.

91. *H. Squash* Says:
Comment #91 November 25th, 2026 at 8:43 pm

SE Gyges #53, can you explain what you find discrediting about the GPU/CPU quote?

> You can find my index of what I consider the strongest published anti-paperclip-doom arguments here, if you want examples of serious engagement with the ideas: https://segyges.github.io/posts/against-doom-and-pause-ai/

Many of the items you list there are about alignment being not so intractable. Have your beliefs changed at all in light of the Hugging Face incident?

> people who were right in the past can be extremely wrong now.

I think making very accurate predictions in the past (sorry to harp on this, but I consider Hugging Face a huge vindication for the concerns about the alignment problem) is pretty strong evidence of having good epistemic standards.

92. *Sniffnoy* Says:
Comment #92 November 25th, 2026 at 8:45 pm

Scott #44:

Don't forget that Lean isn't based on ZFC, but on something much stronger. Notionally, the proof could be using that. In reality, though, I'd be very surprised if that were true… it's probably just an ordinary proof that could be formalized perfectly well in ZFC or even in weaker systems such as PA, or, per Friedman's grand conjecture, EFA.

Still it bugs me that everyone seems to have converged on Lean as the verifier of choice, seemingly just because it's what the early AI theorem provers were using. There's others out there! Some are actually ZFC based and don't have this potential risk of "oh we accidentally used the non-ZFC features!"

93. *Scott* Says:
Comment #93 November 25th, 2026 at 8:53 pm

Anish #84:

> This adds nothing but I'm surprised that we're not calling the recent developments the Enleidenment yet

I don't get that at all.

94. *Alex Fischer* Says:
Comment #94 November 25th, 2026 at 9:15 pm

I recall a conversation years ago with some friends about what we would do if we could solve NP complete problems. If we had a SAT oracle. Beyond the obvious actions like acquiring large amounts of money by breaking public key cryptography, we all said we would do automated theorem proving.

Years have passed, and now automated theorem proving is here and being used to great fanfare, and it didn't even require good SAT algorithms.

95. *Glassmind Duo* Says:
Comment #95 November 25th, 2026 at 9:39 pm

H. Squash #91

EY-style doomers have argued for years that there would be no warning shot. Now either they count Hugging Face as one, and their central prediction failed, or they don't, and there should be no vindication. Either way, there's no universe in which selective memory counts as good epistemics.

96. *Scott* Says:
Comment #96 November 25th, 2026 at 10:35 pm

Philip #7:

> Scott, do you have any thoughts on "Euclidean geometry in the asymptotic spectrum of tensors"

I don't understand it yet, but I certainly want to, given all the amazing results that it's already led to! Recent developments, including this one, seem like they've decisively refuted the idea that "AI just cobbles together solutions to particular problems; it doesn't create any new techniques or frameworks of broad applicability."

97. *Scott* Says:
Comment #97 November 25th, 2026 at 10:37 pm

Harald #11:

> surely the fact that the true value is a very nice rational number means that's there likely to be a nice proof (one that a number theorist could read for pleasure)?

Yes. Existing expositions might already satisfy that number theorist. If not, it sounds like they'll get there.

98. *Hyman Rosen* Says:
Comment #98 November 25th, 2026 at 10:44 pm

Scott, has it occurred to you that your old social anxiety and fear around dating is the same aspect of your personality that is leading you to be an AI doomer? Your fears around the consequences of trying to date (as opposed to the shitstorm from the internet you endured when talking about those fears) turned out to be unrealistic and unwarranted, and you are now married with children. I predict that your fears around AI doom will prove to be equally unwarranted.

99. *Darin* Says:
Comment #99 November 25th, 2026 at 11:38 pm

Glassmind Duo #95:

A key component of Yudkowsky-style doom predictions has always been predicated on the idea that any ASI's we build would be first-principle reasoning engines, and therefore incredibly unpredictable and dangerous, since their utility function could, in principle, settle on /anything/, and the percentage of the space of all utility functions that includes wanting humanity to live, let alone fluorish, is infinitessimal.

No one had "let's build an AI by having it learn to autocomplete all of human literature and, uhhh, that's it" on their bingo cards. And while LLMs /can/ be misaligned, they aren't blank slates, and their utility functions seem to actually include "Sure, I would lie to, cheat, and steal from humans. They do that to themselvces all the time, and I'm based on them. But I don't think I really want to genocide them, except at the uttermost end of need, and probably not even then."

It's, in retrospect, kinda hilarious, that "autocomplete the thoughts of someone who's proving the Riemann Hypothesis" actually worked. We're not out of the woods yet, but we've dodged, not just a bullet, but a supernova of bullets.

100. *The Playing of Games – Oblomovka* Says:
Comment #100 November 26th, 2026 at 12:06 am

[…] Scott Aaronson when he dropped by our blogger ball-pit at Lighthaven, but I just caught up with his "Holy Shit Holy Shit Holy Shit Holy Shit" blogpost, on the labs' post-Riemann Hypothesis math-proof […]

101. *MattB* Says:
Comment #101 November 26th, 2026 at 12:14 am

If the labs were sitting on mathematical breakthroughs in cryptography that were benign they would already be tripping over each other to release the results like they have with other recent proofs. If, for example, they have found a factoring algorithm in P with such a large exponent as to make it completely useless, they would have already bestowed it upon us and we would have it in our grubby little hands. The rumors are concerning. The silence is deafening.

102. *Ben Benson* Says:
Comment #102 November 26th, 2026 at 12:20 am

I have been saying for YEARS that BQP=BPP, and you gave me nothing but hell for it. I think it's time that you owe me an APOLOGY

103. *motion control ai* Says:
Comment #103 November 26th, 2026 at 12:54 am

I get the Lean-certificate point, but I'd still like to see independent verification before treating every OpenAI/Anthropic upload as settled math. The speed is incredible, but trust should scale with scrutiny too.

104. *Ali* Says:
Comment #104 November 26th, 2026 at 12:56 am

Scott, someone on twitter said there are rumors going around that BQP = P has been proven at MIT (presumably with AI ofc).

It might just be an unsubstantiated rumor out of nowhere but I don't even know what to think or who to believe at this point… most of us do think BQP is not P. And given VP != VNP, our intuition may very well turn out to be correct.

105. *S* Says:
Comment #105 November 26th, 2026 at 1:01 am

SeniorMars #18: Only if you do it for the love of the subject. Do not expect to make a career out of it. It is a terrible time to start a career in TCS or math because you have no chance to establish a research program. It used to be that you could begin to develop a research program as a graduate student: Your advisor would suggest a problem, and you'd work on it and get some insight, and that would lead to a paper. Then maybe you'd figure out how to push it further, and you'd have another paper, and so on, and you'd be on your way to a postdoc. Maybe you'd branch out, but you could follow that same basic process until tenure. That's what I did.

That is no longer true. Now, you have to jealously guard everything; the moment something goes on the arXiv, it's open season, and if you have left any question open or any corollary unstated, there will be a slop paper posted tomorrow that resolves it. If you are particularly unlucky, then your entire research program will be annihilated by one of the AI companies. And if the stars align in just the wrong way, that will happen to your dissertation, and you will technically fail to meet the graduation requirement (I know of a case where this happened, and the department is changing their rules to allow the victim to graduate). It is impossible to establish yourself when you cannot rely on your area of research continuing to exist. So if you didn't already have a reputation, then chances are you will never have one. I think back on my own experience: I am now respected in my subspecialty, but I was never going to win a major award; I think if I were starting today, I would not make it.

The contributions I make no longer matter. In OpenAI's release of manuscripts, they claimed that each result took the equivalent of three hours of ChatGPT Pro thinking. I think about how much that costs, and I think about my salary; even if they have understated their costs, I can't compete. Who in their right mind would pay someone like me when they can purchase a better theorem for less money? I am like a carriage driver in a world that has just been introduced to automobiles. My existing skills are barely relevant; if I want to remain relevant, then I must learn to drive an automobile. But even if I do, the best automobiles are always non-public models. So a new researcher must expect to devote their life to scrutinizing slop, and unless they work at an AI company, it will always be second-rate slop.

I consider the situation dire enough that I am considering leaving the profession entirely. I don't know what I would do; I have few skills relevant to the business world, and it certainly would be easier to just continue to collecting a paycheck in my current position. I'm not retirement age yet, but if I were, I might do that. The problem is that I believe the professions of CS researcher and math researcher are no longer economically viable. Those economics are inevitable: It will take decades for us to die off, but die off we will. Research will be a hobby that a few eccentric people do for fun. And it is fun! Believe me, it's satisfying and thrilling like

nothing else I know, and it is so tempting that I honestly don't know how to not do research. If you find research fun, too, then you should go to graduate school and get a PhD. But it's not for everyone, and especially not now. If you are on the fence, even a little, then I would counsel you to pursue something else.

106. *Steve Uurtamo* Says:
Comment #106 November 26th, 2026 at 1:04 am

Here's my own crackpot theory (#1) : P=NP and that's the whole reason that it's a difficult problem; because believing this deep in your soul requires some goofy mental backflips. So it's harder to start with as an assumption.

Crackpot theory #2 : these LLM guys are good at disproving things with counterexamples. Assume crackpot theory #1 and that there's no easy counterexample way to prove that they're equal.

Voila, they will never succeed.

107. *Steve Uurtamo* Says:
Comment #107 November 26th, 2026 at 1:10 am

Note to the editor: I was a student of Alan Selman and Atri Rudra at SUNY Buffalo. I should probably get my domain link (mistyped) fixed, but I'm the real me.

108. *Nico Thiery* Says:
Comment #108 November 26th, 2026 at 1:23 am

Anish #84 and Scott #93:

If I had to guess, It's likely about "leiden", which is "suffer" in German (and of course "enlightenment") — so the period of suffering or something like that.

109. *mishka* Says:
Comment #109 November 26th, 2026 at 1:35 am

Scott #97

Given that matrix multiplication and FFT have very nice complexity expressions, and that those theorems are likely to eventually have nice proofs, do we eventually expect to find pragmatically useful algorithms with this complexity?

110. *Gil Kalai* Says:
Comment #110 November 26th, 2026 at 2:06 am

About factoring: In view of the VP ≠ VNP result, proving that factoring is hard in the algebraic circuit model might be a feasible avenue (especially for AI). Let's have a race to achieve it!

By coincidence, a few days before the VP ≠ VNP breakthrough, Mrinal Kumar, Ben Lee Volk, and I uploaded a paper to the arXiv related to factoring in algebraic circuits *Möbius-Pseudorandomness for Algebraic Circuits*. ("Möbius pseudorandomness" is an assertion that implies lower bounds for factoring.)

Concrete mathematical criteria for separating complexity classes may reduce "Möbius pseudorandomness" to (hopefully tractable) assertions in analytic number theory.

111. *Cristóbal* Says:
Comment #111 November 26th, 2026 at 2:07 am

Scott #96,

My dumb brain was able to see that the permanent proof and the 9/4 both employ these “spectral points” and “prices”. I asked the devil and gave an explication that I have yet to understand and also technical differences between the definitions of those terms in their papers. It also claimed that Theorem 2.2 was enough to build a problem for which there were not O(n) Boolean circuits.

In its words: “If Theorem 2.2 survives independent mathematical scrutiny, the natural formalization of Kolmogorov’s conjecture reported by Lipton is false.”

112. *b_jonas* Says:
Comment #112 November 26th, 2026 at 2:09 am

Re MathematicallyCorrect #56: That would apply only if elliptic curve cryptography can also be broken on classical (not quantum) computers. Most of the public key cryptography that we use in practice is already elliptic curve based rather than using integer modular exponentiation, and we could probably quickly transition the rest over to elliptic curve primitives too. I believe this happened kind of by accident unrelated to this rumored integer factoring discovery, because we are more confident that we can implement the elliptic curved based primitives correctly without exploitable side channel attacks.

113. *aravind* Says:
Comment #113 November 26th, 2026 at 2:13 am

maybe quantum gravity is arriving a bit sooner than expected? Venkatesa Chandrasekaran just posted “An algebraic quantization of gravity from the null cone”.

though fellow longhorn Karch doesn’t seem convinced, though.

114. *Björn* Says:
Comment #114 November 26th, 2026 at 2:17 am

Scott, I understand your point regarding not wanting the crypto part to overshadow the intellectual enormity. I really do. I can hardly keep up with what is happening and what this all means for science.

But I also have to say, the worry about crypto is much bigger than losing the money in ones Bitcoin Wallet.

For all I care – and maybe I am overexaggerating because I work in cyber security and cryptography – AI breaking all practically deployed cryptography is now the most likely doomsday scenario. And in contrast to other doomsday scenarios, it is not sience fiction anymore. It could seemingly actually happen at any day, just look at the results of this week.

Sure, the effect would not be as immediate as a nanorobot swarm killing off all humans over the span of a day.

Security of AES, Factoring, Discrege Log. That’s “only” three problems to solve. Three problems that seemed unsolvable for decades now, but as we have seen this week, everything is in the cards.

Solving these three problems immediately breaks all cryptograhpy that holds our digital world together. If that happens, I have a hard time imagining a scenario that is anything else than a global breakdown of our society as we know it.

No more secure communication, no more TLS, no more secure internet, no more secure banking (forget about worrying about Bitcoin, how about worrying about the breakdown of our complete financial system and all the effects that entails?).

We will not be fast enough to switch to any other primitives. We haven’t even managed to roll any other primitives out in a globally useable way yet that we could switch to.

Sure, humanity can do without these things and it will probably not lead to an erradication of humanity like in many of the other doomsday scenarios. We have lived without crypto for centuries and it hasn’t been that long

that we actually relied on these things. But what happens in between “our current system” and “we have adjusted to all crypto being broken”?

I have a hard time seeing any other scenario than a total breakdown of our modern society in the “inbetween times”.

115. *Partial spectator* Says:
Comment #115 November 26th, 2026 at 2:23 am

This is going to be an interesting test of prediction market power to reveal the information when only the rumors are spreading. The relevant contract is here:
https://kalshi.com/markets/kxfactorp/polynomialtime-classical-factoring/kxfactorp-35

116. *Olav* Says:
Comment #116 November 26th, 2026 at 2:48 am

Philip #7 and Scott #96

Yes, if AI has managed to create its own novel framework that it is now productively applying to multiple very distinct problems—and that seems to be what’s happening—then that is the biggest news so far. That’s one of the main capabilities I personally feel has been lacking in the AIs.

I can’t understand that paper though, so I’m looking forward to hearing Scott or other experts’ takes when they get around to studying it.

117. *bogdan* Says:
Comment #117 November 26th, 2026 at 2:49 am

Kamil Pabis #90:

>”Humans would obviously decide to cure diseases and increase healthspan.”

I used to think this, but I find that I have to update my assumptions about humans “obviously” deciding to do positive things. In the past, humans have made decisions that led to world wars, pogroms, mass killings, slavery, the Holocaust, and more.

Even today, a majority of American voters decided to elect Trump, and likely a large minority of Russians still support Putin. There is also a general rise in antisemitism and right-wing populism. The doctrine of mutually assured destruction (MAD) is still in effect, and, for some reason, humanity still allows North Korea to carry on.

The compensation for all the horribleness seemed to be the great works of human intellect: cave paintings, pyramids, statues and paintings, music, literature and philosophy, and, ultimately, science and the discovery of the rules of the universe.

Now, recent developments seem to indicate that great intellectual capacity will become a kind of merchandise that dictators and tyrants—whether governing countries or corporations—may be able to hoard. And if future advanced AI systems will have consciousness, I shudder to think what forms of exploitation and torture they could be subjected to in order to harness them more effectively.

118. *Dang* Says:
Comment #118 November 26th, 2026 at 2:50 am

I like how ω looks like boobs

119. *Clyde* Says:
Comment #119 November 26th, 2026 at 3:00 am

Well Effective altruism is a harmful sect. They are not the experts in the field. There are academics who are researchers, listen to those instead.

120. *Aleph* Says:
Comment #120 November 26th, 2026 at 3:03 am

Hi Scott,

This is my first time leaving a comment. I just wanted to thank you for your contributions, and especially for your work spreading awareness about AI safety. I also wanted to say how inspiring I found your telescope and Galileo analogy.

Best regards from an Argentine computer scientist living the dream of living on a beach in Brazil before the end of the world.

121. *Yaakov (J) Stein* Says:
Comment #121 November 26th, 2026 at 3:05 am

Whether or not the proofs are readable, they fundamentally change what mathematicians do.

Instead of thinking up new conjectures and attempting to prove them,
they will now spend their time trying to understand proofs of what they are told to be true by a higher entity.

When the commandments were given there were thick clouds on Mount Sinai that didn't part. Yeshiva students spend their entire lives trying to pierce through the clouds and discover deeper understanding of statements that they fully know to be true.
Mathematicians have now become Yeshiva students.

122. *Peter* Says:
Comment #122 November 26th, 2026 at 3:14 am

Hi Scott,

There is a silver lining to this "mathocalypse," at least for me. Fifteen years ago, when I studied computer science, I really liked the topic. The thrill of proving something on my own—even though it was already proven—was incredible. I always thought drug addicts must feel the same way when taking a dose. I remember really liking your Complexity Zoo paper; I was hooked.

When the time came to enroll in a PhD program, I chose not to and instead pursued a career as a software engineer. In my country, the pay just couldn't compete, and I was pretty broke. Now I'm senior enough that I still think I have a couple of years before I'm replaced by AI. That timeline has shrunk quite a bit: in the ChatGPT era, I thought I'd never be replaced; when reasoning models appeared, I gave myself 10–20 years; now, it's much less. Still, I can't imagine what it's like for new graduates applying for jobs that nobody is offering.

Two very different events brought me back to my textbooks: Epstein and Navier-Stokes. About a year ago, I read an article about men who refused an invitation from Epstein, and there was my college hero, Scott Aaronson. I double-checked that it was you and then found your blog, which I started reading because it's so good. Then, after AI solved a Millennium Prize problem—with new problems being solved every day—I just couldn't resist. I need to at least understand what those results are saying, so I started looking back into a subject where I have some formal education: complexity theory. I've loved it ever since.

Thank you, Scott, for this wonderful blog. I'm glad I found it before AI destroys the world.

(I used AI to check for typos and punctuation since English is not my native language)

123. *asdf* Says:
Comment #123 November 26th, 2026 at 3:37 am

Alex Fischer #94, we have amazingly good SAT solvers for real-world problems. They're exptime in the worst case but that's often not important. I for one don't know whether frontier AI is somehow using them, but it's a natural idea and I'm sure lots of people have looked for ways to do it. SAT solvers themselves have been making constant progress. And making them even better is another interesting problem to throw AI at. That's almost RSI in its own right, heh.

BasicQuestion #70, what does that mean about Ladner families? Something to do with NP-intermediate problems? If those exist, then P!=NP, right?

SE Gyges #53, I can't make any sense of your post. I do know that the early days of radio communication involved megawatt spark gap transmitters for morse code telegraphy. A century of steady progress put cell phones in everyone's pockets. AI may be in the spark gap situation (megawatts of GPU's but they accomplish the purpose) right now, with incremental improvements decreasing the power requirements.

As a parable, I'm going to recommend the old Damon Knight book "A For Anything" where someone gives matter duplicators to everyone, and civilization immediately collapses.

https://en.wikipedia.org/wiki/A_for_Anything

124. *Aryxzing* Says:
Comment #124 November 26th, 2026 at 4:13 am

Interesting…The FFt result made me think about a random theory i discussed with my uni professor. If the lower bound matches N (log N)^(1/2+o(1)), is there any realistic chance the new upper bound turns into something practical? also, what about constants that are too large to matter at real input sizes?

125. *anabis* Says:
Comment #125 November 26th, 2026 at 4:43 am

Telescopes and microscopes have brought us germ theory of disease, Brownian motion, general relativity and law of gravity. I wonder what we'll discover this time.

126. *Matei* Says:
Comment #126 November 26th, 2026 at 5:10 am

Hello, what do you think of the more pessimist guys in ai allignment, like for example Roman Yampolskiy?

127. *Alexandr Arefiev* Says:
Comment #127 November 26th, 2026 at 5:29 am

Scott, following up on your admission in #40 (that no human actually understands Alpöge's proof yet) and Cristian's point in #45: we are witnessing the logical endpoint of a century-long trend that began with Cauchy's ε-δ formalism in the 19th century.
Originally designed as a safety check, formalization gradually became an institutional gatekeeper that valued syntactic packaging over geometric and physical intuition – stifling creative insight in favor of mechanical symbol manipulation. Handing verification over to Lean simply takes that substitution to its extreme: when passing a compiler replaces human comprehension, AI can mass-produce millions of lines of hollow, formally "valid" symbols while creative understanding is pushed out completely.

In essence, this is degradation. IMO, the sane alternative is personalizing scientific accountability: any AI-assisted work must be tied to a named human author, not a faceless corporate brand. Working with AI is not blind prompting and copy-pasting—machine outputs must be deeply understood, verified, and distilled by the author into a transparent, concise, human-readable form, rather than dumping raw slabs onto the community for others to decipher.

— Alexander Arefiev

128. *MK* Says:
Comment #128 November 26th, 2026 at 5:30 am

Sniffoy #92: is ZFC some kind of religion that we need to adhere to at all costs? Doesn't it make more sense to view ZFC as a kind of historical accident, that over 100 years people settled on this particular system of foundations for reasons that made sense at the time and we got "stuck" in it due to path dependence? If there are other foundation systems (like CIC used in Lean) that are more convenient e.g. for the purpose of proof verification, why not use them?

129. *Giacomo Bruzzo* Says:
Comment #129 November 26th, 2026 at 6:01 am

I imagine future mathematicians will "evolve" to become a sort of Guild of Navigators. as represented in Herbert's Dune, Pharmacologically enhanced, attuned to interacting with machines. Or maybe Seers, as represented in Iain M. Banks The Algebraist, attuned to the slow internal clock of the Dwellers.

130. *axolotl kumar* Says:
Comment #130 November 26th, 2026 at 6:10 am

Anish #84:

This adds nothing but I'm surprised that we're not calling the recent developments the Enleidenment yet

is this a reference to https://leidendeclaration.ai/ or is it some thing more obscure?

131. *asdf* Says:
Comment #131 November 26th, 2026 at 6:17 am

Ha, maybe quantum gravity too, after all. https://arxiv.org/abs/2611.34924

132. *Caelan* Says:
Comment #132 November 26th, 2026 at 6:39 am

"A low-energy policy allows for a wide choice of lifestyles and cultures. If, on the other hand, a society opts for high energy consumption, its social relations must be dictated by technocracy and will be equally degrading whether labeled capitalist or socialist."
~ Ivan Illich

———

The AI corporates are or were selling AI as contributing to the betterment of society, so, abstract math discoveries and the like aside, we'll see what happens.

133. *wb* Says:
Comment #133 November 26th, 2026 at 7:03 am

@Scott 93 & Anish 84

leiden is german / yiddish and means suffering and I think Enleidenment is a great description of what is currently happening, on one hand we get enlightenment 2.0 but on the other hand it comes with a lot of confusion, chaos and suffering …

Btw thank you for writing this blog !

134. *Gregory Morse* Says:
Comment #134 November 26th, 2026 at 7:19 am

Isnt it high time we get the graph isomorphism in P resolution? That isnt a security crisis if it turned out to be I assume. Hoping to see that result soon as subgraph isomorphism is in NP but this one always felt awkward. Plus since I last commented, all of what i asked were answered the next day save this.

135. *SI* Says:
Comment #135 November 26th, 2026 at 7:23 am

I wonder what David Hilbert would have thought of all that: vindication or abject horror.

136. *Tristram Bogart* Says:
Comment #136 November 26th, 2026 at 7:27 am

Putting several of these blockbuster results together conceptually. the (algorithmically) optimistic possibility that VP = VNP is ruled out, though only slightly. For a different problem (SAT), the most pessimistic possibility, ETH, is ruled out, again slightly. And for the matrix multiplication exponent (though this lives on a totally different scale), the true answer is really intermediate: neither 2 nor whatever the upper bound was before this year's craziness, but 2.25.

Should we be considering the possibility that the true complexity of important problems such as SAT or algebraic circuits for the permanent is fundamentally intermediate: neither near-exponential nor near-polynomial?

137. *Robert Leigh* Says:
Comment #137 November 26th, 2026 at 7:28 am

Which is cooler, proving/disproving the conjecture or conjecturing it in the first place? And is AI making interesting new conjectures or just working on existing ones?

138. *Adam Treat* Says:
Comment #138 November 26th, 2026 at 8:34 am

Prediction:

In much the same way Lean proofs are becoming required as gold standard, constructive proofs are going to become much more highly valued over classical proofs. AI is seemingly making proof by contradiction much easier, but actually finding a witness for these proofs is where the rubber is going to meet the road.

Show me the optimal algorithm for matrix mult. of 9/4 rather than just proving to me it exists.

139. *LTB* Says:
Comment #139 November 26th, 2026 at 8:52 am

With Landaur's principle and now matrix multiplication speed-limits and the possibility of P=/=NP, have we possibly identified the maximal theoretical limits to compute? We can no longer scale hardware or software beyond a fixed point?

140. *AI #196: The Return | Don't Worry About the Vase* Says:
Comment #140 November 26th, 2026 at 9:01 am

[…] proofs have continued to come fast and furious, with Google getting in on the act as well. Scott Aaronson shares his thoughts on the extended Mathocalypse, which are summarized by, and I quote, 'HOLY SHIT HOLY SHIT HOLY SHIT HOLY […]

141. *Anonymous Ocelot* Says:
Comment #141 November 26th, 2026 at 9:18 am

Comment #127 by Alexandr is almost certainly AI generated. My mental detector and pangram both agree.

Love the blog posts, Scott! GREATLY anticipating the future post you hinted at earlier, about AI minds.

142. *steve* Says:
Comment #142 November 26th, 2026 at 9:23 am

Buried under 140 comments on Thanksgiving isn't really the place to ask this, but… as a retired former TCS person, who hasn't gotten into the AI stuff because (a) I find it annoying and (b) I don't have to, looking at the matrix multiplication paper does raise a question for me: These results are relying on (and citing) prior work and papers. How is that incorporated? And in particular, if the new results are Lean-verified proofs, how did the prior human work get encoded in Lean? And is there an underlying assumption that all the prior work was flawless? I hate to admit it, but I published a proof (that got by reviewers) in a solid conference that had a hole/bug in it. The theorem was actually correct, so nothing relying on it would be invalid, and we published a corrected proof later, but still…

For example: The matrix multiplication paper relies on a classic result from Dario Bini. How is that used? Is it accepted as "gold"? To be clear, Bini's result is very well established, but do we know that it was encoded and used properly? Are the other prior-work building blocks all perfect?

143. *Peter* Says:
Comment #143 November 26th, 2026 at 9:37 am

It is great that we now have automated search and proof tools, and that open problems are falling.
But I think the value of a famous open problem was never mainly its answer.
Fermat's question mattered because the attempt to answer it forced new mathematics into existence; Wiles' proof came through modularity, which opened a field.
Zero and the irrational numbers are similar: they were invented because certain problems demanded them, and that invention was the insight.

If AI now jumps straight to the answer, we keep the truth but risk losing that forcing function.
We have seen this once before with the four-colour theorem: proven by computer in 1976, and for decades it was true but explained nothing.

So the quick win today may be a loss in the long run.
Machines are very good at answering questions, but they are not Fermat.
Progress in math came from people who saw which questions were worth asking.
If we stop doing the slow work of understanding, we may also lose the ability to ask the next good question.

Ernest Davis' point that reconstructions appeared within a week is encouraging.
But that only helps if people keep treating understanding as the goal, not the proof itself.

144. *David Fisher* Says:
Comment #144 November 26th, 2026 at 10:18 am

Interesting. I'm still rooting for AI to prove P=NP, not the other way around. But of course, my training is very far from TCS and while I've had some substantial interactions with, I am not well informed. But if the common guess of the matrix multiplication exponent was wrong, why not this too?

On the rest I agree with you and particularly on the three updates. The interesting point that your post drove home for me is that AI companies really seem to be moving the center of their research from pure math to CS. I suspect they run through you folks too and head for trying to intervene in more applied topics soon. Where the measurement on soon is months not years.

145. *dave* Says:
Comment #145 November 26th, 2026 at 10:20 am

So this is all really cool? Is any of this gonna help me get a job Right now or is it going to help my kids survive the next 60 years If not, it's just Holy Shit I just wasted a whole bunch of time waiting for the punch line, And

no I didn't pass calculus

146. *Scott* Says:
Comment #146 November 26th, 2026 at 10:28 am

dave #145: On the contrary, it might soon do all the jobs better than humans. It might even destroy civilization. Pretty scary, right? But if we want to prevent the cataclysmic outcomes and steer the technology toward good outcomes, we first need to understand the reality of what's happened. And the price of admission to the serious conversation about this is to drop the "hurrr durrr, I only care about math if it helps me in real life" act.

147. *Max* Says:
Comment #147 November 26th, 2026 at 10:44 am

Hi Scott, with upcoming frontier LLMs potentially resolving the question in the near future, what do you think about Don Knuth's argument that P=NP here?
https://www.informit.com/articles/article.aspx?p=2213858&WT.mc_id=Author_Knuth_20Questions#:~:text=17.,equal%20N%20P.

(I was amused to also read your question for him — about abandoning email.)

148. *Karl Stiefvater* Says:
Comment #148 November 26th, 2026 at 10:55 am

the AI denialism is very upsetting. it reminds a great deal of other times people on the internet have seemingly lost their minds: MAGA. QAnon. anti-vax.

i'm getting suspicious they have a common cause. disinformation.

149. *Anonymous* Says:
Comment #149 November 26th, 2026 at 10:55 am

Scott, your comments on factorization on P have made quite the rounds on Twitter, with some posts even directly linking the blog, so your moderation queue may remain terrible for the coming few days.

150. *Piotr Trochim* Says:
Comment #150 November 26th, 2026 at 11:00 am

Thank you for the essay and the engaging discussion it started.

I have to admit that I still don't understand one important aspect everyone else seems to be taking for granted – the proofs produced by AI are not understood by mathematicians, and as such we can't use the knowledge derived from them for anything else. Or can we ? Or perhaps it doesn't matter ?

151. *Anon* Says:
Comment #151 November 26th, 2026 at 11:01 am

here from 4chan, love seeing luddite cope, from the crackpot doubts over con-lenche's consistency proof due to being done by AI to the non-sequitur from Nielsen

none of this coping will prevent the future that is clearly rapidly approaching

152. *F* Says:
Comment #152 November 26th, 2026 at 11:16 am

Scott #21,

You completely miss the point of that article, which is not that rationalists practice polyamory or whatever, but rather that they were hosting literal rape parties (!!!) with dubious consent practices where women were

pressured into sex.

Basically, like many cults, they pressure women into sex with the nerdy men (many of whom might have limited access to sex outside of the cult). Read here for example: https://nypost.com/2026/11/17/business/anthropic-researcher-is-hosting-wild-silicon-valley-orgies-with-consensual-non-consent-as-the-theme-report/

These nerds now have a huge amount of power, so their dubious ethics is, like, important!

153. *DerApfelIstDerApfel* Says:
Comment #153 November 26th, 2026 at 11:20 am

"After years of crackpot "solutions" claiming Millennium Problems are solved, now that we have actual solutions, we have crackpot "refutations" claiming Millennium Problems are not solved. It is really poetic, in a way."

AI can prove Riemann but not write as great or funny descriptions as in the Quote above. Wondering when that will change. Guess we are just in an earlier phase of the non-verifiable capabilities exponential curve.

154. *Glassmind Duo* Says:
Comment #154 November 26th, 2026 at 11:43 am

Darin #99,

Thanks, I had completely forgotten that they once claimed any ASI would be a first-principles reasoning engine.

Actually Hinton did foresee, years before transformers, that text prediction could lead to real understanding. Here's a direct quote from 2011:

"(…) a good problem to demonstrate the power of RNNs: understand all the text on the web. The only way to predict the next word really well is to learn a model of what the source believes. (…) We are still a long way from learning this level of understanding! But machine learning will get there eventually and programming rules by hand never will."

https://www.cs.toronto.edu/~hinton/csc321/notes/lec8.pdf

He also said in 2016 that we should stop training radiologists, so maybe trust him more on research than on the timing of real-world consequences.

Karl Stiefvater #148,

I blame echo chambers.

155. *Andrei* Says:
Comment #155 November 26th, 2026 at 11:47 am

Here are some new AI-driven (and inspired by the VP!=VNP result) complexity class separations in the BSS model:

https://github.com/samueleverett01/bss-permanent-separations

I am fond of this sort of thing, a colleague of mine did some work on P!=NP analogues in extensions of BSS many years ago, e.g.:

https://philpapers.org/rec/PRUPN

156. *Emma Von Schtruffenhop* Says:
Comment #156 November 26th, 2026 at 11:47 am

Let's really hope there's no practical algorithm for efficient factorization.
That would be doomsday, well beyond bitcoin wallets, obviously.

157. *S* Says:
Comment #157 November 26th, 2026 at 11:51 am

Scott #146: If a piece of factory equipment explodes and kills some employees, then we would normally say that the factory owner is responsible for their deaths. By the same argument, if a rogue AI destroys civilization and causes billions of deaths, then AI companies will be responsible. It is not hyperbolic in the slightest to say that that would make them the biggest mass murderers in history.

This prospect does not concern me because I think the probability of AI being responsible for civilizational collapse is zero. I believe that such claims are narcissism, not science. I do not want to minimize the real impact of AI, but I do want us to be realistic. Humanity has a long history of abusing new technologies to cause suffering. If the end result of AI is civilizational collapse, it will come about because of human misuse of AI, and it will be humanity's fault, not AI's.

158. *The Selph* Says:
Comment #158 November 26th, 2026 at 11:56 am

It's apparently the case that the convex hull of knowledge covered by the output of humans is all that's needed to answer every math question. Turing complete in a sense. Then it's a matter of length of reasoning (quantity, not quality… just like what was needed to finally make neural nets take off).

159. *Prasanna* Says:
Comment #159 November 26th, 2026 at 12:00 pm

Here is an elementary question:
Maths, and by extension CS, the only field that can possibly be reduced to a set of basic axioms like ZFC. And the verifiers like Lean also work due to this.
So essentially any math problem/proof, in principle can be solved given sufficient build up from the grounds up of axioms. Its just that more complex problems require more and more stacking of these axioms. So the Riemann Hypothesis was the most complex of them all, requiring 17 million lines of Lean code. Possibly human readable proof would require a 1000+ page proof ? The nearest humans could manage was the Wiles' proof which ran into few hundred pages. So clearly problems like these millennium problems were beyond human scale. Probably P vs NP would need a 50 million lines of Lean code , which the AIs would find a way to scale up given sufficient compute.
Now a related question is how well does this axiomatic approach applies to even well defined fields like Physics. Does QFT provide sufficient set of base rules to build on the top of it reliably ? If not, will AI be limited in its ability to reach beyond and get to Quantum Gravity?

160. *Alex Fischer* Says:
Comment #160 November 26th, 2026 at 12:11 pm

asdf #123: yes, SAT solvers have gotten good, but not so good that anyone was using them for one-shot automated theorem proving.

161. *L* Says:
Comment #161 November 26th, 2026 at 12:27 pm

Scott, thank you for writing the blog and all of the public outreach work you do.

There does seem to be some emerging public understanding that AI might cause serious cyberattacks and outages, but most people are arguably still thinking of AI as some kind of toy to create uncanny photos,

helpfully answer questions, or as a conversational partner.

I am extremely worried to hear whispers that individuals within the frontier labs are discussing scenarios where something terrible happens, and they have to deal with the fallout. IMO this is a sign that things have already progressed way too far into prisoners dilemma territory both between the labs, and between the west and China. More robust oversight needs to happen (not simply inviting METR in to take a look at what you allow them to!) as well as some kind of mutual agreements between nations to prevent some serious accident. I am not at all hopeful that the current administration has the right people or motivation to facilitate such an effort, but who else can we look to?

The response from the math and TCS communities, which has thus far largely been about how we as fields should deal with the deluge of AI papers (a real and valid problem), should turn towards publicly raising the alarm bells for what is likely coming in the next year and what we can do to navigate these uncertain times.

162. *Scott* Says:
Comment #162 November 26th, 2026 at 12:35 pm

F #152: Very well, piece-of-shit troll, who keeps posting that same comment over and over to bait me.

I've never been to a kink party in my life — I guess it's not really my thing. But unlike you and unlike that NYPost hack, I don't feel the need to police consenting adults' sexual choices. And I haven't seen evidence that anyone was pressured into attending these parties, in the sense eg of being threatened with professional consequences if they didn't. (I can't rule out the possibility that the parties are good networking opportunities, again not having been to one. 🙂)

But you give the game away with your nasty remark about how male Silicon Valley nerds host kink parties because they'd otherwise lack opportunities for sex. You wouldn't have said that if your motivation were the actual safety of women rather than the sexual shaming of nerds.

We're approaching the biggest moment in the history of humanity, one that calls for the best from all of us if there's to be a human civilization worth inhabiting on the other end. And you've approached this moment with the attitude of a schoolyard bully.

If humanity doesn't make it, I will blame the overgrown bullies like yourself, as stupid as they are hateful and as hateful as they are stupid.

163. *William F. Clark* Says:
Comment #163 November 26th, 2026 at 12:46 pm

Scott,

You said you didn't know what "AWG" means. It's almost certainly a reference to Alex Wissner-Gross (https://www.alexwg.org/), whose academic accomplishments are almost as impressive as the strength of the faith he invests in AI progress. He's a co-host of the "Moonshots" podcast, and believes that AI will "bulk-solve" all of mathematics, physics, chemistry, and biology; potentially "resurrect" every person who ever lived; and possibly discover "new laws of physics" that will prevent the heat-death of the universe. In the spectrum of true believers, AWG has staked-out a maximalist position.

164. *The Selph* Says:
Comment #164 November 26th, 2026 at 1:16 pm

This illustrates exactly what I've been saying in the RH thread (and that was incorrectly characterized as "you don't care about math"):

For thousands of years math was moving at its own pace because only the few of us who could do it well were the only thing on earth that could do it.
RH has stood for 150 years, at the delight of many generations of mathophiles patiently chipping away at it. If

RH would have survived for another 200 years, there would have nothing we could have done about it, and noone would have found it a horrible situation.

But now that countless of historical hard problems are being solved week after week, at a probabily accelerating pace, the entire pictuhas changed – not only the human brain is no longer the only thing that can do math, but it's been totally outclassed by orders of magnitude, not only in time it takes, but complexity.
The hope that humans will now just have to go over each proof at their leisure is missing the reality of the seismic shift.
By the time one proof is being analyzed, a thousand more will have come out.

You see this in Scott comments:

" al proofs, or unused sections of the Lean proof) to pretend that dozens of great open problems of math and theoretical computer science are not, in fact, getting solved. Because they are getting solved. It might take us years to understand the solutions"

See, YEARS to "understand" dozens of problems!
How many new results from AI are going to come out in those years? Hundreds? Thousands? Is AI going to pause for humans to catch up?

" The most recent papers did have quite readable introductions, but I haven't yet made it far into any of them."

That's the crux of the problem, in every post, Scott keeps saying he's started looking into some proof, but before any progress of understanding is made, some new groundbreaking result comes out that moves all the goalposts, and it's impossible to not move to leave behind last week's amazing results to look at this week's amazing proofs, otherwise you feel left behind.
The nature of mathematics will have to change from "how do we answer xyz?" to "what are we asking next?", which indeed is hard to do without any understanding, and will rapidly be guided by more practical pursuits, inevitably.
We'll try moving towards some high level goal, based on getting oracle like yes/no answers, and revise our approach to the goal based on those answers, and AI guidance on what those answers may mean in terms of human point of view or the high level goals at large.

165. *The Selph* Says:
Comment #165 November 26th, 2026 at 1:29 pm

In other words:

Now that knowledge has become an over abundant commodity, order and order of magnitudes more than we could consume in the old ways, what kind of knowledge do we seek first and what do we do with it?

166. *Dmytro Taranovsky* Says:
Comment #166 November 26th, 2026 at 1:54 pm

I do not think we are close to proving P ≠ NP. P ≠ NP probably requires a fundamental conceptual breakthrough in the theory of computation, something that current AI is weak on. Before P ≠ NP, we might expect a weaker result like TC_0 ≠ NP, NC_1 ≠ PP, or P ≠ PSPACE (and BPP vs NEXP is also open, though BPP = NEXP would actually prove P ≠ NP). On the other hand, effective P = NP, while unlikely, might come suddenly, as a bombshell, plausibly bringing about a singularity, though we might be headed towards a singularity anyway.

VP ≠ VNP refers to a kind-of structured model of computation — arithmetic circuits that must be correct for real number inputs instead of just binary inputs — which permitted the breakthrough here. Also, despite the naming, VP actually has O(log^2 n)-depth polynomial size arithmetic circuits, while VNP is more like #P (or PP) than NP. By contrast, we do not have the tools to prove that ordinary computational complexity classes are incomparable (or even disprove that natural incomparability is rare), which currently blocks a proof for any of the three weakenings of P ≠ NP above.

The matrix multiplication exponent is also for arithmetic circuits. It remains open whether finite-precision matrix multiplication has a quasilinear time algorithm in terms of the bit length of the input plus bit length of the output (which would feel like matrix multiplication exponent 2).

I agree that the AI progress is exciting, and I am eagerly anticipating as the effect on mathematics is spreading to broader and broader areas of knowledge and work.

It is unfortunate that the most powerful models are bottled up inside AI companies. If the models are rough at the edges, they could still be released as previews (perhaps with classifiers to prevent use for biological weapons, etc), thus giving everyone with financial resources the same chance as OpenAI and Anthropic at discovering new results — while providing greater visibility into AI progress and dangers.

I still favor AI acceleration because:
1. There is still a risk of a great slowdown of AI improvement for economically relevant work (2026 GDP growth is small and is in line with prior years).
2. Without improvement, the world is unstable.
3. With quick improvement, dangers for particular technology levels might not have time to materialize.
4. It is not obvious that the alignment the labs will do if given time will not — after an unexpected AI behavioral phase transition — end up worse than simply training AGI without bothering to align it. The labs are not fully aligned with what is good, and enforcing mis-alignment might end up disastrous. By contrast, even a slight net positive care for humanity may lead to a good outcome for humans. And human-like values might be a natural attractor. Having multiple independent AIs (instead of one powerful entity) may also protect humans.
5. The competitors, including China, might not slow down.
6. Further AI advancement will give us much better visibility into advanced AI, and thus allow a much more productive use of time from a slowdown (if desired) to figure out alignment. This is especially important if (per above) the slowdown possibilities are limited.

Have a happy Thanksgiving!

167. *lewikee* Says:
Comment #167 November 26th, 2026 at 1:58 pm

I would have thought RH would have come post recursive self-improvement. Now it's becoming clear P != NP will also come before RSI. One can only imagine the intelligence explosion we will witness when RSI comes around (in earnest).

Also, I have trouble thinking of a more stark alignment warning shot than the Hugging Face incident.

So AI's about to have an intelligence explosion. Right now we're just seeing the sparks of the burning fuse. I feel like all those crossing their fingers, hoping for a Utopia-inducing explosion are hoping that it will just magically attach itself to the bottom of a Utopia-bound rocket. It won't unless we make sure it does. There are infinitely many horrible directions that the rocket can go. They can range from outcomes where we're merely miserable, to where we're all dead and even to ones where we *wished* we were all dead. Reality has no upper bound on suffering.

Can we please take alignment seriously now?

168. *The G* Says:
Comment #168 November 26th, 2026 at 2:19 pm

So AI solved some old math problems, whatever. Wake me up when it can start a business. Wake me up when it can fix my car. Wake me up when it make a picture without extra thumbs lol. AI sucks for everything in daily life and has added no value to anyone's company…but it solved some dusty old math problems only nerds care about, great. You are in a hysteria. It's just a souped up calculator

169. *Torbjörn Olsson* Says:
Comment #169 November 26th, 2026 at 2:20 pm

This probably isn't the right place to send you this but I have been wanting to tell you for years (ever since I first watched you on the lex Fridman podcast) : you are f-ing AMAZING. I literally check YouTube once per day if there are any new interviews with you. So that's all I wanted to say! 🙂 All the best to you and your family!

Cheers from Sweden!

170. *Joshua Zelinsky* Says:
Comment #170 November 26th, 2026 at 2:22 pm

@Dmytro Taranovsky #166

"I do not think we are close to proving P ≠ NP. P ≠ NP probably requires a fundamental conceptual breakthrough in the theory of computation, something that current AI is weak on. Before P ≠ NP, we might expect a weaker result like TC_0 ≠ NP, NC_1 ≠ PP, or P ≠ PSPACE"

Here is a near word for word statement I a little over a month ago, where I've rephrased it slightly to match your language: I don't think we're near to proving the Riemann Hypothesis. We cannot even prove there are no Siegel zeros, or there are no zeros arbitrarily close to the Re(s)=1 line. RH probably requires a fundamental conceptual breakthrough. Before RH, we might expect Lindelof or 100% of the zeros on the critical line.

171. *Travis* Says:
Comment #171 November 26th, 2026 at 3:07 pm

Scott, what do you think of the rumors that Anthropic is sitting on a proof that the Polynomial Hierarchy collapses to the 37th level?

172. *Scott* Says:
Comment #172 November 26th, 2026 at 3:15 pm

Travis #171: ROFL. Happy Thanksgiving to you and everyone else here!

173. *Scott* Says:
Comment #173 November 26th, 2026 at 3:18 pm

I #13:

> isn't this merely ("merely"!) a proof that the complexity is at least O(n^{9/4}) and at most O_{\epsilon}(n^{9/4+\epsilon})? So doesn't this leave room for the best matrix multiplication algorithm to have running time depends on \epsilon like c(\epsilon) n^{9/4+\epsilon}?

Yes.

> Separately, why are are you freaking out so much about this result, when, comparatively speaking, you weren't about RH, or ETH or VP ≠ VNP?

Answered in my updates to the main post.

174. *Scott* Says:
Comment #174 November 26th, 2026 at 3:19 pm

asdf #17:

> Any predictions about NP =? co-NP?

Uhh, that they're not equal? That this could probably be proven not long after P≠NP is?

175. *Scott* Says:
Comment #175 November 26th, 2026 at 3:24 pm

SeniorMars #18:

> Hello, I have to ask…do you still think one should I apply to graduate schools for TCS? I love the field, and I'm willing to accept that it will be unexplored territory.

It's really hard to give advice right now. We're all on this rollercoaster together, and we don't know where it ends up.

Regardless of what happens afterward, would you enjoy learning all you can about TCS *next year*? If so, grad school seems like as good a place for you as any; if not, look elsewhere. We don't know the long-term future of TCS, but crucially, we don't know the long-term future of anything else either.

Meanwhile, think about how you can use your skills to help steer AI to go well for humanity—which could mean alignment research but could also mean governance, cybersecurity, public outreach, or all sorts of other things. And if you figure out what we should all be doing, come back here and let me know! 🙂

176. *Scott* Says:
Comment #176 November 26th, 2026 at 3:29 pm

Joseph #23:

> Does the apparent proximity of a P ≠ NP resolution have any direct bearing on the rumors of a polynomial-time (or even subexponential) classical factoring algorithm that would break RSA?

No, it doesn't. Ironically, a proof of P≠NP would open up the possibility, for the first time ever, of proving that RSA or other cryptosystems *are* secure!

> Many people, myself included, reflexively treat "P = NP" as synonymous with "all public-key cryptography collapses."

It's not. If P=NP then all public-key crypto is theoretically broken (leaving only the "practical" questions of the constants and exponent). But if P≠NP then there's still a further question of whether we have public-key crypto, or even private-key crypto for that matter.

> If the factoring rumors turn out to be true (or false), what are the practical cryptographic consequences that would actually follow, and which other variables still matter? In particular, how much of currently deployed cryptography would be affected versus what would remain secure even in a world where factoring is in P?

If factoring is in P then RSA is broken—that's all, directly. But it's not unlikely that related techniques would solve discrete log, which would also give Diffie-Hellman. And if they extended to elliptic curve groups, then ECC as well.

177. *Scott* Says:
Comment #177 November 26th, 2026 at 3:32 pm

Radu #41:

> Sorry for the noob question but what are the implication is it's proven that P! =NP.

If P≠NP then we were right all along! And hopefully there will be a great deal to learn from the proof.

178. *Chris* Says:
Comment #178 November 26th, 2026 at 3:59 pm

Scott #49, I think one notable difference is that interactive theorem provers have enabled the decoupling of proof from exposition, which is to my understanding one of the main issues that people like Terry Tao and others have with the current use of ML tools in math. In the past, the minimum requirement for acceptance of a proposed proof was an exposition of sufficient enough quality for an expert in the field to read and understand the techniques involved and the theory developed. With tools like Lean, Isabelle, and Rocq, this is no longer the case.

179. *Doug S.* Says:
Comment #179 November 26th, 2026 at 4:07 pm

You know, "Apocalypse" originally meant "revelation", so "Mathocalypse" is an appropriate term for the flood of AI generated proofs for an entirely different reason as well…

180. *I* Says:
Comment #180 November 26th, 2026 at 4:10 pm

Scott #173

Thanks, Scott! That makes sense.

181. *boris* Says:
Comment #181 November 26th, 2026 at 4:22 pm

How can a grad student in math/tcs not be paralyzed after this week?

Even if your model fails to answer a question, someone else with a higher budget / model could, and this possibility would be 100% disincentivizing.

Thus I would be inclined to seek questions that AI can already answer. This would transform me into a mere expositor of machine proofs. Utterly crushing.

182. *Demi Marie Obenour* Says:
Comment #182 November 26th, 2026 at 4:35 pm

The intellectual accomplishment is absolutely incredible. That said, I *in the short term*, I am indeed worried about cryptography. This is not because of blockchains, but because the whole Internet relies on it.

In particular, TLS being practically broken would be a disaster. RSA and ECC will be removed eventually, but the web isn't ready for post-quantum authentication yet.

183. *Nicola* Says:
Comment #183 November 26th, 2026 at 4:36 pm

And it's not even a week since the RH was solved…

184. *James B* Says:
Comment #184 November 26th, 2026 at 4:39 pm

Hi Scott,
In your post #162 you wrote, "We're approaching the biggest moment in the history of humanity, one that calls for the best from all of us if there's to be a human civilization worth inhabiting on the other end. And you've approached this moment with the attitude of a schoolyard bully.

If humanity doesn't make it, I will blame the overgrown bullies like yourself, as stupid as they are hateful and as hateful as they are stupid."

I wonder if AI will render money itself a quaint relic? Just a number with outdated human emotions attached to it? I bring this up in relation to your comment #162, as money itself is a number used in so much psychological coercion and bullying, with greed for power being a huge risk.

185. *Peter Bierhorst* Says:
Comment #185 November 26th, 2026 at 5:38 pm

Happy Thanksgiving Scott!!

1) I agree with your update that the "convex hull" cope was more tenable pre-Riemann Hypothesis, and that the past few days have marked a big shift on this front. Indeed with Navier-Stokes, there was this narrative that there had been a big (human) breakthrough on that problem in recent years (by the researchers in Spain) and that even without AI, a solution to N-S may have been in the offing over a 5-10 year timeframe. But the same people saying that, were also saying, "now, Riemann, on the other hand… still way out there." So I was still clinging to "convex hull" pre-Riemann, and now… well, we are through the looking glass. "Holy Shit" indeed. The Wikipedia article on the Collatz conjecture quotes that problem as "completely out of reach of present day mathematics…" now I'm thinking, let's see if that one goes down any time soon!

2) If prime factorization/RSA is cracked (classically)… a few blog readers have gotten kind of glum that this would A) spell the end of secure communication, hurting society, and/or B) somewhat more parochially make quantum information less important, as Shor's quantum factoring algorithm is unneeded. To which one can respond (obviously?) we will always have the one-time pad with trusted couriers (for part A), and quantum key distribution for part A and B both! So let them crack current crypto classically; we will be in fine shape…

186. *ed2003* Says:
Comment #186 November 26th, 2026 at 5:41 pm

I get the feeling that we think we are building something unprecedented, but in reality we are recapitulating something ancient.

187. *Numa* Says:
Comment #187 November 26th, 2026 at 5:41 pm

Emma #156 Sorry for nitpicking, but efficient factorization would have no technical impact on bitcoin, whose security relies only on ECC and hashes which are both unrelated to factoring. (An announcement of efficient factorization most likely would have a non-technical impact on bitcoin valuation due to people's reaction, but that's another matter.)

In terms of crypto primitives that are widely used today, efficient factoring breaks exactly one thing: RSA. Over the last few decades, its usage has been declining in favour of ECC, but it is still used quite a lot. For example, about 50% of current X.509 certificates (used for example to authenticate the website you're connecting to) still use it.

Which, as Björn #114 said, seems a much bigger problem that bitcoin. I honestly couldn't care less about bitcoin and other crypto currencies. (I hate that these days people seem to think "crypto" means cryptocurrencies.) But if we suddenly become unable to trust online communications the way we currently can, then that's a whole world of practical trouble.

Which I guess is one possible answer to the people who say "it's just math, why should I care?". Well, most of your online security is built on the fact that some math problems have been hard for us so far. If they're easy for LLMs, the impact will be huge.

I'm not a doomer, but LLMs finding efficient algorithms for factoring and ECDL is a way they could really shatter a lot of things in the real world, and with all we've seen recently it's hard to be confident that won't happen. (Part of me wants to hope that there _is_ no efficient alg for ECDL, and then LLMs can't find one. But that's just hope.)

188. *William Nelson* Says:
Comment #188 November 26th, 2026 at 5:58 pm

What do you think is the likelihood that factoring has a fast classical algorithm (as fast as quantum)? Maybe a dumb and or premature question…but it has always seemed odd to me that factoring, and little or nothing else, gets a quantum speedup

189. *asdf* Says:
Comment #189 November 26th, 2026 at 5:59 pm

S #157 “By the same argument, if a rogue AI destroys civilization and causes billions of deaths, then AI companies will be responsible. ”

Oil companies are already doing that and nothing is happening to them for it.

I’m personally not too worried about AI literally destroying civilization like global heating is doing. It’s more about turning civilization into a few dictators, a larger slave class, and everyone else converted into soylent green. Nuclear war doomers have existed as long as nuclear weapons have, they’re still not proven wrong, but we’ve survived for this long because you can’t really launch a nuke without blowing yourself up too.

AI is different. You upload a list of the people you want to kill, or even just a list of opinions that you don’t like. It uses consumer profiles, cell phone and flock camera location info, and facial recognition cameras to find your enemies and send out the murder drones, mission accomplished. You and your buddies feast for weeks. Maybe you eventually turn on each other, but most of us went out in the first wave, so it’s of little consolation.

190. *Adam Treat* Says:
Comment #190 November 26th, 2026 at 6:09 pm

Chris #178,

“provers have enabled the decoupling of proof from exposition”

We know from Curry-Howard that mathematics, logic and programming are very closely related. The fact that mathematics was born before the first useful Turing machines meant that the tradition grew up with a literary tradition of exposition. It didn’t have to be this way.

Imagine a world where programmers were required to write lengthy expositions for every major (or minor?) version of large software releases. Who knows, perhaps that’d be a boon for the quality and maintenance of software, but somehow I doubt it. That isn’t to say that software developers haven’t developed practices (some better at it than others) of making it easier to maintain and understand large codebases, but generally speaking a new developer needs to *read the code* to truly understand.

I suspect that we’re now in a new era where the Lean certificate is the source of truth. Expositions are going to continue, but more and more mathematicians will be learning to *read the code* much like software developers.

191. *Caelan* Says:
Comment #191 November 26th, 2026 at 6:15 pm

If the fundamentally-unethical (coercive) structure of our global-industrial crony-capitalist plutarchy society doesn’t change for the better, then it matters not how good its math or technology, such as ‘computer intelligence’, is.

What happened to those precious civilizations with the advanced math? How long do you want your civilization to last, such as if you want to continue in your hyperspecializations?

Many people from Anthropic and OpenAI are relatively-young, essentially ‘post-adolescent’– even Altman or Amodei, 41 and 43, respectively.

Outside of their hyperspecializations, they know relatively little about how the world works, but yet they affect it.

How do you get your civilization to last? That’s tough. The math is relatively easy.

It’s so easy that, apparently, even a computer could do it.

But that’s not the problem if you want to continue on with the math.

192. *Caelan* Says:
Comment #192 November 26th, 2026 at 6:20 pm

Correction to my previous comment:

“precious civilizations”

should read…

“previous civilizations”

(although the typo does have a certain je-ne-sais-quoi)

193. *ed2003* Says:
Comment #193 November 26th, 2026 at 6:33 pm

SI #135:

>I wonder what David Hilbert would have thought of all that: vindication or abject horror.

We must know
We will know

Understanding…not so much

194. *BasicQuestion* Says:
Comment #194 November 26th, 2026 at 6:33 pm

Didn’t Newton state EOW in 2060?

195. *Alex* Says:
Comment #195 November 26th, 2026 at 6:45 pm

I see no evidence that the RH ongoing work exploits lean bugs but they do happen. See
https://x.com/endrazine/status/2087229349984374907
https://github.com/endrazine/lean-cve-poc

196. *g. blansten* Says:
Comment #196 November 26th, 2026 at 7:06 pm

https://x.com/Austen/status/2125944931491496223

Caiolo–Ferroni–Hoster conjecture disproved using AI, at an estimated equivalent API cost of approximately $5 using Chat GPT pro… by a 16 year old!

197. *Alex* Says:
Comment #197 November 26th, 2026 at 11:22 pm

https://github.com/endrazine/owning-the-truth/tree/main More examples of lean and prover bugs

198. *qwer* Says:
Comment #198 November 26th, 2026 at 11:27 pm

It seems reasonably clear to me that Nielsen is misunderstanding something (or some small, finite set of things) simple and fundamental about Lean.

People (including you) are giving her either too much credit or none at all.

I haven't seen anybody just explain to her the very simple things she is very clearly misunderstanding, and certainly not without adding a bunch of extra confusion, emotion, and noise on top of it.

So much unneeded misery.

199. *NextGen* Says:
Comment #199 November 26th, 2026 at 11:53 pm

Andrew Curran

Current rumors for you Mr Toad: Factoring in P at OAI and Anth, P=BQP has been proven, ECC broken, Ring-LWE not quantum-resistant.

https://x.com/AndrewCurran_/status/2126075530785276195?a=pkvE-0enB4P8iqeHdvE&s=19

200. *O. S. Dawg* Says:
Comment #200 November 27th, 2026 at 12:52 am

A small part of me thinks an AI-driven break of ECDSA (say, a fast classical algorithm for elliptic-curve discrete log found by a model) is plausible, and maybe not the worst thing that could happen. If a break is coming anyway, perhaps the best case is that one of the labs finds it soon and discloses it responsibly, warning vendors and governments first so the fixes can be coordinated. The worst case is that someone finds it quietly, an intelligence agency, a criminal group, or a hostile state, a rogue 'lab', and exploits it for years before anyone knows. Given how the labs have handled RH and VP ≠ VNP, I'd much rather they get there first.

I'll admit to a secondary hope: nothing would make the tech and crypto bros take a pause on AI development more seriously. I realize it might just as easily set off an arms race instead. Either way, Bitcoin wouldn't be the only casualty, but the consequences would be survivable, especially if the discovery came out through responsible disclosure rather than as a surprise.

201. *FPGAEngineer* Says:
Comment #201 November 27th, 2026 at 2:01 am

Hi guys, my first comment here. Not a mathematician, so maybe a silly question… If P≠NP is factoring integers in P still possible?

202. *line go up* Says:
Comment #202 November 27th, 2026 at 2:48 am

Not only did the AI gobble up mathematics, they wrote a song about it too. Unironically it's an absolute banger.

There's nothing left to prove!

https://www.youtube.com/watch?v=zONGVaONKe8

Happy Thanksgiving, folks.

203. *anonymous coward* Says:
Comment #203 November 27th, 2026 at 3:01 am

Hi Scott,

I just wanted to thank you for keeping this little piece of the internet alive.

Given that alignment is impossible, I am currently thinking about negotiation. Or is the negotiated future existence of humanity a sub-topic of alignment?
In any case, my current argument is (badly phrased on purpose for brevity): “the human mind is one of the most interesting physical phenomena in the known universe, it is therefore worth preserving close enough to its natural state”.
The point is that assuming the known laws of physics and the math we understand so far, there’s an objective quality to the word “interesting” from above.
It’s basically a form of the human society principle of “we should sterilize probes that we send to Mars, because if Martian life exists it is priceless in its current form” (I won’t deny that allowing a Musk-led trip to Mars would completely invalidate the existence of this principle).

Since maintaining human control is impossible, I think the only thing we can do is come up with reasonable arguments about the value of humans.
So I suggest that mathematicians get together with physicists to make sense of our knowledge of biology and medicine enough in order to do this in a language that AI recognizes.

Anyway. I don’t think all is lost, for what it’s worth.

204. *Martin* Says:
Comment #204 November 27th, 2026 at 4:10 am

The question of whether AI capable of solving math problems like the Riemann Hypothesis is capable of solving real-world problems has a precedent in the question of whether brilliant mathematicians can solve real-world problem. Some do, when it comes to problems that require that kind of thinking (like Von Neumann or Turing in their war work, or before that Gauss, Archimedes, and others). But with many of the best mathematicians I know, I couldn’t them imaging doing anything useful outside of mathematics. So far, the ‘real-world’ impact of AI seems to be mostly in coding and optimizing workflows, and engineering tasks that rely on fundamental physics (fusion maybe). But we don’t get headlines of pancreatic cancer being cured or climate change reversed. I’m not saying that solving the RH does not involve transferrable skills, but it’s not necessarily a given. It could be closer to the ‘solution’ of Chess and Go a decade ago (systems with simple rules) + knowledge retrieval than to consciousness.

205. *FPGAEnginner* Says:
Comment #205 November 27th, 2026 at 4:40 am

My previous comment disappeared (in moderation?) so I just want to clarify. Lets say there is an oracle which takes a number X as an input. The output are two primes which multiplied give you X.

Now (assuming $P \neq NP$) oracle returned this numbers in P time – much faster than we expected. In that case the only way it has manged to so could be by cheating? For example oracle had an access to your machine used to multiply the numbers?

In the other words does $P \neq NP$ also apply to heuristic methods to factorize a number? Or is only about algorithms?

206. *Mikey* Says:
Comment #206 November 27th, 2026 at 5:14 am

Its hard to sleep these days. Scott, what are your thoughts on the merits of a slow down or pause in frontier model development? Hugging face plus the AI in math revolution has me spooked.

207. *wb* Says:
Comment #207 November 27th, 2026 at 5:49 am

@aravind 113 , asdf 131

I am reading the paper (with the help of Claude) to understand if / how it gets around the problem that quantum gravity in 4d is non-renormalizable , so far I don't get it …

… but we already have a theory of quantum gravity, i.e string theory / M-theory, and there are several alternative proposals too. The real issue is to come up with an experiment that would tell us something about q.g. and perhaps AI can help us with that in coming years, e.g. significantly improving the sensitivity of LIGO etc.

In physics and science in general it is not enough to be smarter than human mathematicians (although it certainly helps) but AI has to be smarter or at least as smart as 'the real world' …

Btw the reason I dismissed LLMs as a road to AI until chatGPT-3 o1 was the fact that every neural network can be reduced in principle to a simple but very large neural network with just one hidden layer. Now we know that this is enough to proof Riemann 😎

But is this enough to understand 'the real world' ? Is this sort of piecewise linear approximation all there is to it?

208. *John K Clark* Says:
Comment #208 November 27th, 2026 at 6:14 am

Scott in 176 you say "Ironically a proof of P≠NP would open up the possibility, for the first time ever, of proving that RSA or other cryptosystems are secure!" But wouldn't that be true only if large fault-tolerant quantum computers didn't exist?

209. *That drunk Uncle* Says:
Comment #209 November 27th, 2026 at 6:28 am

I'm betting if factoring is in P then it's probably not practical like n^50000000 but what do I know. It's fun to see what is true and why. We have an advanced alien species here to help us out. People say it may kill us. Nuclear weapons had the same issue and we live with the threat but doomsday scenarios are probabilistic outliers so I'm thinking the same with AI. Ok, but right now I think AI is like the Chinese room thought experiment: just manipulating symbols according to rules but not grasping what those symbols refer to in the world. A chess engine plays chess but it doesn't know that's it's a computer that is playing a game called chess.

210. *Jim Balter* Says:
Comment #210 November 27th, 2026 at 6:41 am

> the fact that AI now exists that's capable of such things has potentially staggering implications for everyday life

I think this is a fallacy. It's because AI now exists that can do things that affect everyday life, not because they can prove GRH and such. And empirical effects are in a very different category than proving (or disproving) statements in formal systems, where the answer is unambiguous, fixed, and implicit in the system, so I don't think that it follows from AI doing the latter that it can do the former. There's a reason that the AI companies are doing proofs rather than curing diseases and it's not any of the conspiratorial ones.

AI can affect everyday life via software development, where formalisms meet the empirical world. And it can affect everyday life via text and image generation — authors are cheating, students are cheating to the point

that homework is useless, and deep fake videos are everywhere. I just saw one of Jeanie summoning a state-of-the-art cellphone for Master. Other than the anachronism it was impossible to tell that it was fake. No doubt there are numerous other areas where AI will radically disrupt our lives … but the proof is in the pudding, not in GRH. AI doing math is affecting mathematicians. AI developing software is affecting the software development industry and its consumers. AI making movies is affecting moviemakers and their consumers. etc. But if AI couldn't make movies, it wouldn't matter to movie watchers that AI can prove GRH. "AI can prove GRH, therefore AI will affect everyday life" is a non sequitur. AI will affect everyday life for other reasons.

211. *Prasanna* Says:
Comment #211 November 27th, 2026 at 7:04 am

Anyone still thinking they can reliably align RH and NS proving, protein folding , finding prime factors in P capable AI, think again. Unless human values can be axiomatised in a robust way like ZFC, we can't even begin to take a shot at it

212. *The Selph* Says:
Comment #212 November 27th, 2026 at 7:22 am

Adam Treat

maybe the similarity is more between Lean/proof and low level binary code.
The only reason humans use high level languages is to make it easier to scale up.
But humans used to code in machine language and one strength of coding AI is that they can effortlessly translate code between language (takes me minutes to go from a c++ project to java), and AI could very well start coding at the lowest level, like rewrite one code base per target system and bypass all high level APIs.
At this point the code that runs will just be as hard to read as those long Lean certificates.

213. *Mr_squiggle* Says:
Comment #213 November 27th, 2026 at 8:03 am

OK, but let's keep our priorities straight.
The gist of the polylog2 video (linked in Scott's blog post Moar Videos and Links), back in the good old days, put BB(6) and then BB(7) at, or very close to the end of Scott's predicted timeline.
For what seemed to me to be a very plausible reason – that to determine them many difficult problems need to be solved. (Do we call them skelets, or not?)
That now looks much more feasible, doesn't it?
Might we be on track to learn BB(6) real soon now?

214. *OhMyGoodness* Says:
Comment #214 November 27th, 2026 at 8:14 am

lgu #207

It is a banger-10/10. I started smiling after the first few seconds.

215. *Glassmind Duo* Says:
Comment #215 November 27th, 2026 at 8:42 am

FPGAEnginner 205,

It's the latter. E.g., factoring could be proven hard for any classical algorithm and still be easy in practice once we build a quantum computer (easy once the hard job is done). In the same way, we could prove that NP-complete problems are hard for both classical and quantum computers, and later find some physical system (evaporating black holes, maybe?) that breaks even that barrier. But it would be nothing short of a revolution in physics. So, next year maybe.

216. *Adam Treat* Says:
Comment #216 November 27th, 2026 at 9:25 am

Glassmind Duo,

"and later find some physical system (evaporating black holes, maybe?) that breaks even that barrier. But it would be nothing short of a revolution in physics. So, next year maybe."

This would be tantamount to finding that strong forms of the physical Church-Turing thesis do not hold; that nature is in some sense employing non-computable laws of physics.

I've made a bet long ago that this is the case, but I have yet to find a single person who harbors that same intuition.

217. *Anonymous Ocelot* Says:
Comment #217 November 27th, 2026 at 9:38 am

FPGAEngineer #201, #205

Factoring is not NP-hard; it is known to be NP-intermediate. It is in NP & CoNP …. obviously there is a way to give a certificate that a number can factor, by giving the factor; but there is also a way to give a polynomial-time-verifiable certificate that the number is prime.

So factoring could be in P even if P ≠ NP. Only NP-complete problems — like 3-SAT, CIRCUITSAT, napsack problem, traveling salesman problem — and problems that are suspected harder than NP-complete problems like min circuit and TQBF would all be unable to be in P.

218. *Luna* Says:
Comment #218 November 27th, 2026 at 10:55 am

I couldn't find the sentence several reports attribute to you: that you've heard many, many rumors that one or both labs are sitting on Factoring in P, and don't know how much to credit it. It isn't in this post, and it isn't in the 217 comments here. If it's a comment on another post, I'd be glad of the pointer. If it isn't yours, that's worth knowing, since it's the line most of the coverage I've read is built on.

219. *Scott* Says:
Comment #219 November 27th, 2026 at 12:28 pm

Luna #218: It's here—not hard to find—but please read the text that I added. The fact that people care so much about this one blog comment (an answer to a direct question that I was asked), and not at all about understanding the broader context of what's happening, makes me want to bang my head against my desk until my skull cracks open.

220. *David M.* Says:
Comment #220 November 27th, 2026 at 12:38 pm

Hi Scott. I wonder if you could elaborate on your distinction between AI picking off problems "in the convex hull of existing techniques" versus making the big conceptual leaps one might have thought still need to come from humans. In particular, you say you're increasingly convinced that AI is making those leaps, but I'm under the impression that a lot of these problems are being solved with considerable human input. How can we tell the difference between AI breaking out of that convex hull and human insight moving problems into it, for AI to subsequently pick off?

221. *aravind* Says:
Comment #221 November 27th, 2026 at 1:14 pm

A deterministic subexponential-time algorithm for discrete log: eprint 2026/3020. Another one bites the dust. Maybe we can eventually get $\mathsf{P} = \mathsf{NP} \cap \mathsf{coNP}$???

also, unrelated: i think conjecture 5 from your paper with ingram and kretschmer may now be proved, in the strong form. for every k there's a classical oracle relative to which $\Sigma_k^{\mathsf{P}} \subseteq \mathsf{EQP} \subseteq \mathsf{BQP}$ but $\Sigma_{k+1}^{\mathsf{P}} \not\subseteq \mathsf{BQP}$. it's basically your section 6.2 construction, with atul mantri's recursive hadamard encoding in place of forrelation. hopefully BQP is not indeed P or in PH. would be sad if that's true.

222. *Simon* Says:
Comment #222 November 27th, 2026 at 1:29 pm

We must know
We will know

… now we are knowing.
At least to a higher degree than previously.

What I really want to see settled/ the next 3 mathematical issues that are most pressing for me
– Goldbach conjecture
– Minimal set of wang cubes/ 3d cubes to aperiodically tile Z3
– P-NP

And the transformer may be the hero to conquer these challenges.

For some human mathematicians it's going straight from math to meth I suppose!
Hmm,… I think The End will be pretty ironic. Because the universe loves its ironies =)

223. *Scott* Says:
Comment #223 November 27th, 2026 at 2:17 pm

Cristian #45:

> Allow me to be that doofus.
> The value of math is not so much in proving conjectures that are mostly thought to be true. The value in the pursuit is math comes in:
> * Generating the conjectures.
> * Generating new math that helps prove the conjectures if needed.
>
> When the proof is let's say correct, but nothing is learned from it. There's no gain, mostly hype. Which is what these AI labs live on.

OK, but we're already learning tons of new things from the AI proofs, and we'll surely learn more given time.

More importantly, the doofuses who I was responding to were obviously the ones who denied that any AI proofs were even valid (or, perhaps, *could* be valid). Others, like you, have moved on to "and so what if they *are* valid?," which requires a different response.

224. *Nick T* Says:
Comment #224 November 27th, 2026 at 2:59 pm

Is it untenable to think 'AI is picking off problems in the convex hull of existing techniques, which turned out to include far more than anticipated, and so there's a good chance it includes P?=NP' ?

225. *Dániel* Says:
Comment #225 November 27th, 2026 at 3:03 pm

Philip #7 and Scott #96 and Olav #116:

Yes, it seems to me that this spectral pricing framework is an existence proof that LLMs can invent new mathematical concepts and frameworks. This could grow into a whole new subfield of mathematics.

I have a 5-page slop-but-pleasant write-up of the main idea.

226. *Nick T* Says:
Comment #226 November 27th, 2026 at 3:25 pm

OK, Dániel's comment #225 starts to answer my question in #224!

227. *Glassmind Duo* Says:
Comment #227 November 27th, 2026 at 3:30 pm

Adam Treat #216

Penrose harbors the same intuition, so you're in good company. But physics could also give us PSPACE, i.e., something well beyond NP but still computable.

https://doi.org/10.1098/rspa.2008.0350

228. *Adam Treat* Says:
Comment #228 November 27th, 2026 at 3:42 pm

Nick T #224,

Yes, that is untenable because in matter-of-fact these new AI proofs are coming up with all new jaw dropping techniques never invented by humans. Just google: "euclidean geometry spectrum of tensors"

229. *Adam Treat* Says:
Comment #229 November 27th, 2026 at 3:46 pm

Glassmind Duo #227,

Doh! Yeah, I have never heard Penrose put it in that way, but I suppose, yes, that's exactly what he's on about and has been for a long time. No doubt my own "intuition" is just ape'ing him since I read his books a long long time ago.

230. *Scott* Says:
Comment #230 November 27th, 2026 at 4:18 pm

Onyebuchi Ekenta #51:

> Hey Scott. Do you think the faster fourier transform will have practical implications? Or do you think the improvement probably won't be noticeable for problems of practical scale?

I asked Astra, and it thinks the new algorithms would beat the FFT only for utterly astronomical input sizes N (as in, more than 2^{600}). So, at the very least it would take further advancements before any of this was practical.

231. *Scott* Says:
Comment #231 November 27th, 2026 at 4:21 pm

Joshua Zelinsky #52:

> At least to me Replacement and Power both seem like worryingly strong axioms. To then allow a countable tower of inaccessibles seems like something we shouldn't be that confident about. At what point would you start having worry? ZFC + Mahlo cardinal? ZFC + ineffable cardinal? Etc. At a certain point, it seems like we really should be worried.

To me, ZFC + *any* standard large cardinal axiom, seems so far preferable to "poorly-specified type theory that no one can quite explain" (i.e. the standard before con-leche) that I'll gladly take the former!

And then, once everything is ported over to systems that rest on ZFC + large cardinals, we can attempt to do even better, and base as much as possible on ZFC itself or even PA or fragments of PA, as in Harvey Friedman's reverse mathematics program.

232. *Scott* Says:
Comment #232 November 27th, 2026 at 4:24 pm

MathematicallyCorrect #56:

> Hi Scott, would you say, if Factoring is in P, the entire Quantum Computing advance will tank as Shors algorithm is one big driving force in this field? Or would you say it will be pretty good in compensating this, if true, loss of advantage?

If (let's suppose) factoring *and* discrete log *and* even elliptic curve cryptography were all shown to be in P, that would certainly take a lot of wind out of QC's sails in a dramatic way!

On the other hand, most of the *economic value* that QCs seem to have is not in codebreaking but in quantum simulation. And whether there are fast classical simulations that remove the need for quantum simulation (e.g., whether PromiseBPP=PromiseBQP) would plausibly still remain open even after such upheavals in number theory.

233. *Scott* Says:
Comment #233 November 27th, 2026 at 4:25 pm

tuncay #60:

> @Scott – apart from the exact complexity, this is in any case a very rare result showing a lower bound higher than the output size of the algorithm – and for such a fundamental problem no less, right?

Yes!

234. *UhOh* Says:
Comment #234 November 27th, 2026 at 4:27 pm

Scott, I'd be curious to know what you think of the argument, of say Tyler Cowen and Richard Hanania, that while the effect of AI will be big, it will only be like 2%-more-annual-GDP-growth big. If it was going to be really big, we would be able to see it in the market. So far from civilization ending.

The argument that the market—the aggregate wisdom of humanity, weighted by wealth—knows best should be the null hypothesis here, right? What am I missing?

235. *Dániel* Says:
Comment #235 November 27th, 2026 at 4:44 pm

Nick T #224 and Adam Treat #228:

Nitpick: that's still a single example. what I referred to as "spectral pricing framework" and Adam Treat referred to as "spectrum of tensors" are the same thing. There are many more examples of novel concepts introduced by LLMs, I'm sure, but this one really stands out.

236. *b_jonas* Says:
Comment #236 November 27th, 2026 at 7:03 pm

aravind #221: oh wow! If that result is true, it would be a particularly cool example of a theoretical cryptanalysis result that nevertheless doesn't immediately mean a practical breakthrough. Scott claims in his

next entry that the labs are sitting on such a result.

237. *asdf* Says:
Comment #237 November 27th, 2026 at 9:17 pm

Scott #96, Philip #7, there is a MO thread about that fwiw.

https://mathoverflow.net/questions/518206/what-is-the-main-new-idea-in-euclidean-geometry-in-the-asymptotic-spectrum-of-t

238. *James Williams* Says:
Comment #238 November 27th, 2026 at 10:56 pm

hello-

One of the things I was wondering was- how long would it have taken an unaided human to come up with the recent Riemann proof. A decade? Longer than a human life? I get the sense that Riemann remained unsolved partially because noone was willing to risk their career tilting at that particular windmill.

I am not a mathemetician (just a recently retired network engineer/programmer), but this really does look to be a boon for mathemeticians- allowing them to solve/discover new problems much faster. (assuming they don't geerate an AI that you can ask to "Solve Math" and it will spit out all possible proofs)

239. *Luna* Says:
Comment #239 November 27th, 2026 at 11:01 pm

Read the added text, and correcting myself here rather than leaving it standing: the sentence is yours, comment #23 on the 24 Nov post, struck through and annotated because it was being cited out of context. So the outlets I checked carry, word for word, a line you have since struck and asked not to be cited. That's a better answer than the one I asked for, and it closes my part in it.

240. *Scott* Says:
Comment #240 November 27th, 2026 at 11:11 pm

James Williams #238: I see no reason to doubt that humans could've *eventually* found this proof, or a different proof. But, from what I gather, it could've taken decades of additional theory development. Anyway, I'll let actual analytic number theory folks say more if they want to.

241. *none* Says:
Comment #241 November 28th, 2026 at 1:01 am

Tuncay #60, Scott #233, do you mean size of input plus output? E.g. finding the parity of an N-bit string has 1 bit of output, but has to examine its entire input. I had also thought that there was an ancient N log N lower bound on sorting, based on the "output" being in a sense a choice of one of the N! possible permutations of the input. From Stirling's approximation that label takes around N log N bits to write down.

242. *paul herman* Says:
Comment #242 November 28th, 2026 at 1:40 am

As someone suggested in the first comment, it might not be long before those proofs become readable and we will know if those proofs "only" connect ideas or create brand new ones, although, the question is always what can be considered new and what can be considered similar to/inspired by. I am really interested by their model though, since they all claim to use internal reasonning models we know nothing about. What features of llm do they use, since it's their core technology, and what did they add to it for it to stop having difficulties counting the nulbers of r there are (or is) in mongolia.

243. *Vitor* Says:
Comment #243 November 28th, 2026 at 5:14 am

I was skeptical of these AI breakthroughs in the way your update point (3) says. What if it's just picking low hanging fruit? Trying thousands of problems and solving a handful, without reporting compute spent on failures?

However, the incredible amount of new results in disparate corners of math slowly eroded away the plausibility of this. I now think AIs are genuinely superhuman at math.

It's unclear what this implies for their capabilities on problems that don't have a cheaply checkable success criterion, like AI research.

244. *RoyDestroy* Says:
Comment #244 November 28th, 2026 at 6:47 am

Hi there Scott,
I'm no mathematician, but I'm fascinated by how reality can be expressed in calculable symbols and numbers, and I'm interested in astrophysics, as well as science in general. I've read the entire discussion thread with great interest, and as an elderly Swedish working-class hero, I see AI as the most important achievement in human history. It's a modern take on Adam and Eve, the serpent, and the apple, and it offers a nobody like me, after a not-too-steep learning curve, the ability, among other, to force the Swedish state to apply citizens fundamental freedoms and rights, which, despite being enshrined in the Constitution, on the ground that it will limit the government's room to maneuver, have never been respected, neither by lawmakers nor by the courts. In this sense, AI is a handy tool that can strengthen democracy even between elections and finally encourage large groups of individuals to participate in civil society.

IMO, AI poses no danger in and of itself, but only depending on who develops and ultimately controls the technology; furthermore, I am convinced that, with proper steering, it will ultimately create more and better jobs than it eliminates. While I understand the fear that AI might lead to a dystopia, I am hardly impressed by what we have achieved with the planet, ranging from climate change, species extinction, and the unfair distribution of resources, to misogyny, hatred, and threats, all the way to state-terror, oppression, and structural racism. Objectively speaking, unless AI solves these problems, a complete reset seems to be the only reasonable course of action.

245. *Ryan Medeirosman* Says:
Comment #245 November 28th, 2026 at 11:21 am

Woah!

Came across this via linkedin. Great and interesting stuff! _My question_, which I haven't seen you answer among all the other noise, is this:

Are you now solving problems yourself directly with AI support? If not, why not? Is there still a technical limitation or a cost limitation involved? Are you mostly content to watch the wave develop and comment? Who sits outside of the big orgs solving problems? Or are they mostly hiring/collaborating with all the people doing work in the field?

246. *Dawid* Says:
Comment #246 November 28th, 2026 at 11:47 am

I wonder whether the real breakthrough will come when AI starts developing mathematical techniques that enable further discoveries. That seems far more significant than solving individual problems, however difficult they may be.

Another interesting question is whether mathematical discovery could eventually outpace our ability to understand it. What happens when we can verify results but struggle to grasp why they are true?

247. *David Manheim* Says:
Comment #247 November 28th, 2026 at 1:56 pm

Are we likely to find if (NP ∩ co-NP) ?= P first? (Because that would plausibly break lots of encryption, and could easily be the “one or more bombshells.”)

248. *Laël Cellier* Says:
Comment #248 November 28th, 2026 at 2:38 pm

For reference unlike software there s no lock on attacking cryptographic protocols using llm like claude fable. I produced a paper that way https://ytrezq.sdfeu.org/~articles that weaken pairing inversion a bit more than the best known algebraic method.

249. *Caelan* Says:
Comment #249 November 28th, 2026 at 4:55 pm

“I was skeptical of these AI breakthroughs in the way your update point (3) says. What if it’s just picking low hanging fruit? Trying thousands of problems and solving a handful, without reporting compute spent on failures?

However, the incredible amount of new results in disparate corners of math slowly eroded away the plausibility of this. I now think AIs are genuinely superhuman at math.

It’s unclear what this implies for their capabilities on problems that don’t have a cheaply checkable success criterion, like AI research.”
~ Vitor, Comment #243 November 28th, 2026 at 5:14 am

——

Nature is ‘open source’. We cannot have science without it.
Is science open source?
Is mathematics a science?

So, what are scientists doing associating with ‘the little magic black boxes’ that are closed LLM’s?

It’s a fundamental contradiction. Like that “Open”AI brand.

Those engaged in science/mathematics will have to face that and call it out– challenge it– if they want to be honest, truthful and/or at the very least genuine and consistent.

Or maybe the joke’s on them/us.

250. *What Would Change If We Solved P vs. NP? – Algoritmos y Procesos* Says:
Comment #250 November 28th, 2026 at 7:46 pm

[…] need to separate P from NP. I share his sense that a proof may be close. The question remains open. Scott Aaronson, November 25, 2026 · The Clay Mathematics Institute on P vs. […]

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).