Who Watches The Watchmen?

OCT 19, 2026

37 4 13 Share

Every story has a beginning, a middle, and an end. The difficulty in faithfully recounting a story detailing a rapid succession of events that individuals lived through, is that it is a tale told in a thousand ways: a collective dream that meshes together, blending facts with fiction, living in the never-never land of symbols and portents. And, as is the case with dreams, sometimes there is a monster at the end of it.

I.
Les Prophéties

My story begins on May 02, 2026.

While idly scrolling social media app X (formerly Twitter), I come across a tweet by Will Depue, previously an employee at the American superintelligence corporation OpenAI. The tweet says (verbatim) “we’re really living through a historic moment for mathematics. it’s pretty incredible to watch”. Reading the ensuing conversations with mathematician Daniel Litt in the replies, it dawns on me that the tweet alludes to the future impact of LLMs in the field of mathematics.

I must admit that at this point of time I had not really been following the developments happening in the space between mathematics and AI, despite being a mathematician myself, although occasional news about progress (eg, on Olympiad problems, various Erdős problems, …) would slip through the cracks and appear on my news feed. Nevertheless, I had certainly heard much about AI itself, primarily from X The Everything App, which happens to be the typical congregation spot for buzz surrounding all things AI. Much of these conversations are no doubt populated and amplified by individuals, like Depue himself, who are based in the San Francisco Bay Area, a place that is intimately tied to the genetics of the AI boom.

Much has been said about how this unending 24/7 AI-news cycle originating from the Bay Area has to do with Silicon Valley VC startup funds being intimately tied to the great financial pyramid scheme of the century, working tirelessly to ensure that the AI

capex breathes while starving every other aspect of the palpable economy and global financial ecosystem of oxygen, all at the cost of leaving world economy hanging in balance, reliant on speculative fortunes of the One Bet. What I had not realized, and could not dream of, was how my relatively quiet and monastic inner life as a mathematician would soon be dragged into the whirlpool that is "Project AI".

The discussions between Depue and Litt went back-and-forth. Litt firmly made the point that, even though solving problems is a significant part of what mathematicians do, the discipline of mathematics is strictly less directionally bound by the kind of precise questions that one can train AI systems on. On the other hand, Depue kept insisting, almost as if he's reaffirming his unshakable blind faith in the second coming of Christ in the form of a Machine God, that the same framework of RL that excelled at solving the Olympiad problems will (verbatim) "more slowly solve the rest [of mathematics]".

At this point, an unlikely character enters the conversation with a response to Litt's rebuttals: Jacob Tsimerman, who would eventually go on to win the Fields medal in a few months. He was already known within mathematical circles to be one of the possible contenders for the Fields medal this year. I'd like to pause here and let my readers know that at this point of time Tsimerman himself already knew that he had secured the Fields medal; this news was conveyed to him in January by the president of the IMU.

Continuing with the story: Tsimerman responds to Litt by construing the domain of mathematical investigations as one that, eventually, centres around solving specific problems anyway, once they are formulated precisely. Therefore, he argued, can one really say with certainty that the impact of AI in the field would not be so significant and severe?

On May 04, Tsimerman goes on to elaborate on his points by a lengthy tweet.

He makes the following claim (verbatim): "problem-solving is an immense, and pervasive part of modern mathematical research. Consequently, if human problem-solving disappears by virtue of the AIs becoming strictly and substantially better at it, then most of the time currently spent by modern mathematical researchers will have to be spent on an activity that is altogether pretty different. Whether such an activity is viable as a professional endeavour is something I am unsure of …" Overall, the tone

of this post seemed rather urgent. Tsimerman spoke like Nostradamus warning of an undeniable impending apocalypse that he had in his visions.

From my vantage, the discourse was rather confusing. Having grown up in the pedigree of Henri Poincaré and Bill Thurston, the claim presented by Depue and Tsimerman that large swaths of mathematics ultimately boils down to problem-solving seemed misplaced at best, and disingenuous at worst. Whether mathematics is problem-solving or not always seemed more a matter of taste than an intrinsic question about the nature of mathematics, to me. But more importantly: there were no major news of AI solving major mathematical problems of general interest at this point of time. So what on Earth were these people even talking about?

I remember my friend, who is also a mathematician, sent me the humorously worded message: "how can a serious mathematician say this "my entire main research field disappears if someone solves the *Zilber-Pink Conjecture* in its more general form"" quoting (verbatim) a line from Tsimerman's tweet.

II.
Start Of Something Beautiful

On May 20, two and a half weeks after the exchanges between Depue, Litt and Tsimerman, OpenAI announced that their internal models had disproved the "unit distance conjecture", posed by Paul Erdős in 1946, (rightly) described in their milestone research announcement as "one of the best-known questions in combinatorial geometry, easy to state and remarkably difficult to resolve".

At the time, it felt like our Nostradamus' predictions are coming true before our very eyes. I would eventually come to realize that these predictions were not proclamations of a prophet, nor paranoid ramblings of a delinquent, but something more sinister: a smokescreen— a seemingly plausible reframing of the essence of mathematics and its impact on the real world, something that our generally naive and narrow-minded intellectuals could have sterile debates over, while unnoticed behind the curtains certain entities could continue their project of large-scale institutional capture and economic disenfranchisation of vast sections of human society.

The graphic novel *Watchmen*, originally a 12-issue comic book series, was written in the 80s by the self-proclaimed English wizard Alan Moore and is now hailed as one of

the most influential comic book literature to date. The central theme of *Watchmen* is a meditation on the concept of power. The book is set in the backdrop of an alternate 80s America, where US won the Vietnam war, Richard Nixon is still the president, and the Cold War rages on. The ones responsible for this divergence in history are the protagonists of the novel, the Watchmen collective, entailing various anti-hero vigilante characters, each with their own conception of power and flawed visions of morality and that of greater good. The main plot of *Watchmen* follows several of these characters as they slowly uncover their comrade Ozymandias' plan to unite the world and avert the impending nuclear armageddon by sacrificing millions of lives and blaming it on a fake alien invasion. At its core, *Watchmen* is a story about the vices of the American society. The world of *Watchmen* is bleak and violent, seen through the overbearingly nihilistic lens of the masked vigilante Rorschach as he roams the crime-ridden streets and backalleys of a gothic-noir version of New York City. Throughout the novel the reader is catapulted across a jarring barrage of newspaper cut-ups, television media, journals, memoirs, stories-within-stories that emphasize a collapse of the information ecosystem not unlike the present world we live in. In face of these existential crises, the heroes of the novel retreat into sex, drugs, religion, delusion, anything to avoid reality. *Watchmen* posits that if a superintelligent omniscient being really did exist, the American government would no sooner reduce it to a weapon and a propaganda tool.

The research announcement by OpenAI on the unit distance conjecture is accompanied by a short teaser launch video. It is here that we are introduced to some of our proverbial Watchmen: visionary AI researchers at the frontier of Mathematics and AI, witnessing the "superintelligence" emerge at their terminal as out of nothing, suddenly realizing that they are the ones upon which a greater part of the responsibility is bestowed to chart out a course forward for humanity into the Great Unknown.

Cinematic music reminiscent of Hans Zimmer's compositions plays in the background. We see OpenAI researchers, previously career mathematicians, Mark Sellke and Mehtaab Sawhney, explaining the unit distance conjecture on an easel chalkboard; camera pans revealing quick flashes of interior design boasting Silicon Valley startup decor. Sellke and Sawhney both recall their initial disbelief upon seeing the model output; Sellke recounts having trouble sleeping upon witnessing the proof. Next, we are introduced to Ozymandias himself: then-head of the OpenAI math

research team, Sébastien Bubeck. Sébastien succinctly explains how AI succeeded over humans here: the proof-tree for this approach was a dark-forest, there were too many delicate decisions to make; by searching through all possible approaches, the relentless blind idiot god Yaldabaoth had found the right idea, like a lightning strike to the ground. He ends with the following remarkable comment (verbatim): "And so what this tells us, is that AI is able to make breakthroughs in science, and can enable breakthroughs not only in mathematics, but in engineering, in physics, in biology, in medicine … (pensively) This result is showing glimmers of a Golden Era for mathematics." At the beginning and the end of the video, Fields medalist Timothy Gowers also makes quick appearances, stating rather matter-of-factly, in his usual British demeanour, the importance of the result that has been proved by AI.

I am far from a combinatorist, but I had heard of the unit distance conjecture before, as it closely relates to the distinct distance problem by Erdős which has been both a source and part of much interesting mathematics. Notably, work due to Larry Guth and Nets Katz on the distinct distance problem used strikingly simple new ideas from polynomial methods and real algebraic geometry. While I did not find, admittedly only upon rather cursory reading, OpenAI's internal model's proof of the unit distance conjecture nearly as gripping as the work of Guth-Katz on the distinct distance problem, it was clear that proof had some decently interesting ideas extending Erdős' original number-theoretic constructions. Furthermore, the experts seemed pretty much unanimous that this was an exciting development. The paper produced by the model was decently written, and it was accompanied by a paper co-authored by several experts summarizing the ideas of the proof very clearly, together with additional commentary by each of the experts — Tsimerman's comments began: "This is a really impressive piece of work, and I would accept it for any journal without hesitation". Bubeck's prediction of the future seemed on the horizon: a Golden Era of mathematics, fuelled by a synergistic collaboration between humans and AI, is now upon us.

Everything flows. All nice things come to an end. The unit distance conjecture would mark the last time OpenAI collaborates with the broader mathematical community in good faith. Since the initial highs, the story of AI corporations' obsession with mathematical theorems has increasingly felt like coming down an up-staircase… and the encounter at the end was crushing.

III.
The Men Who Sold The World

In the weeks following, the discourse on X seemed to return time and again to the topic of mathematics. Comments of the form "Can AI run out of math problems to solve?" were abound. Everyone seemed to be engaging with a performative caricature of mathematics through the lens of AI. At this time, academics around the world had started slowly realizing that GPT-5.5 is able to, once in a while, generate nontrivial mathematics upon prompting sufficiently many requests. It seemed this news had percolated to X. Every article on arXiv with a morsel of contribution from AI would get posted, screengrabbed and circulated across X, as if heralding the arrival of a yet unseen superintelligence.

The next breakthrough came on July 19th from an innocuous tweet by Anthropic employee, previously a career mathematician, Levent Alpöge. The tweet said (verbatim): "hello there the jacobian conjecture is false thanx to my close friend akhil for asking about it and my other close friend fable for working during the world cup final", followed by a string of symbols giving a three-component polynomial map in three variables that disproves the Jacobian conjecture in dimension 3, a problem that dates back to 1939. The formula could be checked on any old computer algebra system, and upon doing so one would soon conclude that it is indeed a valid counterexample.

I remember the tweet appearing on my feed when it had less than 50 likes. I was not familiar with Alpöge prior to this, so I had no idea if the author was reliable, and wondered for a while if it was a prank of sorts. Slowly, everyone including me came to the realization that it is indeed correct. The news spread like wildfire among the netizens of X. My initial reaction upon verifying the solution was… disappointment. I have always thought that an answer to a mathematical problem that is resolved by a collection of symbols without real insight is rather boring, regardless of how old and prestigious the problem is. At the time, I also thought that this was consensus opinion among mathematicians; I now realize that this was a misjudgement on my part. The original tweet by Alpöge came with no provenance on how the answer was reached, what prompts and harnesses were used; a pattern that would repeat over and over in the coming weeks.

Regardless, it was soon found by someone, again using AI, and then confirmed in a follow-up tweet by Alpöge, that there is a much more conceptual (indeed, geometric)

way of thinking about the example. The conceptual explanation nevertheless left something crucial to be desired, an algebraic coincidence that had seemingly no easy intuitive explanation. To my knowledge, no document detailing these points regarding the solution to the Jacobian conjecture has been released by Alpöge so far, despite promising the mathematical community (verbatim) "there'll be some pdf eventually!". After all, let's face it: is it really important how the solution came to be, what threads of human mathematical ideas, what corners of the castle of knowledge and understanding that we have painstakingly built brick-by-brick for centuries, that Anthropic's internal model "Fable 5" trawled through, to get to the solution to this conjecture? No, obviously not… the easy-to-state and truthful headline is machines solved an old prestigious conjecture that humans did not. What more could you possibly want to know?

The resolution to the Jacobian conjecture, studied by countless mathematicians (Zariski, Miyanishi, Gurjar, Abhyankar, Moh, …) lies gathering dust in a 280 character tweet with 43K likes, authored by Anthropic employee Levent Alpöge on the 19th of July, 2026. Much has happened since, and who has time to dwell in the past?

Friday 27th August 1999, Sheraton Moriah Hotel, Dead Sea, Israel. A discussion on geometry is being held, moderated by Misha Gromov, as part of the "Visions of Mathematics: Towards 2000" conference hosted by the Tel-Aviv University. I highlight a conversation from near the end of the discussion:

> P. SARNAK: I think one point that is coming up here, and it came up in your [Gromov's] discussion, is, and I think this is going to be very relevant in the future, is the role of computers in proofs and in mathematics, and it doesn't seem to have been discussed. It's inevitable that computers will appear in mathematics, I mean, computers in proofs, not in science, not in numerical analysis, the role of computers in pure math, in something like the proof of the four-color theorem or the proof of the sphere packing problem.
>
> M. GROMOV: I have a more radical opinion. I believe, that may be in 50 years mathematicians will be less relevant than computers.
>
> V. MILMAN: You see, Peter [Sarnak], if the Riemannian hypothesis would be solved with the help of computers, I would consider this to be a great achievement, because it doesn't matter how it will be solved. When you talk about the four-color

> problem, the only interest, which exists in such a problem, is that if this question is so simple and we don't understand how to answer it, then I would expect that some very young guy will find a missing structure here, and this structure will be more important for mathematics than the four-color problem. And when you put in ten thousand hours of computer time just to tell me that I will be able to find the right colors, indeed I don't care. If some mathematics was developed by it, that is the only thing I care about. [COLLECTIVE LAUGHTER]

Peter Sarnak is the doctoral advisor of Jacob Tsimerman, and is thanked in Levent Alpöge's doctoral thesis for years of mentorship. It is inspiriting to see the pupils venturing out and fulfilling the teacher's dream: a future where computers have an undeniably significant role in novel mathematical proofs. Nevertheless, one cannot help but stop and wonder: what if the dream was really a nightmare?

IV.
Portrait Of Two Mathematicians

On July 23rd, four days after the resolution of the Jacobian conjecture, at the Fields medal ceremony, Jacob Tsimerman accepts the Fields medal, wearing a wide smile and a custom-tailored silver tuxedo. Later that day, during a press conference, he announces that he would be joining OpenAI's AI safety division. He says (verbatim): "I think the world is changing … The math profession as we know it now … I don't think it will exist the way it exists right now."

It would be an understatement to construe Tsimerman's alignment towards AI research (and indeed, towards OpenAI in particular) as one more milestone in the vast project of institutional capture by AI corporations that has been slowly but surely taking place over the last few years. Four AI companies, Open AI, Anthropic, Google Deepmind and Meta has been consistently poaching academics across disciplines. They recruited more than 80 professors from top American universities in recent years according to some dated statistics. The numbers have probably grown steadily since. In 2024, the Nobel prize for physics and chemistry were both given in relation to AI research. Math has not suffered this gloomy a fate: the 2026 Fields medal was awarded to Tsimerman firmly due to his human achievements in number theory. However, his stated decision to take leave from University of Toronto to move to AI research was obviously one that was a long time coming, possibly (and this is speculation on my part) already finalized in January 2026 when he privately received confirmation to have

secured the Fields medal. Indeed, exactly a year ago, in 2025, Tsimerman had co-authored an article titled "A Taxonomy of Omnicidal Futures Involving Artificial Intelligence", where he lists various AI doomsday science-fiction scenarios. In an essay, mathematician Max Weinreich states (verbatim) "In the last two years, [Tsimerman] only took graduate students interested in AI; then, he no longer took graduate students; then, he left number theory to work on AI research."

In a Quanta Magazine interview, Tsimerman says (verbatim) "In my experience, and I think many people's experience, math is very much a goal-oriented endeavor … the truth-and-beauty stuff is there when you're zoomed out, but when you're zoomed in, you just want to win." The article goes on to say that, after he performed poorly in the Canadian Mathematical Olympiad at age 12, he'd spend five hours every day practicing Olympiad problems. Subsequently, he achieved a perfect score in the International Mathematical Olympiad in 2004. After that, he stopped competing because he'd achieved a perfect score and participating again could only damage his record. In 2024, when he realized he could be a serious contender for the final eligibility cycle of Fields medal, his childhood ambitions of the Olympiad days returned; he decided to pursue it single-mindedly: "I worked hard and sacrificed enjoyment to try and win it." By his own admission, he did this by deliberately reorganizing his research area to work on more "impactful" problems, solutions to which could secure him the Fields medal.

On October 16, Tsimerman announced he would be working at the Mathematical AI Safety Institute (MAISI) full-time as its founding scientific director, instead of joining OpenAI, although nevertheless they would still be collaborating with them. He also announced that MAISI will be supported by OpenAI. After all, he who sells the poison also sells you the antidote. It is becoming increasingly clear that AI safety is being used as a smokescreen by AI corporations themselves to keep the intelligentsia busy with trivia while erasing the very real concerns about IP, agency and product liability from public consciousness. On September 24, Edward Snowden, while speaking at the Ethics Forum at ETH Zurich, argued that AI corporations must bear liability for the harms caused by their products just like a weapons manufacturer or a drug manufacturer does for their products. He concluded (verbatim): "I think we need to put Sam Altman in jail, and I think it would be an instructive experience." Indeed, if AI safety is of paramount concern in the near future, the only way to effectively deal with it is by holding the orchestrators behind these dangerous products accountable, not by being a snake oil salesman.

Tuesday, 16th August, 1983, Warsaw. An awkward looking man wearing a plaid shirt and stretch denim jeans receives the Fields medal from Władysław Orlicz at the opening ceremony of the International Congress of Mathematicians. The name of the man is Bill Thurston. Thurston's 1992 article "On proof and progress in mathematics" would come to be one of the most prescient essays on the ecology of mathematics, for a time around three decades hence, to where we are at: the here and now. In his essay Thurston recounts the ecological affects of early success in his career (verbatim): "At that time, foliations had become a big center of attention among geometric topologists, dynamical systems people, and differential geometers. I fairly rapidly proved some dramatic theorems … It was hard to find the time to write to keep up with what I could prove, and I built up a backlog. An interesting phenomenon occurred. Within a couple of years, a dramatic evacuation of the field started to take place … People told me (not as a complaint, but as a compliment) that I was killing the field. Graduate students stopped studying foliations, and fairly soon, I turned to other interests as well." Thurston gives two explanations for this phenomenon:

> First, the results I proved (as well as some important results of other people) were documented in a conventional, formidable mathematician's style. They depended heavily on readers who shared certain background and certain insights. The theory of foliations was a young, opportunistic subfield, and the background was not standardized. I did not hesitate to draw on any of the mathematics I had learned from others. The papers I wrote did not (and could not) spend much time explaining the background culture. They documented top-level reasoning and conclusions that I often had achieved after much reflection and effort. I also threw out prize cryptic tidbits of insight, such as "the Godbillon-Vey invariant measures the helical wobble of a foliation", that remained mysterious to most mathematicans who read them. This created a high entry barrier: I think many graduate students and mathematicians were discouraged that it was hard to learn and understand the proofs of key theorems.
>
> Second is the issue of what is in it for other people in the subfield. When I started working on foliations, I had the conception that what people wanted was to know the answers. I thought that what they sought was a collection of powerful proven theorems that might be applied to answer further mathematical questions. But that's only one part of the story. More than the knowledge, people want *personal*

> *understanding*. And in our credit-driven system, they also want and need *theorem-credits*.

This was a man who had not only made breathtakingly beautiful new discoveries in geometry and topology, but spent quite a lot of time understanding and caring for the ecology of mathematics. Later in his mathematical career, Thurston would overcorrect for his perceived mistakes: he never published his 1986 preprint, "Hyperbolic Structures on 3-manifolds, II" containing his "Fields medal result": atoroidal Haken manifolds are hyperbolic. In 1998 he would put it up on arXiv as an e-print. The arXiv abstract says (verbatim) "this eprint is based on a 1986 preprint, which was refereed and accepted for publication, but which I neglected to correct and return. The referee's corrections have now been incorporated, but it is largely the same as the 1986 version (which was a significant revision of a 1981 version)."

But this is not to say he did not engage with the mathematical community. He did not simply dump his manuscript in a pre-internet era GitHub and vanish from the face of the Earth. Quite the opposite!

By early 1980s, Thurston's ideas regarding geometrization were already well known in the pertinent mathematical circles. In the mid-1970s, Thurston had given several talks explicating his ideas on classification of surface diffeomorphisms, which is a central piece of the puzzle in his proof. He gave lectures on this at Orsay in 1976-77, which would later be compiled in the Séminaire Orsay lecture notes "Travaux de Thurston sur les surfaces". Jean-Pierre Otal recounts in "Thurston's Hyperbolization of Haken Manifolds" that Thurston announced his hyperbolization theorem in Spring 1977 in the middle of his lectures on hyperbolic 3-manifolds. During 1978-80 Thurston began teaching a graduate course at Princeton, and in doing so he started preparing the lecture notes that would become "The Geometry and Topology of Three-Manifolds". In a 1994 autobiographical essay, he recalled that requests for these notes became so numerous that his mailing list reached approximately 1200 mathematicians, to whom he sent instalments every couple of months. Seminars around the world starting working through these notes, often spending weeks understanding a section. On February 1980, Dennis Sullivan gave a Séminaire Bourbaki lecture outlining Thurston's hyperbolization for 3-manifolds fibering over the circle. This would be the first written account of the theorem. Thurston's own manuscript of the proof would be prepared a

year later in 1981, where he would acknowledge Sullivan for persistently scrutinizing and improving on his unwritten proof.

Thus, due to Thurston's efforts in communication and engagement with the mathematical community far and wide, his ideas would disseminate much faster than a mere manuscript of the proof ever could, and breathed life into the field that otherwise would not have happened so swiftly and organically.

Recently, in an interview with Curt Jaimungal, Tsimerman proclaims (verbatim) "I think pretty soon, we are going to have proofs no one understands, and they're going to underlie mathematics, and based on that mathematics, other things are going to work. Does a train run even if no one understands the mechanism by which it runs? Yes it does, I mean, you can look at it run. Already now, a lot of my theorems that I have proven, I don't understand all the steps to …". Like many things Tsimerman says publicly, this is an appeal to normalize engineering epistemology in mathematics, and solidify the narrative that AI is going to be far better at mathematics compared to human mathematicians. If the consensus definition of mathematics is radically altered, this would of course be tautologically true. I'm fairly certain that OpenAI would be quite happy to mould such a vision of the future into reality.

Introduction of AI in mathematics as a vast industrialization project stands in exact contrast to Thurston's sensibilities. Thurston was an environmentalist in spirit. To him, mathematics was indistinguishable from mathematical ecology. He had understood that mowing down flora and fauna of the earth to build theorem-producing factories in their place is an affront to mathematical life itself. In my view, we would do well to heed his call to cultivate mathematics rather than strip-mine the earth for theorems. But I fear it may already be too late. If fate frowns, we all perish.

V.
It's About Sending A Message

00:34 AM Pacific Time, Saturday, August 01. OpenAI announces 10 advances in mathematics and computer science, as a launching ground for their next generation model "Astra".

If you found "00:34 AM on a Saturday" to be an odd time and date to inform the world about your next major product planned for launch, you likely possess formidable

business acumen. The mystery is unraveled a little bit upon reading Sébastien Bubeck's tweet (verbatim), "yes, nonsofic groups exist: this statement is one of the many new beautiful results proved by Astra, our next major model". Obviously, more than one person was working late at night that day; of course, to be clear, I say "late" because that's what midnight is according to their tight sleeping schedules, carefully logged in their Apple watches, meticulously following health protocols put forth by Silicon Valley techno-longevity gurus.

What happened was this: approximately 12 hours prior, an anonymous account on X had released leaked screenshots from a preliminary draft of one of the results proved by OpenAI's model: the existence of nonsofic groups. Independent rumors, credible or otherwise, began circulating that Anthropic employee Levent Alpöge had somehow also obtained the 10 problems list and was prompting Anthropic's internal model Fable to beat OpenAI to the punch. Amidst the chaos, OpenAI had to finish writing up the paper and reasoning walkthroughs as fast as possible and put it up on the website. The shoddy job is rather transparent (and will become even more transparent soon) from the "teaser trailer" associated to the research announcement page, which is half a minute long, animated hovercards only showing Problems 1, 7, 8, 9 and 10 with accompanying pretty pictures. For context, the existence of nonsofic groups is Problem 3, and was by far the best mathematical result proved by OpenAI's models so far.

This is where I enter my own story.

The original announcement article by OpenAI on the existence of non-sofic groups claimed that there had been (verbatim) "no progress on the main result for at least a decade". Mathematician Francesco Fournier-Facio released an article on August 03 significantly shortening the construction, and pointing out that this claim by OpenAI was blatantly false, as their own work critically relied on the results from a preprint by Gabor Kun, and a joint preprint by Andreas Thom and Gabor Kun, respectively. OpenAI has since retracted this claim.

On August 04, I decided to run these preprints by OpenAI's best publicly available model Sol, to check for errors. The reason was twofold: the preprints were unpublished, and it was conveyed to me by some human mathematicians that there were some tentative uncertainty as to their veracity (my understanding is this trepidation has since been completely quenched, as the authors and the mathematical

community has done a tremendous job quickly disseminating the ideas in these preprints). Sol came up with, what seemed to me at the time, to be a non-trivial (but as it turns out, non-critical) error in the statement of the main theorem in Gabor Kun's article. The theorem is a TFAE ("the following are equivalent") statement, entailing four statements. The proof goes (1) implies (2) implies (3) implies (4) implies (1). In the proof, there was an error in the step "(4) implies (1)": this implication could not possibly be true as written. The catch was, OpenAI's article quotes this main theorem as-is, without any footnote pertaining to the error. This immediately raised red flags in my head: the proofs of the 10 results by OpenAI are all accompanied by Lean (formal verification system in mathematics) verification certificates. Surely the system would have found the error? Besides, if a weaker model like Sol can find it autonomously without help from Lean, why can't Astra?

I wrote out the counterexample to "(4) implies (1)" provided by Sol, by myself, in detail, convinced myself it is correct, and then tweeted it out. The tweet got ~120K views and eventually was seen by interns and employees at OpenAI.

The first response in the form of a quote-tweet came from Cambridge University undergraduate student Kevin Barreto, going by the pseudonym AcerFur, who at the time was doing an internship with OpenAI. His response (verbatim), "Anyways, we were aware of these things and of course checked the whole citation tree of results before announcing such results", and then he proceeds to list two examples from the Kun and Kun-Thom papers that Lean had flagged, and how Astra had applied various "patches" to get around these problems. In further public exchanges, upon being pressed by me and the pseudonymous account doomslide, Barreto admits that in the prose proof there were footnotes for pointing issues, adding "But I don't want to speak about it further".

Something was amiss, and I suspected that OpenAI, having created a proof-generating machine, was treating mathematical proofs with the same fast-and-loose "move fast and break things" philosophy as Silicon Valley startup enterprises. As I am sure many recognized at the time, long term ramifications of this ideology taken to its natural limit is terrifyingly dystopian for the practice of mathematics (and, in future, inevitable other arenas of science and technology). Did anyone at OpenAI even understand the proof? Or did they let Astra do all the work, checked if the definitions are the "correct

ones" in the proof tree of the formalization (which can also be relegated to an AI agent), and just dished it out?

The suspicion that this is the case grew tenfold the next day, when OpenAI employee Mehtaab Sawhney posted a tweet saying (verbatim), "We have heard concerns about the proof of the existence of a non-sofic group, particularly its reliance on results of Kun and Kun-Thom. The original lean certificate is end-to-end formalizing every necessary ingredient from those papers … Such issues are commonplace in the mathematical literature and do not affect the results we cite."

I do not think there is anything "commonplace" about a group of people writing a paper using an AI model by typing a sequence of superficial prompts that amount to "please solve this problem", the output of which is then formalized (if that is even the correct order of how things operate), and claimed to be a complete work contributing to the broader mathematical literature, without *any* human in the loop understanding just what the hell is going on. And if they did understand, why didn't they simply put out a footnote explaining why these errors are minor and do not affect the proof? If anything, such explanatory remarks *are* commonplace in mathematics, and contributes positively to the collective understanding of mathematics! There is absolutely no way you can slice the situation that makes OpenAI's math research time look good here.

At this point, it should be obvious that OpenAI had already been framing a corporate narrative for the future of mathematics, one which has no need of humans. Of course, their construction is ludicrous, because mathematics is nothing if not a human endeavour.

VI.
Augustus Horribilis

Over the course of the rest of August, several AI breakthroughs in mathematics would be announced by Anthropic employee Levent Alpöge in increasingly nauseating tones of nonchalance. The tweets came with no insight into the provenance of these results, no detailed log of the prompting techniques and harnesses used and had very little to say about their relationship of the results with the existing mathematical literature. Eventually, Alpöge, who is officially a pre-training researcher in Anthropic, would go on to reveal that all but a handful of these were his personal projects, that he would tinker with Anthropic's internal model "Fable" on free extra tokens given to him on

weekends. Nevertheless, his blasé tweets would rake up hundreds of thousands of engagements on X and it would be unusual if that wasn't doing Anthropic's capabilities marketing a favour.

On August 10, Anthropic announced that they have improved a bound on the density of zeroes of the Riemann zeta function falling on the critical line. The announcement said that a high-school drop-out and former "Thiel Fellow" Jared Sumner, who, it was emphasized, had no substantial expertise in mathematics, persistently prompted Claude to "seriously attempt the Riemann hypothesis". From Sumner's own tweet (verbatim), "8 days ago, while jogging, I asked Claude to solve the Riemann Hypothesis. It didn't. 1.5 days later, it proved >= 67% of the zeros are on the line (prev: 41.6%). Still not sure what that means, but some analytic number theorists seem excited". As reported by Anthropic, his input largely consisted of motivational prompts in the general format of "keep going". Alpöge was one of the people involved in reviewing the mathematics.

The message communicated was clear: "frontier" mathematics is now a spectator sport. You can do it while watching the world cup, you can do it while jogging, you can do it while being hammered at a jazz bar. The days of experts, careful scholarship and mathematics as an act of meditation are over. No doubt the adderall-infused San Francisco techno-optimist hollow men are overjoyed by this news. It was an impeccable marketing strategy. I am sure Anthropic's rivals at OpenAI were watching.

In the days following, Alpöge posted several other solutions to long-standing open problems in mathematics in the form of cryptic viral tweets. Somewhere around that time, an unrelated human breakthrough came into the news: Simon Brendle and Pei-Ken Hung had solved the long-standing question of whether the product of two 2-spheres admits a positive sectional curvature metric, in the affirmative. Alpöge mentioned that he had been working on this with Fable, which produced some plausible-looking results in this direction. In a quote tweet, I suggested (verbatim), "If it does work, let me suggest an idea (which you may ignore). Start a public blog in your name, totally decoupled from Anthropic. Make this your first post; this result is already proved by Brendle-Hung, so it would not be that useful for marketing anyway. Invite expert opinion." Needless to say, my idea was ignored.

On August 23, Alpöge posted a tweet announcing a 108-page construction of a complex structure on the 6-sphere, which was known to be a famous open problem in

differential geometry. The paper, entirely generated by Fable, was accompanied by a 2-page neuralese summary of the construction, which appeared to also be LLM-generated. To date, I have not personally encountered a fully satisfying explanation of the construction, although eventually Philip Engel wrote a much more readable account of the construction. It took several days for me, even with LLM-use, to decode the human-unreadable paper, to convince myself, after drawing many pictures, that at least the 6-manifold constructed is indeed the 6-sphere.

Since then, I have understood a few more things about the construction that are left unsaid in the 108 page tome, and have been able to apply these to produce some limited number of examples of not-entirely-obvious complex structures on almost complex 6-manifolds. I suspect I have only barely scratched the surface in terms of what else could be done with the ideas in the construction, and in large part because no one really knows what the true thought process behind these ideas were.

The worst part is neither Alpöge nor the vast majority of mathematicians and non-mathematicians on X, who made a lot of exciting noises about it, really seems to care too much beyond the headline "AI can do stuff humans wanted to but couldn't". We have suddenly found ourselves at a possible cul-de-sac in the culture of science, certainly already gaining a slow but steady foothold in the public consciousness for years no thanks to a barrage of superficial popular science texts, wherein everybody wants to know but nobody wants to understand. Even if they do want to understand, the meaning of "understanding" seems to be heavily diluted— type such and such prompts to summarize and clean up the exposition and then upon a cursory or not so cursory reading, go "ah, yes, I get it now". This is understanding without putting any effort at finding something deeper or hidden behind the page of symbols and diagrams, a version of frictionless (thus, fast) mathematics. I can't help but suspect that this superficialization (or even commodification) of human intellectual work was exactly the deliberate ultimate goal of announcing these AI generated results on a fast-moving social media site.

Eventually, Levent Alpöge would go on to poach various domain-experts for personal collaborations on prestigious problems outside of his domain of expertise. He would reveal that on September 2025, he had emailed mathematician Tristan Buckmaster to collaborate with him on the Navier-Stokes equation. The outcome of this collaboration would set off a chain of events that would eventually lead to catastrophe for scientific

norms for professional mathematics. After getting, for all intents and purposes, scooped by OpenAI on the Navier-Stokes problem, he would lament the experience as (verbatim) “dream crushing”. Afterwards, he would go on to collaborate with Ethan Sussman on another Millennium Prize problem, namely the 4-dimensional Yang-Mills existence & mass gap conjecture, which is currently a work in progress.

At that time, it was not completely clear to me what exactly is gained by an ex career mathematician of Alpöge’s stature, in prompting LLMs on weekend projects to produce answers to mathematical questions that are too far outside his research area (number theory) for him to fully understand, which are then stored in a locked box with the keys thrown out, or released as hundreds of pages of neuralese mathematics that no one ultimately has any motivation to read, or laundered through domain-experts thinly veiled as academic work. It seemed to me a fundamentally anti-social act to announce these, especially as someone for whom the mathematical community was once home.

I think I may now have a better theory of mind for such behavior. To explain my point of view, I need you to indulge me in a brief digression.

VII.
Of Adam’s Curse

The port city of Duluth, Minnesota overlooking Lake Superior, with thousands of acres of green space and pristine wilderness alongside the lake and over 300 miles of diverse trails for hiking and exploration, is a popular summertime travel destination. The Duluth REU is a 10-week summer research program for undergraduates at the University of Minnesota Duluth, founded in 1977 by Joseph Gallian, and has since grown to have an outsized impact on the American combinatorics and number theory community. Its long list of distinguished alumnus includes Fields medalist Manjul Bhargava himself. Mehtaab Sawhney and Ashwin Sah were among the participants of the Duluth REU during the summers of 2017 and 2018, respectively. In both these years, Levent Alpöge, himself an alumni of Duluth REU from 2013, participated as a research advisor.

> “Time is a river made up of events which happen, and a violent stream; for as soon as a thing has been seen, it is carried away, and another comes in its place, and this

> will be carried away too."
> — Marcus Aurelius

Sah and Sawhney were both rising stars in combinatorics as undergraduates at MIT, where they met in late 2017. Their initial acquaintance would kickstart a long and productive collaboration. During 2017 to 2024, they had produced roughly five dozen papers across vast areas of combinatorics such as extremal graph theory, probabilistic methods, and Ramsey theory. They have collaborated on at least a dozen articles that were accepted at the most prestigious journals of mathematics: Annals, Inventiones, Duke, GAFA, JEMS. They both received the Morgan Prize in 2021 for outstanding research in mathematics as undergraduates, and finished their PhD at MIT in 2024. Afterwards, Sawhney obtained the prestigious Clay Research Fellowship and joined Columbia University as a tenure-track assistant professor. Sah left mathematics, one can only speculate as being disillusioned with academia, and joined D. E. Shaw & Co. as a quantitative analyst.

On the Autumn of 2025, Sawhney would be given access to GPT-5 by Mark Sellke, and it would slowly dawn on him that the system is getting increasingly better at solving important problems in his areas of interest. Eventually, he would join OpenAI on January 2026, to be a part of history: as a Biblical Magi witnessing the infant Christ.

Successful mathematicians usually have humble beginnings. But, sooner or later, they recognize the dark undercurrents of the economics of academia, in which the prestige currency dictates much of one's career trajectory, hiring prospects, journal acceptance, article readership, invitation to seminars and conferences, and so on. Intrinsically, prestige is an emergent behaviour in any creative medium: by and large, society values and look up to as inspiration, individuals who are more creative and original than others. And, as in any discipline, some individuals hone the meditative practice that allows one to obtain ever more creative ideas: the art of "catching the big fish" as American filmmaker David Lynch likes to put it.

The trouble is, once the currency is in place, society tends to reconfigure around the rules of economics, and individuals often start valuing the prestige hierarchy much more than the process by which these alien ideas germinate in the human mind. This is only a flaw in said individuals insofar as every human is afflicted by Adam's curse. A combination of early successes tend to exacerbate this psychological state, and sometimes one falls into a kind of cyclical need for more prestige, often at the cost of

enjoyment of the creative process itself. This is best highlighted by Jacob Tsimerman's own recounting of the time when he started working towards the Fields medal, wherein he meticulously shifted his research project to align with the kind of work that is valued by the prestige hierarchy, maximizing his chances for the Fields medal. Instances of such "perverse incentives" are abound in economic history.

I would like to posit the theory that many academics currently in AI research spaces are fundamentally misaligned by the prestige currency of academia, and the "AI misalignment" they wish to solve isn't a bug, but a feature of their very own creation, a mere reflection of their own selves projected back. During the time this essay is being written, in mid-October, frontier AI laboratories are working to establish that their latest batch of AI systems possess preliminary signs of "research taste" in mathematics. The way they plan to prove it is thus: OpenAI is going to say that by letting their internal model "Bel" run on a single problem in a particular mathematicians' research area, the model can autonomously find and exhaust all related problems in their research program.

The idea that one can strip-mine a collection of problems roughly aligned around a single theme to establish "research taste" should at its face be laughable to any reasonable person. But to an AI researcher living in their own bubble of collective myopia, bound by the very paradigm of goal-aligned research that they infuse and infect these models with by reinforcement learning, taken aback by the paradigm's effectiveness as gauged by various statistical measures of productivity, reaffirmed in their own flawed ideals of human creative and scientific processes— to them, this is unvarnished truth. At the end of the day, I would feel deeply sorry for them, had it not been that they were working with a handful of enormously powerful corporations backed by large institutional and financial structures, which can act on its own accord to establish new norms around aspects of human society with total impunity.

I like to imagine that Ashwin found a more fulfilling life away from academia.

VIII.
Look On My Works, Ye Mighty, And Despair!

By the end of August, boiling anger, concern and ultimately dismay within the mathematical community began bubbling up to the surface. A communal blog for mathematicians, Proofs & Prompts, was launched earlier in the month and slowly a

stream of essays started arriving. Likewise, multiple essays on AI & math began cropping up under the math.HO tag on arXiv; notable among them was Max Weinreich's essay presenting a case for total opposition to the use of AI in mathematics, one that would eventually become a seed for the formation of Association for Human Mathematics in early September. Mathematicians from across the world, with wildly varying worldviews and opinions (even regarding the use of AI and the future of math itself) started pouring in on X, united only insofar as they were a breath of fresh air compared to the opinions of the so called "techno-optimists"— a self-styled fashionable collective of hollow men, eternally proclaiming the same monotone plastic incantation of "more progress" in unison.

In late August, Hugo Duminil-Copin would write a Proofs & Prompts essay titled "Care for a little more AI?", where he makes the prediction that one of the milestone results in percolation theory is likely to be solved by AI soon, and how this may not be the cause for celebration an average person hopes it is. He wrote (verbatim): "In my opinion, the problem lies elsewhere: the current use of AI does not empower us, it petrifies us. These artificial discoveries risk decapitating entire fields before they have had time to develop to their full potential. Even worse, they nuke the mathematical landscape, making it increasingly difficult to inhabit after each blast." A few days later, a GitHub repository managed by Anthropic and committed by Anthropic researcher Justin Leder, would mysteriously appear out of nowhere in the AI terminals of anyone who had asked for it, with an AI generated proof of exactly the result Duminil-Copin had mentioned.

In early September, six Caltech undergraduates, four of whom had previously interned in tech venture capital startups like the YCombinator ecosystem, Sequoia and Neuralink, would go on to announce an AIxMath competition event "Mathathon" with the promotional slogan "40 hours, $2M+ AI credits, solve an open problem." The program was sponsored by OpenAI, Anthropic and several other corporations tied to AI research. Soon, a coalition of current and former Caltech mathematicians prepared an open letter vehemently opposing the event on September 10. Upon going live, it gathered thousands of signatories within hours.

The top was starting to blow off.

Somewhere amidst the ensuing chaos, Sébastien Bubeck reached out to me on August 26 through a mutual on X, asking me to follow him so he could DM me. And so, I did.

He messaged me in a few hours quoting a tweet I had posted. The tweet said "[OpenAI and Anthropic] are knowingly inflicting a ton of damage to the discipline of mathematics because it keeps the investor hype going", which garnered around 25K views. At the time, I had a less than 1K followers on my account. Bubeck wrote (verbatim) "If you're open to it I would like to understand your perspective better, because it is definitely not my/our intention to inflict damage to math, not in the slightest, and I'm very open to update and change course. let me know if you have time for a chat".

I admit I felt rather apprehensive after receiving this message, and couldn't at first really quite believe a representative from OpenAI, a corporation that has practically endless political and financial power, co-founded and managed in the capacity of a CEO by a weapons-grade psychopath, is personally reaching out to a random twitter account with <1K followers for policy suggestions in good faith. Especially because, prior to this, I had been writing faux-antagonistic posts tagging him: for instance, during the nonsofic groups fiasco, I publicly blamed him, being the math team head, for playing corporate product-shipment strategy games with mathematical proofs while writing the actual mathematics badly, though now I know Sawhney was more responsible for oversight on the mathematical end of things. Part of the reason for my antagonism is that it is the key to get people's attention on X, but I digress. Either way, I felt I should be very careful in choosing my words.

My initial trepidations were largely unfounded, and it turned out Sébastien indeed did reach out to me seeking my opinions in good faith. Nevertheless, I am glad I took my time and responded professionally several hours later after careful thought. I paste here verbatim my letter to him:

> Dear Sebastien,
>
> Thank you for reaching out.
>
> I think the project your company has embarked on is fundamentally antithetical to mathematics as a discipline. From private communications with a handful of high-profile mathematicians, I have heard that they were extremely angry at the way the 10 problems were marketed (as a public example, see the footnotes in Fournier-Facio's writeup, his interview, and try to read between the lines), and the fact that there were discrepancies between the Lean formalization and prose proofs and that

the prose proofs were extremely hard to read. I do not expect these to slow you down because I understand this is part of the marketing gimmick of OAI.

However, if you want some suggestions:

- On the positive end I would recommend presenting these "proofs" really as "findings", as an astronomer does by looking at the Hubble telescope, without a confident claim to correctness and "settling an old problem". Most important you can emphasize that maybe there are ways to generalize these findings to discover not isolated examples but conceptual phenomena, which is the ultimate goal of mathematics in my view. This way you can take the mathematical community along as you speed headfirst into the nebulous "progress" into human science that your company envisions, instead of leaving behind humans and gunning for a future where none of these theorems have any value because no one is looking at them. It might help to read, and re-read, Thurston's "On proof and progress in mathematics" over and over before playing the slot machine with infinite compute on these new technologies of the future.

- However as a cynical alternative you can just keep dumping your findings at an even faster rate and larger volume without any heed to what I expect is 50% of the mathematical community that are opposing this new culture, without looking for the detractor's approval, completely crush them all at once, dissolve and declare academic mathematics as an exercise in antiquity so that we (the 50%) can get out faster from this furnace and just become a religion; because no matter what, even if superintellience arrives and has the capacity to solve all of math, some of us will do math "the old way". We just want to do mathematics instead of being endlessly trapped in this artificial discourse.

Anyway these are all my opinion and it's a common sentiment among people I talk to but I am under no pretense of being a representative of some broader math research community. I hope these are of some use. I'm a little baffled by having to be put in the position to suggest policy for OAI given my obvious and open distaste for the corporation, and its influence in my field, and more importantly how I'm quite literally a nobody. But I have nevertheless tried to collect some thoughts and I present them to you as such.

Sébastien responded to me saying that the reason they did a few Erdos problems earlier in the year, then unit distance problem and then the 10 Astra problems is truly to update the mathematical community that these capabilities are real, and notify the community ahead of schedule so they can reconfigure institutional structures. He reassured me (verbatim) "We will NOT continue to release proofs. That would be distasteful and against the principles that we want these tools to empower existing communities. we might do one here and there, if it's like interesting in some ways, but you can rest reassured that we will NEVER do a play on "volume" and try to flood the zone with a ton of proofs that nobody cares about. That would kill the field."

A few hours later, lead reasoning researcher at OpenAI, Noam Brown, reaffirmed this in a public tweet. He said (verbatim), "At OpenAI, we aim to devote time to finding and announcing math results from internal models only when they would meaningfully change people's understanding of the pace of AI progress. Our main focus is shipping great models so everyone can use them to make discoveries of their own." In the replies, someone asked if they'll take a stab at a Millennium Prize problem. He responded "We already did, it didn't solve any of them. But we haven't thrown a ton of inference at them.", foreshadowing events that would transpire about a week later.

On August 30, Sébastien would privately message me by quoting one out of a series of tweets I had made where I revealed (an information I had obtained from two separate reliable sources) that the OpenAI math team was internally torn about what to do with the vast collection of results their models had proven during benchmarking sessions. I had made the point that Ronen Eldan, James Lee and Sébastien Bubeck all had differing opinions, but how it does not really matter, because the corporation they are employed at structurally only optimizes for profit, even if professional mathematics stands to be collateral. Sébastien told me (verbatim), "Just want you to know, there is no money to be made in math, that's really not the goal. also we are giving away 100k seats for free! (Btw if you want one I can give you a pro subscription)". Of course, I declined his generous offer. A short conversation ensued where I underlined my point that it doesn't truly matter what the goal is, for the final outcome to be devastating. I also expressed some limited skepticism about the idea that "there is no money to be made in math". The conversation ended on an amicable note.

At the time, I wondered if Sébastien's offer to give me a free pro subscription was a soft bribe of sorts because my posts were getting too much attention among the X commentariat. Now I see that was just paranoia on my part: the offer was well-meaning, but it comes from the same sociopathic place as OpenAI's general intention to do "good for mathematics" or more generally "good for humanity" while practically, everything they touch turn to ash. Yet, somehow, their employees continue to buy into this altruistic idea not realizing that it's not their intentions but the core values of the corporation they work for, which is aligned towards a project that values only Capital, that dictate the outcome of their actions. An outcome that, at the end of the day, tantamount to widespread disenfranchisement and alienation of individuals.

On October 12, after hanging their dirty laundry in public by turning the solution to the Navier-Stokes problem into a public spectacle of corporate competition, after releasing a batch of poorly-vetted, not even all formalized, AI-generated solutions to 400 open problems in mathematics that annihilated several research programs, after paralyzing countless young mathematicians in name of "progress", after all of this—OpenAI researcher Noam Brown had found in himself the gall to respond, pushing back against a tweet by David Bessis (correctly, in my opinion) claiming that OpenAI does not care about the future of mathematics. His response was (verbatim): "… Many of the OpenAI researchers behind these results are math PhDs or former math professors. If anything, they care too much about the future of mathematics. AI is disruptive but we also think that it will be an invaluable tool for the future of mathematics."

> ADRIAN VEIDT: The Comedian was right. Humanity's savage nature will inevitably lead to global annihilation. So in order to save this planet, I have to trick it… with the greatest practical joke in human history.
>
> DAN DREIBERG: Killing millions?
>
> ADRIAN VEIDT: To save billions. A necessary crime.

IX.
Do Androids Dream Of The Clay Millennium Prize?

In early September, various unsubstantiated rumors began circulating on X that Anthropic had solved "at least two" Millennium Prize problems.

On September 08, mathematician Tristan Buckmaster posted a tweet on Mastodon containing four documents: three of them contained finite-time blowup solutions to three separate PDE's related to fluid dynamics. The fourth document was a statement released on behalf of himself and Levent Alpöge. In the statement, Buckmaster discloses that they were working on these problems along with finite-time blowup for the Navier-Stokes equation, for the past year. Along the way, they had made use of several LLMs. The statement then goes on to apologize for what it's about to say next, and how the headline really should have been a call for the mathematical academia to wake up to the capabilities of LLMs, how their findings are indicative of a "DeepBlue-Kasparov" moment in mathematics, and laments for having external reasons force his hands into changing that intended message.

In the rest of the statement, Buckmaster tells an astounding story. He recalls reaching out to an *unnamed prominent mathematician* at OpenAI to clarify the provenance of his works with Alpöge, which he claims were in personal capacity without any institutional involvements, and to defuse the corporate tension originating from Millennium Problem rumors, which had spilled quite a bit far from social media and into the real world. On September 06, Buckmaster met with this unnamed mathematician and Sébastien Bubeck on a call, in which he was informed that internal OpenAI model had produced a 100+ page proof of finite time blowup for the forced Navier-Stokes equations, settling statements C and D of the famed Clay Prize problem, with "very little human input".

Buckmaster had, then and consistently since, expressed skepticism regarding the "very little human input" part. According to him further chats with Sébastien revealed that there was a team of humans working on this problem and an enormous amount of compute had been used (compare Noam Brown's statement a week ago "we haven't thrown a ton of inference at it"). Buckmaster would go on to make insinuations that OpenAI's breakthrough was possible only after them hearing that Buckmaster was involved in the project with Alpöge. From his recent Numberphile interview (verbatim): "Eventually I said… I sort of did a loaded question to Sébastien, I said "So you agree that you only started working on it after the rumors?" and he said yes." He would also go on to say that OpenAI's unforced Euler equation paper had all the mechanisms and architectural ingredients identical to a different version of the Euler equation paper he and Alpöge produced but did not release yet.

In my opinion, Buckmaster's skepticism is well-placed: I have heard from separate reliable sources that there was a fairly large team, consisting of NDA'd individuals who we may never hear of, working tirelessly on this, often working overnight at the OpenAI offices. Buckmaster goes on to say (verbatim):

> I asked whether the model had been trained on, or had access to, our sessions in Codex, into which we had been putting all our drafts for the whole of this project. I was told the model did not look up user data. I asked again, about training, and I did not get an answer.
>
> Two proposals were offered to me. The first was that we post our Euler result, and that OpenAI post its Navier-Stokes result the next day. The second was that, after posting Euler, I alone write a paper presenting the Navier-Stokes result, acknowledging that an internal OpenAI model had resolved it. Sebastien twice asserted that he wanted Levent removed from authorship, and said it would all be simple if only it were not the case that, and it was so annoying that, Levent works at Anthropic. It was also said that if OpenAI posted after us, they would say that we deserved the Clay Prize, and that we were the "closest humans to the problem". I declined both offers.
>
> I said that if OpenAI released its result in the way proposed I would go public with what happened. The reply was, "Why would you ruin your career?" I replied that I am an academic, and asked why he thought going public would ruin my career. The reply was, "If you don't want me to be nice, then I don't have to be nice".

I must admit that, while reading Buckmaster's account of this extraordinary exchange, I was not only left flabbergasted but mildly impressed. I suppose this reaffirmed my faith in OpenAI's messianic sociopathy that was earlier defused by Sébastien's rather polite messages to me.

Around 12 hours after the release of Buckmaster's statement, OpenAI announced that they have solved the Navier-Stokes problem. Sam Altman personally quote-tweeted OpenAI's announcement tweet and said (verbatim) "one of the most amazing moments for me in OpenAI history was watching this happen over the past week". OpenAI's announcement said "the agents sent 2.7 million messages and used approximately 130 billion output tokens" for Navier-Stokes, and the successful group involved "on the order of 10000 concurrent agents", and that the result arrived about 88 hours after the

first agents were launched. Several individuals estimated the API-equivalent costs at roughly 20-30M USD. In an X post, Noam Brown himself confirmed (verbatim) "Yes, this result cost millions of dollars."

Meanwhile, Sébastien Bubeck wrote a lengthy tweet pushing back against some of the accusations by Tristan Buckmaster. In particular, he said (verbatim):

> I was confused why one would turn an incredible source of celebration (of their achievements!) into such bickering, which is when I said that I did not understand why one would risk their career [over unfounded accusations]. Genuinely, at that moment, I was trying to care for him and do a one last ditch attempt to get a chance to give them all the credit they deserve. I deeply apologize for this extremely poor choice of words, it is the opposite of what I was trying to convey. (I should say that I retracted them on the spot by the way.)

In a response on Mastodon, Buckmaster writes (verbatim):

> Sébastien's statement below is wildly false. Not only did he never apologize, he doubled down on his threat 'if you don't want me to be nice, then i don't have to be nice'.
>
> Sam [Altman]'s tweet that Levent refused to communicate is also false and contradicted by Sebastien himself.
>
> Sam [Altman]'s tweet that they had a different approach is also false. It is probably best the OpenAI due their due diligence before making further claims.

Several other mathematicians rallied in the general direction of Buckmaster's concerns towards IP theft from OpenAI. In a Mastodon post on September 10, Andreas Thom recalls an exchange he had with OpenAI after the non-sofic group announcement, which I paste here verbatim:

> I wrote in an email to Mark Sellke and Sebastien Bubeck shortly after the surprising finding of a non-sofic group using the methods of Kun and myself:
>
> "Another point is that I and a colleague in Dresden were discussing the expander matching problem and various extensions of the work with Gabor Kun actively over the last months with ChatGPT, so that we are of course curious if that was part of

> the training data or accessible to the reasoning process. There is a certain (frankly unacceptable) lack of transparency here; and I fear it will damage the communal process of math more than the new AI-generated results will benefit the subject."
>
> Mark Sellke's complete answer was: "Regarding your conversations with ChatGPT: that did not happen."
>
> I had explicitly asked about two different things: (1) whether our conversations entered training data, and (2) whether they were accessible to the solving process. The categorical answer now looks as though it addressed only direct access under (2). No such qualification, explanation, or evidence was given. I take this as dishonesty to say the least.

Similar concerns were raised about Anthropic, in relation to the Jacobian Conjecture. In a pre-print titled "On endomorphisms of affine spaces and the Jacobian problem" by Alexander Borisov, Ofer Gabber and Adrian Vasiu, where the third author makes the following statement

> Very unfortunately and without me being aware, the in-progress version of this work dated January 14, 2025 was shared publicly at the link [...] on January 16, 2025. My understanding is that after 18 months it became a feed of the AI models.

All the disparate circumstantial evidence surround these events lead me to conclude that we are probably witnessing the biggest IP theft project in the human history, perhaps unintentionally and as a consequence of alignment issues innate to the products shipped out by AI labs, and that the AI labs are desperately trying their best to dismiss these concerns as either bad-faith accusations or unfounded paranoia from the users, compounded by techno-optimist pundits on social media trying to undermine these issues by attacking the very concept of IP itself. The latter arguments would go along the lines of the following caricature: "so what if AI has stolen your academic work without attribution, at least it'll find a cure for cancer!".

My plea to applied science industries (biotechnology and pharmaceuticals, materials science, defence contractors) and the financial sector (for instance, trading firms) that are considering integrating top of the line AI models into their operations is to take the mathematicians' concerns about intellectual property theft seriously. Do you really want your proprietary research, trade secrets and commercially sensitive data freely

roaming around the servers of AI corporations, with little material assurance as to how that information might be stored, accessed, or used?

> "I would add a Tolkien analogy and say that Math is fighting the Battle of Helm's Deep for the human role in the future of STEM right now"
> — Francesco Maggi

X.
Encounters At The End Of The World

In the midst of the Navier-Stokes chaos, past acquaintances and interns of Sébastien Bubeck came out openly on X with more accusations of unscrupulous behaviour. Sébastien reached out to me again on September 10 with a casual joke. I told him to speak frankly. I paste here some of the discussion verbatim:

> BUBECK: 100% honest answer: personally I would like to spend my time on the other areas I care about, which are probably not your cup of tea, but like, healthcare. if we have a good machine, accelerating medicine seems like a good idea.
>
> NIHIL: And your corporation would botch it up exactly the way it's botched up math academia. We know this.
>
> BUBECK: :-(
>
> NIHIL: You have left a trail of destruction at your wake. I mean I don't blame you. But that's what it is.

At this point, Sébastien asked me for (verbatim) "One Name, One Person" that I would trust within the mathematical community to make the right decisions, whom the mathematical end of OpenAI's explorations could be handed over so that it could be decided from within. I thought for a long time and came up with names of four mathematicians who could be put in a committee to decide the best practices for mathematicians moving forward. But I struck a bargain:

> Since you solicit my suggestion, here is my concrete offer. Here is a list of mathematicians, including some extremely serious people, who are opposed to Mathathon [list of 1K+ signatories in Open Letter about the Mathathon]. In the list

> you can see, for instance, that there is a person who has a 100% chance of winning the Fields medal in 2030, if it still exists.
>
> My offer is this: withdraw funding from Mathathon as an act of goodwill. This will show not just to me but to a large section of the community that OAI is serious about preserving mathematics instead of destroying it, and that you, Sebastien Bubeck, are acting in good faith.
>
> If this is acceptable, I can tell you who the 4 people you should consult are. Two of them are American mathematicians, and will very likely be willing to talk to you.

The next day, OpenAI scientist Dan Roberts announced that they are withdrawing funding from Mathathon. I kept my end of the bargain by giving Sébastien the four names I had thought of. Eventually, a nine member committee called Advisory Group on Mathematics and Artificial Intelligence (AGMAI) would be formed, supported by the Institute for Advanced Study. None of the four names I had suggested would be there, but perhaps that was expected for various reasons.

On September 29, AGMAI would announce guidelines for responsible release of results generated by AI labs. On the top of their article, they would state clearly (verbatim):

> At present, some frontier AI labs are testing advanced mathematical problems on proprietary models that remain inaccessible to the broader scientific community. Our recommendations are formulated with this practical context in mind. However, ideally, they would not do so. We want to state clearly from the start: we do not endorse this practice, and we ask them to stop testing advanced mathematical problems on proprietary models.

On October 06, OpenAI would release a GitHub repository containing around 400 mathematical results, some of which settled (or killed, depending on how you look at it) entire research programs that had been guiding certain fields for a decade or more. OpenAI would go on to declare that they have been collaborating with AGMAI, and use that a soft-endorsement for their release.

The magnitude of impact on mathematics that this release had is hard to summarize in a few lines. I refer the readers to this Proofs & Prompts article to get an idea of the reactions from the broader mathematical community. Suffice to say, working

mathematicians were not particularly thrilled by this development. Most of the proofs in the repository are unreadable to a human mathematician without repeatedly interfacing with the very AI models that created them. The task of understanding these proofs is closer to that of a cataloguer than a creative artist.

A fair share of the proofs are not formalized, which comes into question the internal AI lab epistemology regarding benchmarking. If, as some claim, these problems were indeed benchmarking problems, how is correctness of the output evaluated? Presumably, one offloads the proof into some collection of agents, with the parent system given the task to convince these agents that the proofs are correct. Upon the completion of the process, it is decided that with some sufficiently high probability, the proof is correct, and then the session is shelved. This is all speculation on my part gathered from private correspondence with various people; the truth is we do not know how these labs operate.

On October 07, OpenAI would retract three articles due to a critical sign error that kills the proof strategy in each of the results. Human mathematicians pointed out that the sign error is actually rather incidental: the proof strategy itself is unsound. I would like you to imagine a scenario where reward-hacked AI models would make subtle errors in drug design or critical infrastructure or algorithms that, if went undetected, would put consumers at risk.

In such a scenario, which seems increasingly plausible in the near future, would it be attributed to people who underplayed “AI risk”, or to the negligence of those at the helms of these powerful models? I think it would be prudent for investors in these large corporations to exercise some caution.

OpenAI and Anthropic recently announced a contribution of $5M each to a Simons Foundation initiative to, in OpenAI’s words, “help mathematicians examine, understand, and build on results produced by AI”. These corporations further promise access to their models and compute-resources to researchers supported by the Simons Foundation.

While limited and meagre contribution to a small section of the community in compensation for demolishing the wheels and pinions of academic mathematics, annihilating research programs that distinguished mathematicians and their students were working on, and sowing the seed of volatility and chaos within the mathematical

community— could seem like a gesture of goodwill, I worry that this takes us one step closer to the great institutional capture by private capital that these corporations have been actively undertaking, a project that vastly overshadows that of Elsevier, and one that will spill over into reshaping, and ultimately controlling, other aspects of the human society soon enough.

XI.
Epilogue

The story I have told here is a memoir: an account of events that transpired in the space between mathematics and AI, as witnessed from the limited vantage of someone that got a chance glimpse into the eye of the storm. Some of the details are so exceptionally surreal that I felt compelled to preserve them in writing, as faithfully as I can, in roughly the order of occurrence.

But more importantly, I wanted to illustrate that every step of the way it was the deliberate decisions from employees of these giant corporations that eventually led to the rapid rupture in the values and principles that kept the mathematical community held together. When speaking about AI, I often see people criticizing the technology: AI could never do this, or that. After living through AI corporations' obsession with mathematics in the span of the last six months, I have come to realize that it's not the tool that is dangerous, but the *people* who are at the helms of the tool.

It would be far too simple if the individuals involved were plain evil. It is more the case that their very involvement with these large corporations — which are guided by great forces of Capital beyond individual reckoning, which by means of a great churn that turns human conception of value and economic productivity upside down, effectively aim to establish a new world order with these corporations themselves at the center of it all — when suddenly faced with the great responsibilities and their misplaced altruism, end up exacerbating their innate sociopathic tendencies while making decisions for certain large sections of society. Therefore, despite the claims they make in their individual capacities that they merely want to empower humanity by rapid shipment of these powerful new tools, in reality they are very much like the Watchmen of Alan Moore.

On August 28, I had asked Sébastien Bubeck: "What do you see as the future of professional mathematics? I think you have a vision of it when you mention 'what's

coming'. This is a genuine question from someone far below". I think his response (verbatim) was enlightening:

> Call me an optimist, but I think:
>
> 1. there is a chance for much faster progress. There are so so many questions that we want answers to, and now we have a chance to see them in our lifetime. It will definitely be faster paced, and not as tranquil as it used to be.
> 2. what I was personally always drawn to (and so were the masters such as Gauss) is the "power" that a strong theorem gives you on the actual real world. Suddenly you can do things that you couldn't before. This was a bit lost I would say in the last century, but there is a chance to reclaim an extraordinary impact of math on the real world.

I never really agreed with the point of view that we, collectively as mathematicians, want "answers to questions": it had always seemed to me that the confusion *is* the point, the recognition of how much we do not know, and the process of fumbling around in the dark. Even after the AI revolution, I stand by this perspective: for every question we obtain an answer to, I think it would be prudent to ask ourselves if the questions we had asked were really the right questions we meant to ask in the first place. To me, mathematics will always be a raid on the inarticulate: a blind and ill-defined search for things that may or may not exist just on the periphery of the visible, "with shabby equipments always deteriorating in the general mess of imprecision of feeling" (T. S. Eliot, "Four Quartets").

Sébastien's second point intrigues me the most. At the time, I had thought that perhaps he is speaking of the great decoupling of mathematics and natural sciences that had happened a century ago, and maybe to an extent he did! But after so many weeks what catches my eye now is the word "power". In the six-part documentary film "Can't Get You Out Of My Head", English journalist Adam Curtis frames the deep history of artificial intelligence as follows:

> The laws of human thought that George Boole had created would become the central structure of all thinking machines, computers, because it fitted perfectly with the binary switching systems inside them, either zero or one, and it was used by the machine to create endless branching pathways of binary logic, called algorithms. Out of that was going to come the dream of artificial intelligence:

> machines that could think independently, that could then order and manage the world as a rational system, not driven by the dangerous ideologies of the past.
>
> But back in the 1960s, as the engineers began to build the first neural networks, what they had forgotten was that the system of thought they were creating inside the machines did have its own history. That it had been born out of a time when science had become deeply involved with questions of Power and Control in the British Empire. That what lay behind the computer logic was the aim of simplifying human thought, which would finally allow you to colonize the last free outpost: the human mind. But unlike the old empires, where power was visible, this power would be hidden in remote places — in the servers.

Created as a result of an accident at a US government particle physics and weapons research base in New Mexico, Dr. Manhattan is a nigh-omnipotent and nigh-omniscient being in *Watchmen* who grows increasingly distant and apathetic towards humanity, primarily due to his omniscience and nonlinear perception of time. During the course of events in the graphic novel, he is used by the US government as a weapon and eventually manipulated by Ozymandias to go along with his destructive plan.

Dr. Manhattan is portrayed as a detached and indifferent god who does not see himself as slighted by the humans who weave a web around him and complies unbothered with the events that unfold, as all of it is already visible to him ahead of time, like a clear stream of water:

> "Why does my perception of time distress you? Everything is preordained, even my responses. We're all puppets, Laurie. I'm just a puppet who can see the strings."

The jury is still out on whether or not AI will prove to be one of the consequential discoveries made by humanity to date. In concluding this essay, I instead urge the readers to consider the possibility that, in the hands of for-profit corporations and governments hungry for power and control, administered by flawed men that profess altruism but betray a sociopathic indifference to human consequences, it becomes nothing more than a silicon Leviathan: a vast and monstrous centralizing power that will eat through centuries of human values, intellectual traditions and, most important of all, the human freedom of thought.

> “The ultimate hidden truth of the world is that it is something we make, and could just as easily make differently”
> — David Graeber (1961-2020)

Subscribe to nihilunbounded

Launched a day ago

In death's twilight kingdom

Type your email... Subscribe

By subscribing, you agree to Substack's Terms of Use, and acknowledge its Information Collection Notice and Privacy Policy.



37 Likes · 13 Restacks

Discussion about this post

Comments Restacks

Write a comment...

Norman 1d

Liked by nihilunbounded

It's entirely unclear what the motivations of these transhumanist techbros even really is, or the academics who they've managed to rope into their circles despite some obvious ideological rifts, witnessed by the Bubeck-Alpoge-Buckmaster debacle (to clarify: with Buckmaster & Alpoge primarily playing the questionable role here). Alpoge seems to be oddly obsessed with involving himself in the proof of major theorems in the capacity of something like a prompt engineer or intellectual bystander. What this accomplishes for him or anyone else remains to be seen. He certainly does not give the impression of someone with a genuine intellectual curiousity for these results, but rather a desire to stamp his name on the big conjectures & prizes of mathematics, even if he has to destroy the discipline in the process to do so. In fact, I think this point needs to be emphasized. Alpoge seems to have his eyes on the Fields medal (lol) or Millenium prize (lol again), and doesnt mind what he does to mathematics or science to get there. I can only speculate here, but one almost gets the

impression that in a state of perhaps AI psychosis or meglomania, individuals like Alpoge may perhaps fancy themselves to be the last generation (and hence the best) of mathematicians. One can imagine a caricature of the great Alpoge who takes out ALL of the Millenium problems in one final apocalyptic flood of LLM slop mathematics, and in the process, "solves" mathematics once and for all.

LIKE (4) REPLY SHARE



Maxim Raginsky 12h

Liked by nihilunbounded

Excellent essay, thank you for putting the time and the thought into it! My only quibble is with the invocation of "engineering epistemology" in the context of Tsimerman's quote. This is only the image of engineering epistemology refracted through the sociopathic vision of Sam Altman and his fellow travelers. Engineering does not end with the design and synthesis of new artifacts; once these artifacts are embedded in the human milieu, responsibility and care (in the Heideggerian sense) begin. Too many of these self-proclaimed engineers forget this. For this reason, I'd add Ivan Illich's _Tools for Conviviality_ to the required reading along with Thurston's "Proof and progress in mathematics."

LIKE (2) REPLY SHARE

1 reply by nihilunbounded

2 more comments...

© 2026 nihilunbounded · Privacy · Terms · Collection notice
Substack is the home for great culture