WRITINGS ON AI AND MATHEMATICS

Contents

1 The AI dissenter viewpoint - Tasmin Chu - 9th August

In the following essay, I would like to give a reasonable defense of what I call the AI dissenter viewpoint. I have argued along broadly similar lines in two previous essays.[1] My views are continually evolving and I recognize that a no-AI future for mathematics is unlikely to succeed, but I think it would be intellectually dishonest of me to water down the argument and propose some nominally more pragmatic scheme. So let me push the Overton window a little bit further and offer my honest thoughts.

The thrust of my argument is the following: the introduction of increasingly capable LLMs has a disastrous effect on the *social context* and *professional environment* through which human mathematical understanding is produced. Moreover, we are not alone in the crisis we now face, and we should expect LLMs to affect a range of white-collar and scientific professions in the long-term. As citizens of a broader society, we have moral obligations to avoid collaborating with AI companies and improving their models with the aid of our extensive mathematical training; in fact, we should take an explicitly adversarial position to these companies and the future they are creating. When we use LLMs to generate **new proofs** of mathematical theorems, we are directly benefitting AI companies, who stand to financially and politically benefit. The more important[2] the theorem, the worse the impact. For moral reasons alone, we should not use LLMs to generate new proofs, even if increasing scientific progress might otherwise be a *good* use case for AI tools.

We should **not** embrace the use of LLMs to maximally increase mathematical progress, but instead safeguard a social context which preserves and increases human mathematical understanding. Human mathematicians must extensively cooperate to establish reasonable professional norms around AI use in this new regime. In the near-term, we should preserve the current model as much as possible until a consensus coalesces; that means an **ongoing moratorium** on asking LLMs to prove novel mathematical theorems. **Do not collaborate with AI companies, do not defect, and do not pollute the commons with LLM-generated mathematics.** Since non-defectors are obviously punished by the existence of defectors, cooperation and coordination is the only path forward. In every regime, the current model of mathematics will clearly substantially change. (For instance, I believe it is likely that the "ownership model" of mathematics will collapse, and so we should redesign our incentives to preserve human understanding of mathematics.)

1.1 The current regime: Humans competing with LLMs

Right now, in the ongoing ownership model of mathematics, we have created a situation where human mathematicians are now directly competing with LLMs. So we need to redesign our incentives, fast. Humans who are doing mathematics without AI use are spending labour and time; humans prompting LLMs are spending money on tokens.

Concern: LLM capabilities are quite spiky; their capabilities not only vary field-by-field, they vary problem-by-problem, such that an LLM may be capable of proving Hard Theorem A, yet spit out nonsense when asked to prove Easy Theorem B. Moreover, different models have varying capabilities, and some mathematicians have access to frontier models while

[1]With some important differences: after reflection, my ethical objections to LLM use have only increased. For instance, I no longer think the "ethical (collectivized) model for using AI" I outlined in my first essay would actually be ethical. (It goes without saying that it was never practical.)

[2]As determined by social consensus, first chapters of introductory math textbooks, the zeitgeist of the mathematical community, whatever.

others do not. As a result, vast inequities result for mathematicians in different research niches and with differing access to models.

Concern: Mathematicians are directly disincentivized to communicate and disseminate their ideas, proof sketches, and new theories with other people, because they can be scooped by LLM users. People are incentivized to either quickly publish lower-quality work in an LLM-accelerated environment (slop mathematics), or to fully flesh out a theory entirely alone and then drop a 100-page mathematical monograph. Long-term, high-quality projects are now significantly more risky.

Concern: Various actors (AI companies, social media users, mathematicians) will **pollute the commons** in the near future by quickly producing and generating mathematical work using LLMs and releasing it without refining, disseminating, communicating, and understanding that work carefully. In other words, they will produce **slop mathematics**. This is already evident in the appalling phenomenon of important mathematical counterexamples being dumped on social media by users who clearly have no intention of ever producing an ArXiv preprint or publication; instead, the mathematical community steps in to pick up the pieces.[3] In short, we should expect that all the worst aspects of the publish-and-perish model (the production of 'slop' mathematics, underverified proofs, etc.) will be exacerbated in the near-term LLM regime.

Concern: Early-career mathematicians will leave en masse in the next 5 years. They will respond to a field-wide sense of uncertainty and fear if reasonable professional norms do not coalesce quickly. Many of these mathematicians already face suboptimal working conditions of severe financial precarity.

Concern: Mathematicians with serious ethical concerns about AI companies are directly penalized in the near term; this fuels discrimination in our field. In other words, a person's ideological lens or political sensibilities could unacceptably curtail their career opportunities.

Broadly: My sense is that using LLMs contributes to a social ecosystem in which human mathematical understanding is broadly degraded, disincentivized, and punished. The degree to which human mathematical understanding is preserved is largely due to individual actors consciously or pro-socially moderating their LLM use.

1.2 Arguments against LLM uptake in the mathematical profession

I would like to present three different arguments for avoiding using LLMs in our profession. I should add that clearly, LLMs are complex enough tools as to create a *moral spectrum* of responsible use; but I believe the conclusion of the three arguments below is that in our profession, *the less we use them, the better.*

The social context lens (the preservation of mathematical understanding)

The introduction of highly capable LLMs imperils human understanding of mathematics, because human understanding of mathematics happens in a social context. We should expect

[3]We should also consider the dark shadow of this phenomenon, when users publish incorrect LLM-generated proofs on social media (often because they lack the mathematical training to read such proofs) and take no accountability for the errors, even when they're pointed out. This kind of thing directly erodes scientific discourse and wastes everyone's time; it's slop.

a serious loss of understanding even if we only allow LLMs to *generate* and *verify* proofs, while we *digest* these proofs as an active and living community.[4] Let me offer some arguments.

Argument 1: When humans generate mathematics, it is *useful* for mathematical understanding. *We understand mathematics more when we do it ourselves.* If we increasingly delegate the generation of mathematical proofs to LLMs, that already has a drastically negative impact on our understanding.[5] Of course we spend vast amounts of time understanding other people's work, but frequently we do so with the aim of eventually writing *original mathematical work* ourselves.[6]

Argument 2: People need to believe they add value to our field. If LLMs are pervasively used to *generate* research mathematics and eventually become more capable at generating new proofs than human mathematicians, then many people will not choose to work in mathematics. Long term, many people are not attracted to working on *solved problems*. Does reading AI-generated mathematics full-time sound like a *profession* or a *hobby*?

Argument 3: When we outsource the work of generating *research mathematics* to LLMs, we degrade the working conditions of our field. The labour of doing original mathematical work is creative, important, and enjoyable. Is refereeing papers your favourite part of your job? What about papers written by LLMs?

These are serious concerns which I cannot quiet even if we wholeheartedly pivot to incentivizing the *communication* and *digestion* of proofs (which I believe we should: see below). I think if humans increasingly relinquish the work of generating autonomous proofs, that is an incalculable loss to our understanding and our profession.

The labour lens

AI companies have extracted value from the scientific community's labour by using our papers, textbooks, and research output as training input for frontier models. Simultaneously, they exert pressures that in the near-term will result in mathematicians losing their jobs and our working conditions deteriorating. How is that fair? To state the obvious: AI companies have not turned to our field because they are benignly interested in increasing the scientific progress of our field. They use their models to prove theorems because when they announce a nonsofic group exists, they increase shareholder value and establish market dominance.

---

[4]Here, I am referencing a framework advanced by Terence Tao, who has argued there are three core components of mathematical problem-solving: proof generation, proof verification, and proof digestion (or understanding).

[5]For similar reasons, sometimes an unsolved problem can be more useful for human mathematical understanding than a solved one (for sociological reasons alone). When problems are unsolved, people try to work on them. Compare this to solved-yet-dead fields.

[6]As a metaphor: One could imagine mathematics as a mysterious, opaque language that we access through unclear means (mathematical sensory perception, physical intuition, formal proofs, mental thought, etc.) Right now, we are a community who all speak this language to various degrees, and many of us spend immense amounts of time *listening*, *understanding*, and *learning* this language from each other. But we *learn* it so that we can one day *speak* it (that is, generate useful, *not necessarily novel*, mathematical proofs), and all of us expect that speech production eventually becomes part of our regular linguistic experience. In other words, mathematics today is a living language.

Now imagine that we were able to learn many new words in this language through artificial means, and discovered that it is vastly more efficient to access this language through such means. Imagine we completely relinquished speaking this language in light of that: clearly it is just more efficient to *digest* and *learn* the language collectively than to continue speaking it ourselves. The language becomes a dead language; we study it the way we might study Ancient Greek or Latin, able to cobble some words and phrases together, but primarily acceding the production of words in this language to more authoritative source material.

People don't learn dead languages very well.

AI companies are attempting to create models which outperform humans at almost all economic tasks.[7] If they succeed, almost all of human society stands to be affected. This is so unbelievably inequitable that it is poised to be one of the greatest wealth transfers in human history.

It is incumbent upon us to **organize** in light of this and to **hold the line**. We must defend our material interests and our livelihoods, particularly when AI companies used *our* corpus of work to create these highly capable models. Many of us produced scientific work with the understanding that it would become *the collective intellectual property of humanity*; we did not produce it anticipating that it would become fodder for private interests and the creation of models which disrupt the production of exactly that scientific work.[8]

The ethical lens

In mathematics, we have a tendency to treat AI like it is some massive exogenous force which came out of thin air to disrupt our profession. But highly capable LLMs were not produced ex nihilo; they were produced by an industry with intense financial interests, appalling professional norms around safety, and executives with so little message discipline they say things like this in public:

Weirdly enough, if you think that this moment is, I don't necessarily believe this, but a lot of people would say we're living through this kind of eclipse of the human intellect where we're in the final days of humans being the primary actors on this planet, um, and that soon machines will rise. [. . .] It's a little bit like, it's, in that sense, it's a very beautiful time period to live through because in a Dionysian way, there's a lot of ugliness about it, but there's a beauty in the ugliness of when a star dies, it grows super big into the red giant, right? And it's like that, where you, as you watch this final flowering of humanity and the birthing of the machine intelligence, it's like you see this greatness in human effort.[9]

Right now, thousands of people are attempting to build artificial general intelligence which outperforms humans at all cognitive tasks. Is that a good idea? Is that democratic? Set aside for a moment empirical claims about whether or not AI companies succeed at their stated goals. Should they even be trying?

Here is my sense of the situation. **It is scientifically irresponsible to initiate a scientific project which, if it succeeds, leads to disastrous consequences for human society.** Why should we not spend our time doing frontier human genetic engineering or creating super-contagious pathogens? We could make huge scientific progress in those domains. The reason why is that if we succeeded, we would have done something deeply unethical.

In light of that, **we have moral obligations to avoid collaborating with AI companies, and in fact to resist them as much as possible**. When we use LLMs to generate novel mathematical proofs, AI companies benefit. When we offer our expertise to help mathematically benchmark models, we feed AI hype.[10] When we collaborate with AI companies, we are aiding and abetting the Manhattan Project of our time.[11] The labour of

[7] See OpenAI's charter.

[8] Intellectual property alone is an ethical minefield: AI companies have disrespected intellectual property to an extraordinary degree. I think the scientific community at large is (rightfully) reluctant to lead with the logic of Disney and Nintendo, but again it bears mentioning that AI companies have extracted enormous amounts of wealth from harvesting our labour in unheard-of ways.

[9] This is a verbatim quote from Dean W. Ball, the head of strategic futures at OpenAI.

[10] Even if all we offer is a sober assessment of the limitations of their current capabilities.

[11] A turn of phrase borrowed from Max Weinreich in this essay.

mathematicians is critical for helping AI capabilities improve; it's past time for a boycott. I believe that we are desperately in need of a global AI pause. We can either lend our support to the movement for an AI pause, or continue to manufacture consent for AI companies.

Of course, many mathematicians are not directly building LLMs or working for AI companies, and we may console ourselves that we are just downstream beneficiaries of this irresponsible scientific project. Maybe we can just use LLMs to resolve our own scientific curiosities and finally make some progress on our favourite pet conjecture. After all, we tell ourselves, scientific progress is morally neutral. Whether humans or machines generate proofs is morally neutral. But this is clearly not true. *The more important the theorem, the more these companies stand to benefit. The more we use LLMs, the more social consent we manufacture for this dangerous and rogue industry.*

The fact that increased AI capabilities are broadly bad for mathematicians is a corollary of a more general proposition: *increased AI capabilities are bad for everyone.* So we can't use LLMs, even if nominally a proof is a proof, whether it comes from a human or a machine.

1.3 Rebuttals to the view above

I now address some rebuttals to the view above.

LLMs are not that good at mathematics yet. LLMs may be good at [y] thing that was never really the essence of math (e.g. problem-solving, finding counterexamples to conjectures, tedious computations of technical lemmas), but it will not replace [x] aspect of our mathematical work which is actually the essence of mathematics (building theory, teaching younger mathematicians, clearly expositing mathematical proofs).

Unconvincing. What happens if this intermediate regime does not last? What happens when [x] aspect of our career can be performed by LLMs too?

Of course, the one kind of labour that LLMs cannot do for us, definitionally, is the human digestion of mathematical proofs. But the human digestion of mathematical proofs occurs in a social context. See the **social context argument** above.

Human mathematical understanding will just become much higher-level. LLMs will function similarly to calculators and computers.

Low-level details are also key to understanding things deeply. Moreover, LLMs are substantially different from calculators and computers in their capabilities. See the **social context argument** above.

Mathematicians are paid to produce theorems and it is unfair to taxpayers and the public to not maximally accelerate mathematical progress with LLMs.

I think this is a terrible argument. The main way the public benefits from government-funded math research is from the *ancillary benefits* of having a large, public research mathematics community. I think it is pretty easy to argue that human mathematicians are more useful to the public than LLMs. Mathematicians can explain mathematics to the public (in the classroom, on Youtube, in popular science magazines). They teach undergraduates across scientific disciplines foundational tools like calculus and linear algebra. They educate and train graduate students on problem-solving skills which transfer to industry and academia. The public largely doesn't care about the production of new mathematical theorems; the main benefits they receive are indirect benefits like those above.

LLMs may prove to be cheaper at producing novel mathematical theorems than human mathematicians, but the indirect benefits the public receives will only decrease if we replace mathematicians with LLMs.

If the theorem economy is artificial[12], then there is no harm in simply producing as many new theorems as possible with LLMs in the interest of satisfying our mathematical understanding.

Theorems may have little *market value* (in plainer terms: the majority of mathematical work is useless). However, that does not entail that it is consequence-free to produce as many theorems as possible with LLMs. See: the *social context* and *ethical anti-AI* arguments above.

If we institute some kind of professional ban on using LLMs, that actually decreases and hinders human mathematical understanding (possibly to an enormous degree). It is incoherent to defend human mathematical understanding while asking mathematicians to desist from LLM use, and to inhibit scientific progress, which is nominally the goal of a research community. In other words, scientific progress is one of the positive use cases of artificial intelligence.

I think this is one of the most interesting and compelling rebuttals to the arguments I have outlined above. It is by far the strongest.

One possible rebuttal is to reiterate the *social context argument*. Already, a vanishingly small fraction of human society is interested in learning the proofs of famous theorems. If LLMs become more capable than humans at generating mathematics, I think many people will lose interest in mathematics. If our profession is hollowed from the inside out and mathematics becomes a hobbyist endeavour, then it's not really the frontier that matters anyway. *We can know; we won't know.*

But I think the more intellectually honest thing is to concede the point. Yes, we won't make the most mathematical progress possible without the use of LLMs. But we have compelling reasons to refrain from using them anyway. See: the **ethical anti-AI argument** above.

People can just lie and use LLMs anyway. Mathematics is a competitive environment and actors who don't use LLMs will be punished. A worldwide total ban on AI use is unlikely to succeed. There are morally grey methods of AI use (literature searches, learning classical mathematics, etc.) which would undermine the efficacy of an outright ban. Even if we control the use of LLMs in some countries, other countries may not abide by such agreements. In short, the AI dissenter view is unrealistic and unlikely to succeed.

I concede all of these points. I expect that the actual course of events does not land us in the AI dissenter's ideal world, and at best in some intermediate regime.

Nevertheless, I would like to shift the Overton window to a more radical place by being as intellectually honest as possible. From a pragmatic perspective, I think if we don't have the intellectual courage to articulate 'naive' points of view, then we have no chance of them every succeeding. From a moral perspective, I think the AI dissenter viewpoint is just correct. To state a fairly obvious observation in moral philosophy, there is no reason to believe the *morality* of a given action is at all related to the *facility* of performing that action. But

[12]In the sense that most theorems are not useful products for taxpayers or for industry, and our professional norms have created the theorem economy.

actually, I think we still have very good options for protecting our field from AI companies; see the *radical proposals* below.[13]

If LLMs outcompete humans at research mathematics, then we should just allow LLMs to largely monopolize the work of generating research mathematics (that is, producing proofs). The job of a mathematician could instead be about digesting and disseminating frontier AI-generated mathematics. The work of our profession will be teaching each other math, educating undergraduates, service work, and community.

I actually find this very interesting. Broadly, I think the mathematical community is coalescing around redesigning the incentives in this direction, and I support this initiative. These sorts of labour are valuable. The work of it assembles into the labour of a real profession. I think it is an interesting rebuttal which addresses concerns about both *labour* and *the social context* by which human mathematical understanding is produced.

There are two issues with this proposal to my mind. In this theoretical regime, the work of *generating* original research mathematics now plays a marginal career in our careers. That is an enormous loss.

The other is that this course of action does not address the *ethical anti-AI arguments*. Mathematicians will still have aided and abetted AI companies in improving AI capabilities. When novel mathematical theorems are proved by us with extensive AI use, then AI companies stand to gain *enormously*, in terms of both stock value and sociocultural sway. Every time you use AI to prove a new "big" theorem, you are helping AI companies (even if the ownership model is long dead).

1.4 Some radical proposals

I now suggest some proposals which can help address the AI crisis in mathematics. Critically, supporting these proposals does not require too much anti-AI buy-in; I believe these proposals could be supported by a coalition of mathematicians.

(1) **Move away from the ownership model and the current theorem economy.** Redesign the incentives to encourage human mathematical understanding. Incentivize work like: mathematical communication; education; and textbook writing, instead of just the production of original mathematical work.

Even those of us who are attached to the human *generation* of mathematics should support this initiative, because this initiative directly disincentivizes defection. It curbs the creation of slop mathematics. People will be much less motivated to use

[13]On a personal level, I feel there is a sense in which it is fine if I do not succeed as an AI dissenter. If the working conditions of mathematics in the future become

- "Use LLMs as much as possible to pump out theorems and produce new mathematics" in the near-term; and
- "Read interesting LLM-generated frontier mathematics in my spare time while the thrust of my job becomes teaching undergraduates who offload cognition onto LLMs where possible, knowing that I have aided and abetted AI development" in the long-term,

then I'm not very interested in doing mathematics. For me, the answer will be simple: I'll just leave and pursue another career which is less sensitive to automation. So will many other PhD students, postdocs, and people without job security; they'll go off to work in the relational sector, finance, the military, and for AI companies. When I talk to other PhD students my age, the overwhelming consensus is that we will leave if working conditions deteriorate in the LLM-uptake regime. So if we all do nothing and give into entropy, the outcome is simple: early-career researchers will leave. The research mathematics community will be impoverished as a result, but I won't actually have to suffer the consequences; you will.

LLMs to thoughtlessly produce novel mathematical theorems if they aren't rewarded for it. Already, too many theorems were proven for the community to be able to reasonably digest them.

(2) **Limit the number of papers that an individual is allowed to output per year.** This is an immediate way to punish actors who want to flood ArXiv with LLM slop, and to encourage higher-quality work. This essay by Mario Pasquato expands on this proposition at greater length.

(3) **Determine the leverage we have over AI companies and exercise it.** What can we do to oppose this industry? Options could include: a field-wide boycott on collaborating with AI companies; deliberately polluting the commons with bad and faulty mathematics to wreck models; restricting frontier mathematics from being accessed as training data by AI companies; demanding direct remuneration from AI companies; building coalitions and organizing politically; writing public essays and speaking to the media.

(4) **Organize at your own university.** I reiterate here urgent proposals from mathematician Max Weinreich: *Every college mathematics department needs a committee, formal or not, to regularly discuss and make recommendations for AI use.* ***You*** *can start this committee.* ***Time is too short*** *to coordinate a national project from the top down or to wait for someone else to do it. Departments hold the keys to ensuring that human mathematicians may at least persist as a minority in the mathematical world in many ways. Departments can lead by establishing anti-AI policies for student work, by reserving hire lines for mathematicians who eschew AI, by valuing AI-free papers more highly in tenure promotion, and by increasing resources for talks, seminars, and conference attendance. [emphasis mine]*

(5) **At an individual level: avoid using LLMs to prove novel mathematical theorems. Don't defect and don't pollute the commons.** Every time you do this, you directly benefit AI companies and you marginalize the production of human-generated mathematics. You behave unethically. If you use LLMs, limit their use to applications which don't directly compete with humans. In all cases, act conscientiously and pro-socially, and don't engage in mathematical arbitrage.

1.5 Thoughts on cooperation and defection

Clearly, it is extremely difficult for humans to cooperate on a large scale even when it is in their best interests to do so (see: all of human history). But any path forward requires extensive cooperation. Everything from here on out is game theory. People can make locally rational decisions to globally disastrous effect. If you don't make your own views known in a public forum, then sympathetic actors are unable to coordinate with you. Within mathematics, we are obviously deeply divided as a community on our views about how to proceed in an era of increasingly capable LLMs, and every person will have to make concessions.

I would like to push back against this manufactured consensus that we all have to start using AI tools or risk falling behind. I have been told it is exceedingly naive to resist these tools and to ask people to cooperate and not defect. Probably this is true. But I think it's also *exceedingly naive* to embrace the use of LLMs in mathematics and expect good long-term outcomes for our field. I do not find the people whose argument is to submit to entropy *practically* convincing. I do not find even thoughtful and nuanced AI proponents *morally* convincing. I find AI companies morally despicable and I continue to dissent.

All of us know that LLMs are certainly not better than the mathematical community in aggregate, yet. What happens when LLM capabilities at mathematics dramatically increase? An entire industry exists which will only meet its financial targets if it succeeds at exactly that. Thus, we will be on the back foot if we have not already built a coalition.

It takes a certain kind of person to form a union. It takes a certain kind of community to cooperate *en masse*. Our working conditions involve relentless competition; thus, from the outset, it seems unlikely that we are that community.[14] But I'm not so sure. Only time will tell.

[14]A few of my thoughts here being informed by this comment of Mario Pasquato: *Academia is unfortunately very competitive and individualistic, so the battle is uphill. In the future we may have an external enemy to rally against, so this may change. It takes a certain type of person to start a union and a certain type of community to join it en masse. Mathematicians don't look like it, but who knows?*

2 Care for a little more AI? - Hugo Duminil-Copin - 30th August

Disclaimer. *The effects of AI are vast, and there is much to be said about its broader impact on society. One should also remind that mathematicians have a role to play in advancing research on AI safety. I do not feel sufficiently qualified to address these wider questions here, so I will instead focus on the direct effects that the current use of AI is having on the mathematical community and provide my experience as a practicing mathematician.*

The progress of AI in mathematics is spectacular. I will not venture here to predict what AI may accomplish in the future, but I can already say that, without question, frontier models are far better than I am in many mathematical tasks. To mention at least one, they possess encyclopedic knowledge of my field and neighboring domains, while I sometimes struggle to remember some of my own proofs.

Some problems with direct and urgent applications will obviously profit from advances in AI. Their solutions will enable further progress in related sciences and contribute (one hopes) to the greater good. It would be absurd to deny this.

But what benefits some areas of our discipline may be ill-suited to others. It is entirely legitimate to ask whether we truly need and should really welcome the rapid proliferation of AI generated proofs in mathematics.

2.1 The $\theta(p_c) = 0$ conjecture

So far, I have been fortunate not to be directly affected by an AI solving a problem on which I was actively working. But lightning has struck close by. In April, some colleagues went through this painful experience: ChatGPT Pro proved a difficult problem in percolation that we used to call, among ourselves, $p_c < 1$. It now seems only a matter of time before the most famous conjecture in our field, $\theta(p_c) = 0$, also falls to the bulldozers. Indeed, it became clear that AI will probably prove $\theta(p_c) = 0$ before humans do. It is certainly an excellent candidate for frontier models as there may be a simple proof, something hidden that everybody missed. Will the proof come from an employee at OpenAI or Anthropic? From another company eager to "help" mathematicians in distress? More prosaically, from a mathematician armed with their favorite AI model? Or even from a casual user with virtually no mathematical expertise? Only time will tell.

A very simple statement, not much background, and a high difficulty (for human) is precisely what made the $\theta(p_c) = 0$ conjecture so beautiful. This conjecture stands out among percolation problems because it is inspiring and intriguing, not because solving it would obviously unlock a vast new area of mathematics.

My own story with $\theta(p_c) = 0$ is instructive in this respect. I no longer work actively on this question, but there was a time when this conjecture inhabited my mathematical dreams. During the early years of my career, I returned to it regularly. As you can guess, I did not find a proof. And yet I consider my relationship with this conjecture a success. A success measured neither in theorems, nor talks, nor prestigious prizes. A success far more precious to me.

Let me provide a small sample of the encounters made possible by the conjecture. I met my main coauthor in percolation theory, Vincent Tassion, because we were both trying to solve the same simplified version of it. It was while trying an analogue of $\theta(p_c) = 0$ for the Ising

model that I started working with Michael Aizenman. Together with Michael and Vladas Sidoravicius, we did prove the conjecture for Ising. More importantly, we were able to breathe new life into a powerful tool that Michael helped develop in the eighties: the random-current representation. This tool later enabled us to solve a multitude of other problems, in particular the triviality of the model in four dimensions, and it is now used by many. There are many other examples I could cite off the top of my head. My failed attempts to prove the $\theta(p_c) = 0$ conjecture generated dozens of ideas that I later repurposed in other contexts, leading to discoveries I would never have imagined making.

A mathematical question is much more than a theorem waiting to be proved; it is a source of momentum. First and foremost, it is a lighthouse in the night: it illuminates and guides mathematicians in their wanderings, both esthetic and scientific.

Seeing the beautiful story of $\theta(p_c) = 0$ coming to its final words would obviously not be the end of the world. But I had hoped that one day, a young mathematician, through the originality of their thinking, would come and show us what all of us had missed. I could then have marveled at the creativity of the human mind. But depriving me of this pleasure is obviously not the main issue here.

2.2 What is my problem with the way generative AI is being used in mathematics today?

Certainly not that AI-generated proofs are systematically unreadable. This is not true. The proof of $p_c < 1$, for instance, is short and particularly elegant.

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.

Let the lighthouses built for us shine a little longer. We owe it to our predecessors. Let us not profane our most beautiful conjectures by using them as mere "benchmarks" for the next frontier model.

What are we supposed to tell young PhD students entering a field? Which problems should they work on? How can they build mathematical abilities when the most accessible problems may fall at any moment under the battering ram of compulsive users of AI models? What Holy Grail should they pursue for the next several years if the proofs of major conjectures might suddenly appear, from one day to the next, on some social network?

What is at stake is not only how we choose problems, but the culture and ethic that shape our profession. The social consequences for our community, especially for the younger generation, are devastating. Some will argue that mathematicians are simply an endangered species. This may be true, but I do not believe it.

2.3 A profound misalignment between AI and mathematicians' interests.

When mathematicians say that the process matters more than the solution, this is not an empty statement. The richness of what emerges from repeated attempts, failures, detours, and encounters is extraordinary. I obviously cannot speak for everyone, but I am deeply convinced that what I have described from my own experience with $\theta(p_c) = 0$ resonates with the paths many of my colleagues have followed.

Understanding the mathematical world requires both personal and collective exploration. The same formal concept takes on a different shape in the mind of each mathematician, through a long and gradual process of appropriation. The ideas that live and circulate within our community are the true jewels of mathematics. Like Peter Scholze says in the context of the Leiden declaration:

In my experience, mathematical ideas, like children, must be nurtured and grow over the years. Just like I do not want my children to be educated by AI, I am pondering my mathematical ideas without use of AI...

Let us recall that our mission as mathematicians is not limited to producing theorems, we must before anything produce understanding. We remain the first link of the chain turning very abstract notions mastered by few experts into concepts that become part of the culture of next generations.

Ideas need time to emerge, to be refined, tested against one another, understood by multiple mathematicians. That is the price we must pay for ideas to mature in our minds and ultimately become genuine intuitions. Once properly mastered, these intuitions can be passed to other mathematicians, then teachers, and kids. This is exactly what happened with 0, the = sign, negative numbers, infinity, fractals, etc. And this process is ongoing with more recent concepts.

Unfortunately, it is precisely this slow process of intellectual development that is threatened by the way generative AI is currently being used. The consequences for mathematics, education, and culture will be massive if nothing is done.

2.4 What should we do?

Mathematics is far from being an isolated case. One of the fundamental questions raised by generative AI is how far we are willing to delegate our cognitive abilities to a machine. I have always regarded mathematics as one of the highest forms of human intellectual exploration: a way of pushing the limits of what we are able, individually and collectively, to understand, imagine, and abstract.

If we mathematicians lay down our arms and accept delegating our creativity, even partially, to machines, what other form of intellectual adventure will not follow? Instead, we must persevere in our human exploration, accepting that it takes time, that it sometimes leads us astray, and that it often remains fruitless.

Let us keep on the traditions of a discipline several thousand years old, and teach how to search, doubt, experiment, fail, and restart. Training future teachers, engineers, and researchers is impossible without personal experience of what it truly means to do mathematics ourselves. In my case, it is this experience that I like to share with students, school kids, and the public, more than the precise statements of my results. As a community, we have a responsibility to continue playing this role. In full honesty, we have even a great deal of room for improvement here.

While some areas of mathematical research may very well embrace the possibility of speeding up the resolution process leading to spectacular applications in other fields, we should not succumb to a purely utilitarian vision of our discipline.

I believe that many areas of mathematics simply do not stand to benefit much from the assistance of AI. I certainly do not have a straightforward rule of thumb for determining which ones, and even if I did, I am in no position to provide a general recommendation. In fact, it

has never been the intention of this text to offer a solution (one that I do not have). Still, I sense that many mathematicians share a similar view of recent developments, and I wanted to give voice to it, in the hope that some of them might recognize in these words something of what they themselves feel.

I will conclude on a more personal note: in my own research, I have chosen not to rely on artificial intelligence to replace my creative process.

3 Tao's Mathstodon - Mining open problems - Terence Tao - 8th September

I wrote recently about how the collection of good, fruitful open problems is now being mined in a non-renewable fashion, leading to the potential scenario of these problems becoming scarce. This may seem unintuitive at first, since the set of possible problems one could ask is infinite. Perhaps the following analogy can help: a country or region can suffer a critical shortage of drinking water while simultaneously being surrounded by a massive ocean.

One can easily generate any number of open problems in mathematics at will, such as working out the $10^{10^{10}}$th digit of π. But the vast majority of such problems are not worth focusing attention on: they show no particular propensity to reveal any further insights or connections to other questions, or may either be too easy or too impossible relative to known techniques to learn anything from the exercise.

Working out whether a question is actually worth highlighting is a lengthy, deliberate, and subjective process, often informed by historical experience on what good mathematics was generated (or not generated) while working on earlier problems of this type. In particular, being aware of the "difficulty landscape" in a field - what questions are very easy to answer with known methods, which ones can be solved but only with some effort, and which ones are impossible - is of crucial importance in making such determinations. And some problems only become interesting after an external connection is made. For instance, there could hypothetically be an unexpected connection between the Riemann zeta function and the 10^{10^n}th digits of π for various n... at which point the previous question of determining the $10^{10^{10}}$th digit suddenly becomes relevant again. (I should emphasize though that this specific scenario is incredibly unlikely to actually be the case; I use it only as a hypothetical iilustration.)

Every new advance in mathematics, whether it comes from technique, technology, or infrastructure (such as access to libraries of past literature) reduces the difficulty of solving problems. This is generally a good thing; but it comes at the cost of flattening out the difficulty landscape of a field, to the point where one can no longer discern its geometry to the extent that promising questions can be extracted within the range of applicability of the tool. Often this effect is counteracted by the ability of such a tool to enlarge the radius of the sphere of results one can plausibly reach, creating new boundaries to fruitfully explore.

However, one notable feature of the current AI era is the absence of any definitive such boundaries. While AI tools have flattened the difficulty landscape now in many areas of the subject, thus destroying the ability to locate promising new problems in that area, there are no clear frontiers that are separating the "AI-feasible" problems from the "AI-hard" problems (which certainly still exist, given that the difficulty level of problems are unbounded, and can even be undecidable). This is in part due to the rapidly changing nature of the technology, but also compounded by the refusal of AI companies to disclose their negative results, or reveal the process towards obtaining their solutions.

In fact, it is now the identification of a promising problem which is the scarce and precious resource. We have now seen that even the rumor of someone working on a problem can trigger a massive amount of AI-powered effort to flatten it before the original research project has time to reach its full potential. The incentives may now be pointing in the direction of no longer sharing any promising research directions with the broader community, which would

reverse centuries of traditions of open science and do serious long-term damage to the future of the field.

In short, the indiscriminate use of powerful solution-extraction tools can achieve the immediate short-term goal of solving problems at hand, but at the cost of sustaining the ecosystem for the next wave of progress, or in understanding the progress already obtained.

While it may be technically infeasible to completely prohibit the use of automated tools to perform indiscriminate solution extraction, I believe that we can still designate many classes of problems as being desirous of a careful analysis that not only solves the problem, but identifies insights from the solution process, and learn more about the difficulty landscape for nearby problems, and for which raw solutions without such analysis would be of negligible or even negative value for these purposes. This is analogous to how a modern food donation drive no longer accepts arbitrary contributions even when they are verified to be technically edible, but instead maintains explicit and socially accepted standards on what level of contributions are actually sought.

4 Duet for the end of math - Matilde Marcolli - 9th September

Is the rise of mathematical proficiency in the latest corporate models of AI a harbinger of the end of mathematical research as a profession, as some prominent members of our community have recently argued? In the middle of a deluge of claims of vanquished Famous Conjectures™ and solved Big Problems™, as mathematicians we are increasingly and loudly asking (and being asked) how to justify our existence. Does the rise of AI-generated mathematics imply that humans and machines are engaged in a dramatic duet for the end of time? (Mathematical time at the least, if not the civilizational collapse that some AI developers have repeatedly been warning us about.)

There seem to be several strange assumptions behind this reasoning and these conclusions, which reveal that something must have gone badly wrong, since a very long time, in how our mathematical community perceives the goals and the values in our profession.

Why is the automated solving of some well known open mathematical conjectures sending so many shock waves of panic in our professional community? This points, first and foremost, to the disproportionate role that the idea of Famous Conjectures™ and Big Problems™ has come to play in the field. When and how did we start viewing mathematics as a zero-sum game, with some predetermined scarce trophies out there that one alone will conquer while all the others will fail? Who convinced you of that? When? And why did you let it happen?

The process of mathematical creativity is incredibly varied, rich, and complex, and the solving of certain specific narrowly defined outstanding problems is but one particular aspect of it. It is, however, an aspect that has been endowed with a disproportionately large attribution of value. Career advancement, prizes, professional status, and ultimately power are tied up primarily to conquering one of a narrow set of goals that someone else (those who have priorly acquired that kind of power status in the field) have singled out as valuable. These goals are achievable, generally, through a technical hyperspecialization within a very clearly defined subdiscipline (and therefore a very narrow subcommunity) of mathematics. Since these became the primary and unquestionable goals, we reoriented the whole functioning of our community, training of students and postdocs, hierarchies of journal publishing, control of reward attribution, around this particular system of values. We are getting to the point, to paraphrase a famous aphorism, where it is becoming easier to envision the end of mathematics than the end mathematicians' obsession with Famous Conjectures™.

And yet, if we take a look at another field of science that is very close in methods and in spirit to mathematics, namely theoretical physics, we begin to see how it is actually possible for things to look very different. In theoretical physics, for obvious reasons, what someone else was not able to do one or two centuries ago has absolutely no relevance on what drives the development of the field today. This is of course because our understanding of the physical universe has so dramatically changed with respect to a century ago (in 1926 it was not even known yet that neutrons existed) to make such comparisons irrelevant. Mathematics by contrast has a long persistence, so that it is indeed possible to have open problems that trail along with the community for centuries without losing relevance. The question remains though, on why they should catalyze so much value attribution. So let's look a little more at how our next of kin in theoretical physics view the question of value attribution. Since I have been most of my life in between these two communities, I have been observing since long time these profound sociological differences. In theoretical physics I see a system of value attribution that focuses very strongly on a balance between the originality of ideas and their degree of connectedness, meaning that an idea that is highly novel and that at the

same time connects to many genuinely different questions and problems that many other people are thinking about typically receives very high value attribution. This points in a completely different direction from the value attribution that we have been forcing upon the mathematical community, because it privileges broadness over hyperspecialization, it privileges abundance-driven collaboration over scarcity-driven competitiveness, and it regards technical muscle-flexing as secondary to ideation, so that the perspective of automating the first is not perceived as threatening the latter.

Most importantly, these different systems of value are not an intrinsic reality, ordained by the gods or by the natural order of the universe, but are fairly arbitrary choices that are made by our human communities, and that can and should be questioned and not given for granted. It is an act of propaganda, fueled like all such acts by an entrenched system of power, that maintains the view of goals and values as immutable and unquestionable. We have often heard mathematicians speak of some kind of "mathematical Platonism", where mathematical concepts and their properties have ontological existence in some unspecified "world of ideas" and are simply being "discovered", not invented, by the human mind. Depending on our personal background, we may have different reactions to such claims, from just viewing them as an exceedingly metaphysical rephrasing of the undeniable "unreasonable effectiveness" of mathematics in the sciences, to regarding them as a benign delusion, to be viewed with a sympathetic eye if not enthusiastically embraced. This is not so harmless and inconsequential, though, and rather than a mere delusion, it is akin to a subtle propaganda instrument. Because if mathematics is indeed predetermined and enshrined in stone in some olympian other-world, determined by more-than-human forces, and all we can do about it is uncover it, then it really is a zero-sum game of competing armed conquerors over a virgin but immutable territory. Ironically, the ultimate conqueror of this other-than-human landscape may turn out to be itself other-than-human, hence the widespread panic about artificial minds out-counter-exampling human mathematicians in a conquering spree worthy of Napoleon, Alexander the Great, or Genghis Khan, laying our precious beloved Famous Conjectures™ in ruins.

Overtime a small set of problems with a varyingly long history came to dominate the mathematical imagination. Some of them, like Fermat's last theorem, spanned the course of centuries until a final bout of virtuoso arithmetic geometry brought about its triumphant resolution. Others, like the Riemann hypothesis, still haunt the mathematical night, amidst hopeful claims and long programs. In addition to these widely known cases there is a fair number of other conjectures which you are likely to consider crucially important problems if you happen to belong to one specific narrow subfield of math, and you have likely never heard of if you don't. What is it about all of these famous and less famous, invariably hard, Big Problems™? There is an undeniable sense in which these questions are interesting. Typically they are because, in the process of thinking about a particularly hard question, a lot of new mathematics gets developed that advances the field as a whole. The Riemann hypothesis is currently unsolved (at the time of this writing), but many people, in the process of thinking about the problem, have developed a very significant amount of very interesting mathematics. That is certainly a valuable goal. But is the Famous Conjecture™, invested with its aura of sacred object, actually needed for that goal? The idea of kettling mathematical research into narrow streets overseen by the dominant presence of Big Problems™ gained prominence with programmatic efforts like the Hilbert problem-list and its more recent millennial revival. While some of the problems in Hilbert's list were broad in scope, the understanding of what constitute a Big Problem™ is increasingly reflecting a competition system of pain and rewards

that is simply a system of power, and that can easily become detrimental to our creativity and to the inner life of the mind.

Now we are all faced with a new reality in which sudden unpredictable burps of artificial intelligence, often with unauthorized access to ongoing unpublished work of human mathematicians gobbled into their garagantuan training sets, can liquidate our centenarian Big Problems™ in a moment's time. One may be tempted to say, to quote another famous aphorism, that these AI-solutions "fill a much needed gap" since probably quite a bit of very interesting mathematics that would have been developed while thinking about these questions will now no longer be pursued once it becomes detached from its value attributor. Is having sizable parts of the development of mathematics entangled with a small number of very specific questions really helpful to the field? Doing mathematics consists of many different activities and the creativity is everywhere along that process, but we choose to reward it in a very uneven way. It is our right as a community to decide what we want to value, why do we exercise this right so narrowly? What if, instead of attributing importance to a mathematical result solely on the basis of whether it cracks a question on which N people have previously failed for M years, with N and M sufficiently large, we would attribute value more like theoretical physics does, on the basis of originality of ideas, collaboration, and broad connectedness between different areas of mathematics. It is possible to envision a genuine plurality of values that much more closely represents the richness of mathematical thinking. Why do you care about what someone did not prove a century ago? Why is that the only source of value? Why do we grow into the world of professional mathematicians, and in turn educate the next generation to prioritize working on what others may attribute value to, rather than listening to our own inner voice. How do we become enablers of a vision of mathematics as a conquering battle field rather than as a shared voyage of self discovery?

What is for us humans the act of doing mathematics? I'd say mathematics is two very important things at once: a language to make sense of reality and a form of self-expression, much like the arts. When seen from this perspective, the existence of artificial intelligences that can generate solutions to mathematical problems becomes essentially irrelevant, and is just left to individuals and their ethical considerations whether to regard AI as a useful scientific tool, an opaque proprietary instrument of corporate greed and oppression, a new possibility for increased accessibility, an uncontrollable danger, or all of the above at once. If you are a poet, the fact that a large language model can compose a dochmiac metered poem in Ancient Greek upon demand (I checked) is entirely irrelevant, because that would simply not be *your* poem. So don't ask yourself which are the Big Problems™ and Famous Conjectures™ you should be working on and what is your reason of existence now that an artificial intelligence can get them before you've even finished your morning coffee. Ask yourself rather what is the mathematics that lives only inside you and nowhere else. The one that makes you dream, like the music that no-one has ever heard, that wakes you up in the middle of the night, playing itself inside your head, and demanding to be written. Your mathematics, that is like the poem that composes itself in the mind and explodes onto the page because your life demands it, the gesture on canvas that crosses the boundaries of thought and action. Mathematics, like all of these other activities, the music, the poetry, the painting, is a form of self-expression, that we all collectively share because we are human, like we share our appreciation and performance of the arts and our profound need for them.

Current forms of artificial intelligence have become very proficient at simulating human products, including the arts and including mathematics, because they have been fed every single product of human civilization that could be shoveled into their training data. What

comes out of it may be good or evil, useful or not, dangerous or tame, but there is one thing it will not be: the thing that lives right now inside your mind and that is asking to be expressed, the thing that only you can give a voice to.

You don't normally worry that the existence of an AI capable of writing acceptable texts in human languages will imply that every single novel in the world will end up having already been written. You do (I hope) recognize the absurdity of such a thought. And yet large numbers of mathematicians now worry that every theorem will have been proven by AI, leaving the mathematical profession in tatters. This narrowing of the horizon that we have all implicitly accepted is astonishing! Yes, if all we see in mathematics is a few shiny trophies consisting of Famous Conjectures™ (old and well seasoned) sanctioned by a small number of established individuals (old and well seasoned too, if possible) then we may surely risk running out of all such rewards, through the intervention of new technologies. Well, the sooner the better! Good riddance! Because with that way of envisioning mathematics being shattered before our eyes, maybe we can finally start to have a conversation that the mathematical community should have had a long time ago. We can and should come to accept the idea that the work of a mathematician can be more like that of an orchestra conductor or of a music composer and less like the virtuoso violinist, single instrument performer, whose enormous skills and creativity spin to interpretive perfection music composed by someone else. The skills of the orchestra conductor are different from those of the solo instrumentalist, with a less hyperspecialized technique and a broader sense of the orchestra as a whole, of how sound circulates and different instruments complement each other. The world of music knows how to distribute value more evenly to such different skills. It is not difficult to imagine a path of mathematical education and professional performance that is more like the conductor and the composer, and less like the virtuoso single instrument performer, or at least that can acknowledge that such different roles exist. If we insist on focusing our entire profession solely on the tackling of conjectures as the unique goal and reward, then we may be out of business quite fast, not because of the end of mathematics but because of the end of our imagination.

5 Tao's Mathstodon - On being like children - Terence Tao - 10th September

"Toutes les grandes personnes ont d'abord été des enfants. (Mais peu d'entre elles s'en souviennent.)" [All grown-ups were once children... but only few of them remember it.] Antoine de Saint-Exupéry, "Le Petit Prince."

An impromptu game of "who can name the largest number?" is an example I often point to as a way in which young children teach each other about non-trivial mathematical ideas and facts, such as infinity and the concept of proof by contradiction (embodied in this case by one child eventually realizing that they can defeat any other child's guess by adding "plus one" to that guess, thus demonstrating by contradiction[15]that there is no largest natural number).

Prematurely putting such games to an end by immediately providing the nominal goal — for instance, providing these children with a brief lecture on infinity and the unbounded nature of the natural numbers — defeats the entire "purpose" of such curiosity-driven play, insofar as play *has* a purpose: to explore, to learn, to develop, and to simply have fun.

Children eventually become adults, and "put away childish things". Significant portions of our time become devoted to optimizing "important" and immediate goals: weighing the costs and benefits of various decisions, and acting on them accordingly. Unstructured play and exploration become restricted to one's hobbies or personal activities only. This is so normalized in modern adult life that most of us simply take for granted that this sort of goal optimization is what anything "serious" is necessarily about.

But science — particularly basic, curiosity-driven sciences such as pure mathematics — is a notable counterexample to this general rule: one of the rare and precious disciplines in which childlike wonder and exploration can be put to "serious" use. We can ask questions which appear frivolous and of no direct application (such as the global regularity problem for the Navier-Stokes equations, if truth be told); but if they are well-chosen, then scientists or mathematicians can explore them very fruitfully with the right combination of adult discipline and expertise, and childlike whimsy and fearlessness. We are still trying to achieve the goals we set for ourselves eventually; but the best outcomes arise when the journey is unhurried, convoluted, and serendipitous. Optimizing prematurely to reach the nominal goal can extinguish these unexpected detours, stories, and discoveries before they have a chance to manifest themselves.

Many in my profession are drawn to it in part because of this ability to direct these childlike impulses to productive use, which leads surprisingly often (by an extremely indirect pipeline) to actual advances in science, technology or engineering as per Eugene Wigner's "unreasonable effectiveness of mathematics". But, perhaps in order to impress the other adults observing (and, through public funding agencies and the like, financially supporting), we often downplay the "childish" aspects of the craft, in favor of the more tangible "serious" outcomes — including the hard, objective target of solving some designated open problem.

Until recently, this polite fiction has worked, because human mathematicians knew, possibly subconsciously, to use both childlike and adult skills when solving these problems, even if they might only publicly admit to the latter. But modern AI tools, when directed without such expert supervision, can now be aimed all too easily at the ostensible goals of the field,

[15]To be pedantic, this is a proof by negation rather than a proof by contradiction, but that distinction is not the point of the anecdote I wish to make.

without any incentive to take the slow, whimsical path. The flag is captured, the goal scored, and the problem is solved; but at the cost of lessons learned, insights gained, collaborations formed, and new targets located.

In this modern era of heavy AI use, it is important to remember what it is like to be a child.

6 A Severe Misalignment of AI in mathematics - Terence Tao - 11th September

I am proud to be among the list of 25 initial signatories — all Fields Medallists — to the declaration below, which grew out of discussions between ourselves over the last week. We have also posted our declaration on this web page, and (similarly to the Leiden declaration) invite further signatures. (It is unfortunate that we did not have the time to have a more consultative process, as with Leiden; but we decided that the urgency of the situation was such that we needed to release a statement sooner rather than later.)

See also this recent article in the Economist regarding our declaration, and a brief interview with James Maynard on this topic. A French version of this declaration was published in Le Monde.

Over the last few months, the mathematical capabilities of LLMs have improved dramatically, to the point that they can solve major outstanding problems in many fields of mathematics. However, the push by AI companies to solve mathematical problems as a benchmark is detrimental to the science of mathematics, and to the mathematical community. The goals of the AI companies and the goals of the mathematical community are severely misaligned. We see these as part of broader alignment issues impacting other scientific and creative professions, as well as the whole of society.

Research mathematics deals with understanding basic structures of shapes, numbers, and natural phenomena. Over the course of generations, it has built a large corpus of sophisticated ideas, methods, abstractions, and other tools to comprehend the mathematical landscape. In turn, modern technologies and sciences are based on mathematical tools.

Famous problems have often served as landmarks and lighthouses against which one can measure an improved understanding of this landscape. Solving one of these problems has been a certain sign of new insights and interesting methods, which would then be studied by a community of mathematicians, through a long and arduous process of talks, discussions, simplifications. At the end of this process, one will ideally find a textbook presentation of the results suitable for any graduate or even undergraduate student to study. Some of the mathematical ideas pursue their journey even further to become, decades or centuries after, tools that are understood and used by the whole population.

The mathematical community functions, in many ways, as a miniature version of humanity. It consists of individuals using a wide variety of different approaches, joined by core values. The most precious resources of our profession are students and ideas, and these we nurture with great care. We feel responsible to let them grow to their full potential, until they can live a life of their own in the mathematical world. For students we often suggest problems with the core intention of developing skills making them well-positioned for advances in research and elsewhere. Our ideas we disseminate in talks, private discussions and careful writeups, connecting them to the previous ideas of others. These processes invariably take time and are based on human interaction.

In recent months, the success of AI in solving major mathematical problems has made headlines even outside mathematical circles. But solving problems is only a tool and proxy for achieving the primary goal of conceptual understanding and insight. Forgetting this in the world of AI may turn the tool against the primary goal. Indeed, the mass production at faster and faster pace of “true/false” statements could destroy fertile ground instead of breathing life into new ideas.

Often these solutions are announced in a rush, leaving no time for a proper writeup, the isolation of new methods and ideas, and citing relevant previous work of others. As in all creative professions, this raises severe attribution and plagiarism questions. Moreover, without the willing mathematicians who must take care of their development and integration into the mathematical canon, AI-conceived ideas would never become fully alive and the crucial human transmission chain between mathematicians would be lost.

We are witnessing a general threat to intellectual work, with misalignment between the outcome of the use of AI and its initial purpose. In many fields and activities, years of training have traditionally served not only to produce a final answer or product, but also to develop understanding and the ability to formulate new questions and ideas. However, building on a vast body of previous human work, AI systems are becoming increasingly capable of producing the results of such work directly, and these goals cease to align. The issues the mathematical community faces now are similar to issues that other scientific and creative professions are facing, and indicate issues that all of humanity might face: how to make sure that, as AI changes the way work is done, we do not lose sight of what that work was meant to achieve in the first place.

AI offers the potential of enhancing and accelerating genuine mathematical study and understanding. Mathematics as a profession will need to adapt to these changes in several ways. However, whether these changes ultimately benefit the field or have a destructive effect will in large part be determined by the decisions of the humans in control of this new technology.

These issues must be addressed urgently, in the mathematical community, by the companies developing these technologies and, more broadly, by a society that will confront similar problems in many other forms of intellectual work.

7 After Math - Silvia De Toffoli and Eamon Duede - 12th September

On September 8th, 2026, OpenAI announced that it had produced an AI-generated solution to the Navier–Stokes existence and smoothness problem, one of the seven Millennium Prize Problems. The announcement kicked off debate over credit allocation and the respective contributions of humans and machines to the result. Moreover, the announcement intensified already circulating comparisons with earlier AI conquests in domains believed to otherwise exemplify human intellectual prowess. In a recent statement, Tristan Buckmaster, one of the mathematicians involved in the Navier–Stokes saga, wrote: "This is a Deep Blue–Kasparov moment."

Existential questions for mathematics follow naturally: if AI can now provide answers to questions at the very frontier of mathematics, is the discipline on the verge of being "solved" as many have said of chess and Go? Like chess and Go players, should mathematicians just "keep playing" and rearrange their practices?

There is something right about the "keep playing" response. As philosopher C. Thi Nguyen (2019) has been insisting, the purpose of playing a game is not exhausted by its aim (winning). The real point is not only the outcome but the process. This is perhaps clearer with a party game such as Twister than with chess: the aim of playing Twister is certainly not winning. But something similar also applies to deep intellectual games, like chess and Go. For instance, playing a game of Go well can be an achievement in defeat.

But in the context of mathematical practice, this feels like an unnecessary retreat. Instead, we can make a stronger move: reject the characterization of mathematics as a game that makes the retreat seem necessary in the first place.

The question, then, is not simply what comes after math, once AI can answer its hardest questions. It is also what we are after when we do mathematics in the first place.

The narrative that AI has "solved" mathematics rests on two assumptions, both seductive and plausible, but both wrong:

(1) AI really did solve a problem in mathematics.
(2) Mathematics is only about solving problems.

The first assumption is wrong because to really solve a mathematical problem, providing a mere answer (even if formally certified) is not sufficient. What is missing is an intelligible proof that human mathematicians can understand and use to advance the aims of mathematics. And, as we will argue below, even if AI were to give us just that, the story would not be over because the second assumption is wrong. Mathematics is clearly a much broader enterprise than just problem solving. Mathematicians strive to develop new concepts and theories, to ask and answer new questions, to unify disparate areas, to educate and sustain scholarly communities, and to produce work that is valued for its beauty and depth.

We can (and should) therefore reject the narrative of AI defeating humans at mathematics and start thinking hard about what mathematics really is and what we want it to be.

7.1 Not All Answers Are Solutions

OpenAI produced an answer to the question of whether Navier–Stokes can develop a singularity: yes. In The Hitchhiker's Guide to the Galaxy, Deep Thought produced an answer to life, the universe, and everything: 42. Neither is exactly what we wanted.

Of course, OpenAI gave us much more than "yes." Deep Thought offered only a number, whereas OpenAI produced two artifacts that many are willing to call proofs. The first is a Lean formalization certifying validity. This was accompanied by a manuscript that appears to contain the corresponding informal proof. So, why is this still dissatisfying? The reason has to do with the underlying notion of proof itself.

There are, in fact, two notions of proof: a logical notion and an intelligible notion.

Modern logic characterizes proof in terms of deductive validity such that a proof can be checked by a mechanical procedure that does not itself require understanding of the mathematical argument. A Lean formalization meets these standards exactly and a Lean formalization of the Navier–Stokes result is therefore a genuine and important contribution: by meeting the demands of the logical notion of proof, it secures certainty.

But mathematicians also want something else from proof. They want understanding (Thurston 1994). They want to know what makes a proposition true. This kind of knowledge trades in mathematical ideas that they can grasp, communicate to other experts, connect with existing knowledge, and use to make further progress. This is the intelligible notion of proof. As of now, it is not clear that OpenAI's result has given the mathematical community the kind of value that one expects from the intelligible notion of proof.

Genuine proofs are at the same time logical and intelligible proofs. Historically, the two notions have tended to run together. This is because no mathematician could produce an enormously complicated logical proof without first grasping some of the key shareable ideas that made the theorem true. The logical notion of proof was primarily used to verify the correctness of intelligible proofs (Burgess and De Toffoli 2022).

But with AI, these two notions can now come apart dramatically. We can end up with formal proofs that float free from any intelligible proof.

This is not a criticism of formal proof. The converse problem is at least as serious. An intelligible mathematical argument can convey a grand idea while failing to establish that the result is actually true. Jaffe and Quinn (1993) famously used Thurston's geometrization theorem for Haken three-manifolds as an example: a major insight accompanied by insufficiently complete proofs could become a "roadblock rather than an inspiration." And one motivation for Hales's Flyspeck formalization project was to verify that the intelligible (but hard to check) proof presented for the Kepler conjecture was, indeed, a genuine proof (Hales et al. 2009).

Therefore, falling short of either the logical or intelligible notion creates roadblocks where genuine proofs clear the way for mathematical progress. A real mathematical solution requires both logical correctness and intelligibility.

This is particularly clear in the case of the seven Millennium Prize Problems. They were not selected because mathematicians merely wanted seven answers, but rather because they wanted fruitful solutions. The Clay Mathematics Institute itself explains why proof matters in the case of Navier–Stokes: "Because a proof gives not only certitude, but also understanding."

What OpenAI has given us is an answer. But it is not clear that they have delivered a fruitful solution. Perhaps, we will find that they have, but at the moment, the situation is far from clear. A genuine solution will provide adequate grounds for believing the result but also an intelligible mathematical argument that allows the result to become part of mathematics as understood and practiced by mathematicians.

Nevertheless, if it turns out that what OpenAI has provided is a mere answer, this is not enough to dispel the existential threat that mathematics is facing. Future AI systems are likely to produce genuine proofs that are at once formally certified and fully intelligible to mathematicians. So, current concerns that mathematics is on the verge of being "solved" by AI are not fully dispelled by simply insisting on genuine solutions rather than mere answers.

7.2 You Need More than Solutions to "Solve Math"

If future AI systems will produce genuine proofs, logically correct and intelligible, like those produced by "master" mathematicians, it would still be incorrect to think that mathematics would have been "solved" as some say that chess or Go have been solved.

In chess and Go, we accept radically uneven competition between humans and machines because both are, in the relevant sense, playing the same game.

But mathematics is not (or at least not only) a game. To begin with, there is no winner. Mathematics is not an adversarial game with determinate conditions for victory. It is certainly true that mathematicians compete with one another for fame, prizes, jobs, and credit. Chess players do those things too. But chess players also win chess. There is no corresponding condition for winning mathematics. There is no mathematical checkmate.

In mathematics, it is more natural to treat AI as an assistant rather than as a competitor. As Jeremy Avigad (2026) puts it, "We should keep in mind that AI is nothing more than technology, designed to serve our purposes. It is misguided to think of mathematicians as competing with AI; when we drive a car, we aren't competing to see who can go faster, and when we use a phone, we aren't competing to see who can speak louder."

But there is a deeper, and in many ways prior, problem with the competition framing. It requires accepting the assumption that solving problems is the activity by which mathematical success should be measured.

Genuine problem solving is certainly one of the principal aims of mathematics. It is not, however, its only aim. Mathematics is a body of knowledge engaged with, interpreted, and digested by a scholarly community and not a registry of results in the abstract. This simple point has even motivated an entire movement in the philosophy of mathematics: the philosophy of mathematical practice.

Terence Tao (2026) lists many goals of mathematics beyond problem solving. These include developing new theories and techniques, understanding the world, sustaining a community, training the next generation of mathematicians, contributing to cumulative knowledge, and creating works of aesthetic value. Of course, these have been positively correlated with genuine solutions.

But AI breaks that correlation, for the same reason it separates the two notions of proof. So, even genuine solutions would not satisfy us.

This is not moving the goalposts but recognizing that any specific goalpost is inadequate. If mathematics is a game, it is an infinite one.

This attitude is not reactionary. We reject both the concession that logically establishing a theorem is sufficient for a genuine proof and the reduction of "AI for mathematics" to proving theorems. Accepting this, we may find many opportunities for AI in mathematics to support human mathematical flourishing.

7.3 Aftermath

We do not deny that, if OpenAI's announcement is correct, this is an extraordinary achievement. But we should get clear about what type of achievement it is. At this moment, it is an answer, not a solution. And, even if in time the result reveals itself as a genuine solution, we have argued that, in the practice of mathematics, solutions are not everything.

The urgency of rethinking what we value in mathematics is already being recognized within the mathematical community. In a recent declaration initially signed by 25 Fields Medallists, mathematicians warn of a "severe misalignment" between the goals of AI companies and those of the mathematical community.

Mathematicians need to do more to examine their norms. The priority norm is not the problem here (though it is likely a separate problem). The current failure to distinguish genuine solutions from answers, and the growing focus on problem-solving alone, are. This way of thinking is inspired by the current credit economy in mathematics and, as David Bessis recently discussed in his blog, the credit economy needs rethinking.

AI presents mathematics not with an ending but with a choice about what mathematical practice should become. If mathematical success comes to be identified too closely with the production of certified answers, mathematics risks adapting itself to precisely those features that are easiest to benchmark and automate away.

If, instead, mathematicians treat AI as a technology for advancing its long-standing and centrally human purposes, the technology may come to contribute to an accelerated flourishing and enrichment of the discipline. The important question is, therefore, not whether AI will defeat mathematicians, but which mathematical ends we want AI to serve.

What remains in the aftermath is not merely leftovers for humans to scramble for once machines have devoured all of the real problems. Rather, it is an opportunity to clarify what mathematics is all about. We should ask again what we are after when we do mathematics.

8 Wimbledon, the U.S. Open, and the future of mathematics - Steven Strogatz - 12th September

WIRED reported today that I began to cry while talking about AI and mathematics. But the article didn't explain what moved me.

The truth is, I'm not entirely sure myself.

(A disclosure upfront: ChatGPT helped me write this post. I'm tweaking it a little to sound more like myself.)

First, thanks to WIRED reporter Isabella Ward. She had the difficult job of distilling a long, wide-ranging conversation into a timely piece about the amazing events of the past week. And she had to put it together in one day! A lot inevitably got left out.

OK, so why did I get choked up? It wasn't about AI taking my job. I'm 67 and nearing retirement. I also love playing around with AI and find it exhilarating. I use it in research, brainstorming, and learning—and, as you can see, to put this post together.

So what was it then? I think it was when I tried to convey the magnificent 4,000-year story of our subject to the reporter. Mathematics, I said, is more than its results. It's a human way of seeing and understanding, a conversation in which one generation explains to the next not only what is true, but why. The pleasure of that understanding, and the new discoveries it makes possible, are part of what mathematics is for.

It's an honor to be part of that long conversation—all the beautiful and important work, and the benefits that have accrued to humanity.

Then, suddenly, the tidal wave of AI hits math in 2026.

Something about the grandeur of the tradition and the suddenness of the change may have conspired to get me worked up.

But I'm not sure it was that. It might have been when I started riffing about Tristan Buckmaster. I don't know him, but from what I've read over this unsettling past week, he reminds me of many mathematicians I've known: pure, clear, and scrupulous, committed to truth, and perhaps unprepared for the corporate world.

During the WIRED interview, I mentioned Wimbledon four times as an analogy. To me, Buckmaster represents the Wimbledon of mathematics: white clothes, tradition, etiquette, and the ideal of being a gentle person. How you play the game matters.

The U.S. Open is different: louder, brasher, full of individual style, crassness, energy, and orientation toward the future. It's also more open and democratic than Wimbledon. I get what's appealing about that.

With the men's final coming tomorrow, I find myself wondering whether AI could make mathematics at the cutting edge more like the U.S. Open than Wimbledon. It could open the gates, allowing millions of people without years of specialized training to pour in and make discoveries.

That's thrilling. And I don't want to be the stodgy old guard. So yes, I love the U.S. Open ... and I also love Wimbledon. (I went for the first time a few years ago, and got choked up then too!)

For the past two years, Alex Townsend and I have been writing a book that traces the marvelous ideas that brought humanity to this dizzying moment: the power unleashed by supercharging math with computers, the gifts it has given us, and what may be yet to come.

So all of that, I think, may have been what set me off: the grandeur of a 4,000-year human tradition, the inconceivability of what may lie ahead—and the possibility that something precious may disappear along the way.

9 Deep theorems were scarce and difficult and so became an effective mechanism to identify deep thought. AI has broken this system. - Bryna Kra - 13th September

For generations, mathematicians have treated the production of theorems as a clear measure of success. The stronger the theorem, the deeper the proof, the more surprising the connections, the greater the achievement. Positions and prizes are based on these theorems and the mathematicians making the breakthroughs set the directions for future research.

But theorem production was only part of what we cared about. It was a proxy for something harder to measure: understanding. A major breakthrough meant that years were invested in learning a subject and uncovering hidden aspects, accompanied by work to make the answer apparent to the community. Deep theorems were scarce and difficult and so became an effective mechanism to identify deep thought.

AI has broken this system.

Artificial intelligence has lowered the cost of producing sophisticated proofs. As the models improve, the list of deep conjectures that fall will grow. Someone with little knowledge in a field can now generate a manuscript that reads like a polished article, cites the literature, and combines techniques that would have taken years to master. The machines are producing solutions faster than the mathematical community can read them, much less digest them.

This week the changes arrived in my field. The Nivat conjecture is a beautiful problem at the intersection of combinatorics and dynamical systems. Imagine an infinite grid of square tiles, with each tile colored either red or blue. Look at every rectangular window of a fixed size, say n tiles wide and m tiles high, and count how many different color patterns show up in that window. The conjecture says that if there are at most nm such patterns, then the entire coloring repeats under some fixed, nonzero shift of the grid. In other words, some precise limit on local behavior gives rise to simple global behavior, periodicity.

Though this statement can be explained without formulas or higher mathematics, the conjecture resisted our efforts for nearly three decades.

In work that we first circulated in 2012, Van Cyr and I proved a partial result, showing that periodicity holds when the number of patterns is at most half the area of the window. We translated the combinatorial question into one about the geometry of dynamical systems, studying ways in which information in the system does (or does not) propagate. Progress on challenging problems does not happen in a vacuum, and our work began by studying and building on earlier results in the literature [1, 2, 3]. Soon after, Jarkko Kari and Michal Szabados developed a different approach, viewing the combinatorial question as an algebraic object, using certain types of polynomials to find the infinite patterns.

To those of us who worked in the area, this was a tantalizing situation. Two different methods to approach the same problem seemed to shout to us: find the conceptual bridge between them. We tried, without success, and moved on to other problems. Colleagues continued pushing the boundaries of what we knew, combining the varied techniques in increasingly powerful ways [1, 2, 3].

But synthesizing different approaches is one of the ways in which frontier models excel. A machine does not need years to absorb distinct areas of mathematics before it can find the connections. It can translate notation, compare partial results, and try thousands of

combinations without being discouraged. The Nivat conjecture has been proposed for inclusion in Google DeepMind's Formal Conjectures project, turning the problem into a target for automated reasoning and making more people aware of the question.

The manuscripts have started to arrive. This week I received several purported proofs of the full conjecture from researchers making their first foray into the subject. Some acknowledged using AI, but only for polishing the English or checking the proof. When I asked if the authors would meet over Zoom to explain their arguments, none accepted the invite.

Silence does not mean that the authors used AI and a proposed proof is not a theorem. But merely disclosing AI use and checking correctness of the proof, both of which are essential, miss the point. Suppose these results are correct. That is a mathematical contribution. Yet it opens more questions. What did we learn? What did the authors contribute? What impact will this result have?

A proof is more than a certificate that something is true. Instead, it is a story, a picture, an insight, an explanation. A proof highlights novel ideas and opens new directions for what we should ask next. It becomes part of the toolkit of the community. A deep theorem changes how we think, not because of its statement, but because of what it teaches us. As Bill Thurston wrote on MathOverflow in a 2010 response to a question about what mathematicians do: "The product of mathematics is clarity and understanding. Not theorems, by themselves."

These new texts certainly contain knowledge, just as a data file of bits of 0's and 1's contains information. But human authorship is more than the string of words that prove something. It is intellectual responsibility, understanding of a new concept, the ability to respond to questions, the sorting of new ideas from existing scaffolding. And perhaps most importantly, it is incorporating the result into the corpus of our understanding.

I welcome AI-generated proofs and believe that human-only proofs will become rare. This is a profound change of perspective, but mathematicians have always used external tools and machines. AI is already a powerful mathematical instrument and we, as a community, need to incorporate it responsibly into our practice of mathematics. The question is not if machines should participate in discovery. They already do. The question our community needs to address is what we value in our mathematics now that certain kinds of discovery are no longer a scarce resource.

The tradition of journals publishing novel results, hiring committees rewarding those publications, and prize committees recognizing the person who completed the result do not suffice. Peer review is a slow process that relies on the unpaid expertise of a small group of overworked researchers. Any single submission can be handled, but the current volume is drowning the reviewers. The careers of young researchers depend on producing theorems, with incentive to produce as much as possible as quickly as possible. The time for deep reflection, necessary for deep understanding, does not happen when someone types the question into a model and quickly produces a manuscript.

All of these issues existed before AI entered our field. Unfortunately, we no longer have the luxury of addressing them in a leisurely manner.

Mathematics is at a watershed moment. Our incentive structures are misaligned with the prolific output of highly accessible AI models. Careful verification and stellar exposition will not happen when there is no reward for those tasks. Producing a paper is no longer enough: authors must be able to explain the proof's mechanism and how they arrived at this point. Journals need to distinguish and credit the roles of discovery, proof, formalization

and explanation. Exposition must become an important component of any major intellectual achievement. We advance mathematics not by what we write, but by what we learn, check, apply, and teach others.

Finding the right concept, crafting the right definition, and formulating new questions have always been part of our intellectual work. The creation of collective understanding, the ability to communicate ideas in a lecture, and the sharing of ideas informally that spark research have always been admired. Such achievements are harder to quantify than theorem production, and so have been treated as by-products. But these are the key aspects shaping the future of our field and their value is at least as important as knocking down the next conjecture. Our incentive system needs to be realigned to reflect what we truly value, and not just what we can easily measure.

Mathematics is the canary in a coal mine for all parts of intellectual life. Its claims can be carefully checked and so we are witnessing the disruption in real time. But when AI disrupts the tried and tested standards in a field like mathematics, what will it do to fields where evidence and interpretation are more contested? Science, law, public policy, and art will all have to find ways to assess what is valuable amidst the abundant output.

The arrival of machine generated mathematics does not replace the role of human expertise. Instead it reveals why we valued that expertise. The goal of our discipline is not simply to resolve conjectures, but to enlarge our capacity to reason and open our minds to new ways of understanding. It will take our entire community to create new standards, discover new ways of creating mathematics, work with AI models to open new horizons of knowledge, and develop the tools to approach them. It's truly an exciting time to be a mathematician.

10 A beginning for mathematics - Daniel Litt - 14th September

Three years ago, AI systems could not reliably add two numbers. A year ago, internal models at OpenAI and DeepMind received the equivalent of a gold-medal score on the IMO. Now, these systems are autonomously resolving major open questions. It's hard to imagine this trend continuing for another year, but I expect it will. It is clear that this will require a radical rethinking of our profession.

A few weeks ago, I gave a talk titled The End of Mathematics. If you only read the title[16], you might guess that this talk was about how, soon, AI will "solve" math. That's not what it was about. The talk instead laid out a gloomy vision of the future, in which, despite the possibility of AI systems that are robustly superhuman at mathematics, the design of our institutions causes human understanding of mathematics, and possibly even mathematical progress in the abstract, to stall. I think we will avoid this future, but I also think it is plausibly the default if academic mathematics does not adapt. Despite my relative enthusiasm for the use of AI to do mathematics, I share this view with many of its detractors.

Here I want to lay out, instead, a positive vision of the future of mathematics, and the human practice of mathematics. I claim we can deepen human understanding even as the production of interesting mathematics becomes less dependent on it.

This essay will take as a premise that AI systems that are robustly superhuman at most or all aspects of mathematics will be here soon. But the concrete changes to our institutions I propose only require accepting the weaker premise that the production of mathematical text is becoming increasingly disconnected from mathematical understanding.

10.1 What are we even trying to do here?

I think it has now become clear that there is no consensus in the mathematical community as to what our goals are. Some of us want to solve problems; some of us think of mathematics as play or as poetry. For some: "Wir müssen wissen – wir werden wissen."[17] Some of us think we are penetrating the mysteries of the platonic realm. Some of us think the goal is to embody love of and understanding of mathematics,[18] and to transmit that love and understanding to the next generation.

My personal, if self-referential, answers are:

(1) We're trying to produce and understand high quality mathematics.
(2) We're trying to produce high quality mathematicians.

These goals should be construed broadly. What high quality mathematics consists of has changed quite dramatically over time; we come to its definition as a community. We are not just training PhD students to do research in mathematics. A substantial part of our job,

[16] I regret choosing this title.
[17] Hilbert's full opinion is as relevant today as ever: 'We must not believe those, who today, with philosophical bearing and deliberative tone, prophesy the fall of culture and accept the ignorabimus. For us there is no ignorabimus, and in my opinion none whatever in natural science. In opposition to the foolish ignorabimus our slogan shall be Wir müssen wissen – wir werden wissen ("We must know – we will know").'
[18] I owe this phrasing to Peli Grietzer.

though perhaps an underemphasized one, is to educate the general public about high quality mathematics and mathematical thinking.[19]

Whatever our goals are, we've operationalized them primarily through *proving theorems*. Almost all papers or PhD theses have a main theorem, and ostensibly a proof of it. But it should be clear that the goal of mathematics is not to prove theorems; if it was, it would be trivial to automate. A computer or monkey could easily start at the axioms of ZFC and iteratively apply deduction rules to them, with no attention whatsoever paid to their meaning. It has had particular significance when a theorem resolves an *open problem*, especially one that has resisted substantial effort. Again this is easily automated; our computer or monkey can simply conjecture all mathematical propositions in alphabetical order.

The general attitude of our community towards a technology that can prove theorems and solve open problems suggests that these operationalizations of our values are at best incomplete.

The prospect of automating mathematics by enumerating all conjectures, and all proofs of ZFC, is probably not so disturbing to you. But let us for a moment assume the computer or monkey is very smart; perhaps it understands the results it is proving, and writes beautiful expositions thereof. Perhaps it has a good sense of what we find interesting, and is primarily focusing on those questions. Perhaps it has, in the course of enumerating theorems of ZFC, answered many of our most pressing open questions, and is asking many more fundamental open questions. Is there still a need for human mathematicians?

I think so. This machine might produce answers we value, but it would not, in itself, produce human understanding of those answers. In fact I think we are at the beginning of an incredible, wonderful explosion of mathematics, and if we value human understanding, there will be more need for human mathematicians than ever before. But the profession will have to change.

In the course of this change, we will have to decide what to hold on to and what to throw away. Some things I would like to preserve: learning seminars; serendipitous conversations that spark an idea; students knocking on a professor's door to chat about math. A robust community learning exciting new mathematics. Thousands of people that, together, slowly start to resolve their confusion.

I worry that much of what has been written on this topic, including some of my own past writing, focuses too much on trying to preserve the precise shape of the *institutions of academic mathematics*, rather than our values. How can we preserve the journal and peer review system?[20] How can we protect the arXiv? How can we keep our role as gatekeepers? If you have internalized the fact that existing AI systems can produce relatively high quality results for the marginal cost of a few dollars, the idea that any semblance of the current equilibrium can survive what's coming is absurd.

As we try to find a new equilibrium, we could try to chase the edge of model capabilities. Right now AI systems arguably underperform us at theory-building, asking questions, exposition,... so we could prioritize and reward those skills. I think this is unwise: compare the speed at which the academy adapts to the speed at which model capabilities improve. We

[19]Note that this list consists mostly of internal and social-relational functions (understanding, coming to a determination of what's interesting, training, and so on). This is in contrast to our operationalizations (proving theorems, solving problems, etc.).

[20]This system was already close to breaking before AI; it is overdue for radical reform.

need to consider the endgame. If the models remain incapable in some domain, we can adjust later.

Before I propose some relatively concrete steps we can take, let me remark on what we're trying to protect mathematics from. There is a lot of anger at AI labs, and certain individuals at those labs. But whatever our judgment of the labs, we need a plan that does not depend on AI capabilities disappearing. The basic issue is not the labs' behavior, ethical or not.[21] It's the technology itself. I think there is some belief that the labs will "move on" from math next year, be nationalized or broken up, or that a financial bubble will pop, somehow returning things to normal, or...

But there is no way our institutions can survive unchanged when anyone with a laptop and a few hundred dollars can generate what would have been an Annals paper last year. AI does not care if you are anti-AI.

10.2 Producing high-quality mathematicians

The most urgent question our profession needs to answer right now is: what should our students be doing? It's now possible to produce a PhD thesis one hasn't even read; in terms of demonstrating understanding, mathematical text is worth the paper it is printed on.[22] The text no longer reliably conveys a signal about the person who produced it.

In my view we should welcome interesting mathematical results regardless of provenance. But our institutions have historically relied on the same signal to indicate both mathematical progress and mathematical expertise. These now must be distinguished.

I propose the following reconceptualization of the goal of a mathematics PhD: to become a world expert on some interesting, deep topic, and to be able to convey that interest and understanding to others. Part of operationalizing this might be a thesis, but the degree would be awarded primarily on the basis of a rigorous defense, in which the student explains the topic to their examiners until they are satisfied. While we might require the topic to be original, its provenance – AI or not – is irrelevant.[23]

How different would this look from current PhDs? I think students would still meet with an advisor, who might suggest a topic. That topic could be explored with AI assistance, or not, but the student would be responsible for understanding it; it might be much more open-ended and larger than the typical PhD is currently. The student would be trained to ask interesting questions and try to resolve them, by whatever means. To keep students on track, there might be regular meetings in which the student is asked to independently work through an unfamiliar example, apply a technique in a new case, etc.

The allocative aspects of our job (hiring, graduate admissions, etc.) are in dire need of reform if we want to retain human mathematical expertise. Broadly speaking I think we should focus on rewarding skill in the parts of our jobs that cannot be automated: the internal (e.g. understanding mathematics) and social-relational parts, and operationalizations that hew as closely to those aspects of the profession as possible. For example, **talks** and **sustained mathematical discussion** now demonstrate understanding much better than papers. Once

[21]Obviously some of it has not been ethical. But even if every lab had behaved perfectly, the capabilities of AI systems would still force us to radically adapt our institutions.

[22]Which is not to say the text is necessarily uninteresting.

[23]This is a practical necessity. There is no way to enforce restrictions on provenance, and attempting to do so will only create incentives to conceal use of AI. But I find it unlikely that someone whose only contribution was to push a button, and who did not engage deeply with the material, would be able to pass a rigorous defense.

AI systems improve at exposition and "digestion," this will be even more the case. We already interview faculty hires; we must now do the same for graduate admissions.

I think we should try to foster a **robust seminar culture** in which speakers are expected to explain their topic to the audience's satisfaction. Much has been written recently (by myself among others) about the fact that we are primarily interested in understanding, not merely the truth value of mathematical statements. If that is the case, let us make sure we actually understand each other.

Right now the use of AI systems to do mathematics above some minimum bar relies on the fact that our community has produced many open conjectures, whose interest is evidenced by the existence of human mathematicians who care about them.[24] The recent importance of this fact suggests to me our community plays a very important function that we have, arguably, underrated: namely, figuring out what is interesting. It is not entirely clear to me how to operationalize this, but one possibility might be to reward the construction of **research programs** (either with help from AI systems or otherwise) that persuade others of their worthiness.

To be clear, I am not saying that AI systems will not be able to ask interesting questions, make interesting conjectures, pursue interesting programs, and so on. I think they most likely will, resulting in the production of an abundance of PDFs. The contents of some of those PDFs may even have important applications. But others will primarily be of interest because they tell us something fundamental about basic mathematical objects, and accrue value only if we can and do engage with them. It seems to me that it will be up to us to **build a community** of researchers to do so, and we should reward mathematicians who do. And even if the AI is asking excellent questions, there is no reason to think it will ask the same questions we would.

All of these changes are oriented towards increasing the amount we talk to each other about mathematics. It seems to me that this would be positive even in a world with no AI.

I think there is room in this world both for mathematicians who, like me, are enthusiastic about AI, and for those who do not use it. But as the models begin to produce huge quantities of mathematics, it will not be possible to avoid their outputs entirely.

10.3 Producing high-quality mathematics

As we think about how to reshape our profession, it's important to understand that, whether one likes it or not,[25] it's impossible to stop people, amateur or professional, from pushing a button to produce mathematics. The idea that we will persuade people not to play around with math, or that we will be able to "reserve" problems for graduate students, is just not realistic.[26] And we shouldn't want to do this!

[24]To be clear, many open conjectures are less interesting than one might have hoped, post hoc, and are generally not an end in themselves. They are often meant to measure our failure to understand some object, but they are sometimes resolved without improving that understanding.

[25]On balance, I think I like it, though I am sometimes annoyed to find slop PDFs in my inbox. It took me some time to understand that these PDFs expressed a need for understanding; a person elicited them, often without being able to meaningfully engage with their contents, and needed to know that someone could engage, and that someone cared.

[26]That we cannot reserve a problem for a graduate student does not mean we can't give them the opportunity to work on it. This is compatible with the reconceptualization of a PhD outlined previously.

There is now more interest in math than at any other time in history. We should be ecstatic for mathematics's sake, even as we are concerned about mathematicians and mathematical expertise. And by and large, the value of this button-pressing comes from the mathematical community. If a conjecture falls in the woods and no one is around to hear it, who cares?[27] For the abundance of new mathematics to have value outside application, we will need an abundance of new mathematicians. And for results with applications, we will want people to be capable of understanding their assumptions and consequences.

I wrote above that solving problems and resolving open conjectures is an incomplete operationalization of our values. But nonetheless it is important to solve problems and resolve conjectures! The provenance of such solutions only matters insofar as it intersects with the existing structure of the profession (incentives, prestige, and so on). It is obvious that structure needs to change in any case.

Mathematics used to be the cheapest of the sciences. I think the biggest change we are facing is that now, some portion of our questions will be answerable via a cash injection. I know some of my colleagues find this distressing. Previously those questions might have brought together a research community, led to interesting auxiliary developments, and so on. This contingent progress may now no longer occur.

But don't you believe in mathematics!? There will always be more to learn. If a basic question can be resolved for the cost[28] of a nice dinner, we should be delighted. But that's only the beginning. We will ask what the answer explains, and what it helps us understand. It will lead to many more new questions, some of which can in turn be resolved for the cost of a nice dinner, and others which renew our confusion and lead to the development of a research community.

Our industrious new helpers will be churning out an unbelievable amount of math, pursuing our interests or perhaps their own. We will have our own questions, and confusions; sometimes they will be resolved by the models, and sometimes they won't. Sometimes the answers will be complicated, and we'll devote a learning seminar to them. Sometimes progress will be minimal, but the question itself will be so motivating it gives rise to a research community.

A student will be confused. They will knock on their professor's door. Maybe the two of them will ask a model for help, or maybe not, but first they might spend some time at the blackboard thinking through the question. And the model might give them a beautiful explanation, but we all know that's not enough; no one can understand mathematics for us. We have got to do the work.

There is so much more to learn – an infinite amount. We've always been at the beginning, and we always will be.

[27]Some have suggested that interest in using AI to answer mathematical questions may soon fade. It is hard for me to see how this will happen as long as questions we care about remain unanswered.

[28]By this I mean marginal cost. Michael Groechenig points out to me that it is unclear that we should directly compare the cost of a machine proving a theorem to the cost of a human doing so, as the products of this work are arguably different. Only one of them produces understanding and expertise in a human being, which I think we might value independent of the result itself.

11 Why I do mathematical research - Emily Riehl - 14th September

Like many mathematicians, I was drawn to the field as a little girl by its aesthetics: questions that I found endlessly fascinating and arguments that struck me as delightful and elegant. As I grew older, I came to appreciate how personal mathematical aesthetics are. One of the great strengths of our community is the breadth of our collective view of what mathematics is. (Even better: those views are not static. My sense of what mathematics is has changed considerably in the 15 years since my PhD, but that's a story for another time.)

The point of this preamble is to clarify that what follows is a personal view of the purpose of mathematical research, or more precisely one aspect of it. It would be too strong to speak of "the" purpose of mathematics, even using category theorists' "the", so I'll instead describe a purpose of mathematics, focusing on one particular perspective.

Of the many purposes of mathematics, my favorite is to search for the simplest explanation of a particular phenomenon. This often requires inventing a clarifying abstract language to isolate a common pattern from distracting specifics: not necessarily the maximal level of generality but the "right" one. As with mathematical aesthetics, the simplest proof, most clarifying abstraction, and right level of generality are also matters of taste — and that's a good thing! While mathematicians tend to agree about what's true and false, we have countless different opinions about what is intuitive or interesting. (I take a similar relativist view of the foundations of mathematics, preferring to occupy a mathematical multiverse — again a topic for another time.)

To give an example, when I was first learning about sets and functions, it really bothered me that the inverse image preserves both unions and intersections of subsets, while the direct image preserves only unions. The direct image seemed simpler to define, so why was its behavior more complicated? Several years later I learned that the inverse image can be regarded as a functor between powersets, and that functor admits both right and left adjoints, while the direct image only admits a right adjoint. Since left adjoints preserve colimits and right adjoints preserve limits — by a proof that strikes me as straight from the book, expressed exactly at the right level of generality — I now consider my set-theoretic confusion resolved. (Your mileage may vary.)

My mathematical tastes tend towards this sort of "abstract nonsense" (a phrase that, like "queer," is often used affectionately by insiders). Mathematics like this is sometimes called "theory building," whose aim is to make it easier for more people to hold increasingly complicated mathematical thoughts in their heads. (Bill Thurston, in his famous essay "On proof and progress in mathematics," offers manifolds as an example of a unifying definition that makes insights easier to communicate to non-experts.) As Michael Atiyah put it in "How research is carried out":

The aim of theory really is, to a great extent, that of systematically organizing past experience in such a way that the next generation, our students and their students and so on, will be able to absorb the essential aspects in as painless a way as possible, and this is the only way in which you can go on cumulatively building up any kind of scientific activity without eventually coming to a dead end.

It's not always obvious from the outset what the long-term aesthetic value of a newly proposed theory will be. The two contributions that I am most proud of are the theory of ∞-cosmoi, developed with Dominic Verity, and simplicial type theory, introduced in joint

work with Mike Shulman. In both cases, I think our most important contribution is not the specific axiomatization — there are good reasons to tweak the axioms, and others have since extended both theories productively — but the accompanying methodology and proof techniques. (As Tim Gowers notes, theory building and problem solving aren't as far apart as they are sometimes thought to be.)

I think what Atiyah means by "eventually coming to a dead end" is something like the aftermath of Babel, where the frontiers of mathematics have extended so far in so many directions that experts in different fields are no longer able to talk to one another. I am not suggesting that we need to pace the frontier of mathematical discovery to let those attempting to consolidate these findings catch up. But I do think that time spent reaching back to bring others up to the frontier is as valuable as time spent pushing it forward.

I have been working on a metaphor to help describe the mathematical universe to the general public: mathematicians are simultaneously constructing and exploring a mansion so vast that no architect has seen the floor plans and no contractor is directing the work. Instead, individuals and small teams wander down dark corridors to see whether a stuck door can be opened with a bit of force applied in the right place, or with a key someone has just found. Opening a new door is cause for celebration, but the work is just beginning. Someone needs to go inside to reinforce the ceiling and try to turn on a light; someone needs to see whether the door opens onto a closet or an undiscovered wing of the building; and someone needs to relay the message to teams in nearby hallways and on to whoever is drawing the map.

The impetus for that radio conversation and for writing down these thoughts is the rapid advance of AI's mathematical capabilities over the past several months. Our community is now facing the kind of shock that has confronted countless other occupations. But at the same time, as mathematicians, we are part of a very long conversation that has persisted for millennia, through no shortage of upheaval, and the community it has produced is as robust as it has ever been. It's hard to predict what the future might bring, but I plan to keep searching for better ways for us to talk about mathematics with each other.

12 The technical debt of AI-generated mathematics - Henry Cohn - 15th September

In this essay I'd like to discuss the role of AI-generated solutions in mathematics and why many mathematicians are justifiably concerned about a mismatch of values. Let's start with the groundwork, so we're all on the same page regarding what mathematics is and what is needed for progress.

12.1 The role of understanding

Academic mathematics is about understanding, not just about getting answers. Answers can be useful, but they are not the goal or the primary value of this endeavor.

Why do we care so much about understanding? Humans are naturally curious, and understanding how the world works is part of what makes life meaningful, but there's a deeper reason. Thousands of years of intellectual history have shown that understanding and utility are inseparable. You might think you can draw a sharp line separating mathematics from its implications, where you ask mathematicians what you want to know and they tell you the answer you need without having to explain why, but you can't.

This is a universal pattern, not about you personally. Nobody can figure out what they need to know if they are missing crucial understanding. When non-mathematicians formulate a mathematical question, it's not likely to be the question that would actually most help them. When mathematicians propose an answer in isolation, it's also not likely to be what the rest of the world truly needs. This is why applied mathematics often requires deep collaboration between people with different expertise.

Of course we're all finite beings, with limited time and abilities, and we can't understand everything. We have to work together, and even when we do, human knowledge consists of patchwork with enormous gaps. These gaps are unavoidable, and AI knowledge will be no different, because you can't fill infinite space in a finite amount of time.

What gives all of this knowledge value is understanding, not just facts. Understanding lets us make use of knowledge in meaningful ways and extend it more broadly.

12.2 How do we develop understanding?

You can't be taught to understand anything deeply by passively absorbing knowledge. It's just not possible. It can't be done by a skilled teacher, a personal tutor, a book, a video, or any other form of instruction. These can all help guide you in the right direction, but nobody can lead you passively through the process. Ultimately, achieving more than a superficial understanding requires engaging deeply with ideas and exploring them yourself.

This is what math teachers mean when they say math isn't a spectator sport, and what Euclid meant when he apocryphally told King Ptolemy that there is no royal road to geometry. It's why math classes assign problem sets and PhD students write dissertations. Learning facts isn't enough: you have to investigate them actively, question them, reconstruct them yourself, and build your own understanding.

When you make a discovery, the knowledge doesn't transfer seamlessly to everyone else. It's not simply a matter of explaining what you did and letting everyone else read it. It can take a shockingly large amount of work to document your discovery in a form that is truly useful for

others, and even after all that work, the collective effort required for the community to master and internalize it is enormous. If that effort doesn't successfully take place, then the discovery is just words on a page, rather than a more substantial contribution to human knowledge. We often idolize people who make breakthroughs, because they have done something difficult and essential, but it's just one step in a crucial larger process.

Mochizuki's purported proof of the ABC conjecture is an instructive example. He's a brilliant mathematician, and I haven't studied his work myself, so let's give it the benefit of the doubt and assume he really has thought through a valid proof. That would be a magnificent individual achievement, but the broader mathematical community has not been able to fully understand his papers on this topic, he has not been able to bridge this gap in understanding, and as a result his work has not had the impact one would hope for.

In short, writing down a clear and beautifully explained solution is only a small part of the process of advancing human understanding. Writing down an impenetrable solution is barely progress at all, even when done with the best intentions.

12.3 Open problems and AI

Mathematics has many open problems, questions that have been asked but not yet answered. People sometimes imagine that these problems are declarations of failure and requests for help, but that's usually not the case. In many cases, they are questions nobody has tried to answer, or problems where we expect progress can be made.

The reason people curate lists of open problems is as an invitation to others. It can be very useful to have suggestions for fruitful lines of investigation, especially if you are taking up an unfamiliar topic. Providing these suggestions is a service to the community, and it's worth doing even though others might make a breakthrough you had your sights on. Ultimately, we're all playing for the same team.

When a mathematician works on an open problem someone else suggested, they are benefiting from the mathematical community's guidance. In exchange, they do their best to help the community understand and internalize their solution, and they pose further problems when they can.

The net result is that everyone benefits, in a way that transcends competition. For example, Maryna Viazovska won the Fields Medal by solving a problem I had tried to solve for many years. Of course I would have liked to have solved it myself, but her solution greatly enriched human understanding, and it opened up many lines of work I could contribute to. It was obviously a net win for everyone, including me.

It would be unreasonable to say "I hope nobody solves this problem if I can't be the one to do it." However, the primary value of a solution lies in the insight and understanding it provides, not in the mere fact that the problem was solved.

Current AI models are much better at solving mathematical problems than at communicating the solutions well to humans. These might sound like similar skills, but they aren't. AI often produces cryptic, messy, ad hoc, poorly motivated solutions full of elaborate calculations and unnecessary complications. Basically, they often produce math slop. It may be correct, and it may contain genuinely novel and important ideas, but the process of going through it carefully can be awful. Hopefully not as difficult as solving the problem from scratch, but it can be an exceptionally unpleasant and time-consuming process. By comparison, human work often goes through a slop stage early on, but we try hard not to leave it at that.

Part of the reason this work is unpleasant is social: the community gives less credit for following up on or polishing a discovery, even if you understand it better. I expect these norms will evolve over time, perhaps quickly. Part of the reason is psychological: I just don't enjoy reading poorly explained mathematics, even if with considerable effort I can extract something much more appealing, and I'm not the only one. And of course there's the legitimate extra work of carefully examining the slop, which is difficult and time consuming whether or not you enjoy it.

Furthermore, there are coordination issues in how the field should organize this examination process. Traditional refereeing can't handle limitless slop, and chaos is not a viable solution, either. The field's procedures will of course have to evolve, but this will work best if everyone does their part to minimize the communal workload.

I hope future AI models get better at avoiding slop, and I expect they will, but it's unclear whether they will ever be as good at communication as they are at problem solving. Mathematical truth may be easier to train for than effective communication with humans, which is more subtle and less objective.

This communication issue is why many mathematicians are unhappy when someone has AI solve an open problem, dumps a slop proof on the internet, and then abandons it, without any attempt to understand it themselves or communicate it effectively to others. They believe they are making a contribution, but often they are primarily creating unpleasant work for other people.

Of course it doesn't always work this way. I've interacted with a number of non-mathematicians who have done a very responsible job of documenting AI-assisted work in a productive way, and I'm delighted that AI is broadening who is able to contribute to mathematics. Anyone who works to make a meaningful contribution is welcome to participate, and this aspect of AI is unambiguously good. But I've also seen examples on the internet that make me roll my eyes, where AI companies and individuals are loudly taking credit for slop contributions that are genuinely counterproductive for progress in the field.

12.4 Technical debt

The equivalent of this problem in software engineering is well understood; it's called technical debt. Sometimes you can rush to write a program that works, but it's poorly engineered. The architectural decisions are unwise, the documentation is inadequate, and the testing framework is minimal. The code accomplishes the narrowly stated goals you were aiming at, in the sense that it functions, but it's not written in a way that can be extended and built on the way you'd hoped. This technical debt will have to be paid if you want to make further progress, and paying it may take lots of work. You can't necessarily do it by patching up the existing system; instead, parts of it may have to be redesigned and rewritten from scratch.

Math slop causes technical debt in mathematics, and in some ways it's even more insidious. In software engineering, a functioning program with technical debt still works, and deliberately accepting technical debt may be justified if you truly need it to work immediately. In mathematics, my experience is that getting the answer is usually a negligible form of "working." There are cases where humanity can benefit from a factual answer whether or not we understand it. For example, knowing that a cryptosystem is insecure tells us to stop relying on it, and knowing that a counterexample exists keeps us from wasting time looking for a proof. But even in these cases understanding is important, and in most cases it's crucial.

Accruing technical debt hinders understanding. We have seen that understanding always takes effort, for reasons that go beyond slop or technical debt. Compounding that difficulty unnecessarily is a major problem.

To a certain extent technical debt is unavoidable: nobody has the wisdom or foresight to avoid it completely. However, in the past it was kept in check socially, since the people creating the debt were largely the same as those who would have to pay it in the future. You don't want to create future work for yourself and upset your colleagues at the same time, and so you have an incentive not to do so unnecessarily. In contrast, it's a problem if people have a button that creates technical debt without their knowledge.

We live in a world with an infinite supply of slop: anyone can type a prompt and produce output on whatever topic they choose. Pushing the AI button is generally not a significant contribution, since we all have buttons and can push them as we see fit. The contribution comes from what you do with the output. Sometimes it requires little or no additional work to be useful, and sometimes it requires an awful lot.

When someone prompts AI to produce a slop solution to an open problem and declares victory, they are creating more work for other people. It can feel to mathematicians like saying "No, don't spend time on the ideas and directions you feel are likely to be most productive. I want someone to devote time to making sense of this solution to a problem I arbitrarily chose, so I can feel good about having been the one who pushed the button." And it feels even more galling if they are crowing about their great achievement, as if this were the end of the process of understanding rather than the beginning.

This is of course a caricature, and I'm sure no one really means it the way I described. I imagine they expect understanding will follow naturally once an answer has been provided, in which case the work really would be finished. But math slop really is a problem, even if it's factually correct.

Nobody can tell you how you have to do mathematics, but if you just want to push buttons and output slop for the internet, then you need to understand that you are actively gumming up the works. Instead, you should take responsibility for your contributions and work in good faith to help integrate them successfully into human knowledge and understanding. This is admittedly easier said than done, and it's something we all struggle with, but we all need to take responsibility for it.

So what should you do if you aren't sure how to understand your contribution or take responsibility for it? You can try to talk with people in the field and help as best you can, rather than abandoning it. Maybe you can shed light on the process that led to it. If you had access to unusual computational facilities, you can use them to help with understanding it. You can try to follow up and clarify with AI; maybe it will produce more slop, but maybe it will help. There is no universal solution, but the key point is that if this project was worth doing and publicizing in the first place, then it's worth doing responsibly.

13 Fast math/slow math - Ben Antieau - 15th September

To resolve a central tension in the practice of mathematics by academics, I suggest the parallel with fast food and its attendant slow food movement.

13.1 Three hearts

Mathematical work comprises many activities and experiences. Among my favorites are:

(1) Reading a single paper over the course of several months.
(2) The thrill of the chase, when the scent of a theorem, discovery, or the right definition is in the air.
(3) The moment a PhD student shows me their first theorem or paper.

The first and second constitute gaining and creating knowledge. The third is about transmitting meta-knowledge: the ability to perform 1 and 2.

Mathematics for me has always been about the joy of personal understanding and the engagement in bringing others around to that understanding, usually slowly over the course of many months or years.

In recent months, a deep unease has been growing within the mathematical community around the role of language models in our research. The fear is fundamentally economic. LLMs can produce mathematical statements and proofs, some certainly interesting.

Are we ourselves, mathematicians, obsolete? Will our students fail to get jobs? Will our funding agencies, or even our universities, abandon us? Is mathematics in fact solved?

No. Mathematics is not the collection of true statements implying, by complement, the collection of false statements as well. It cannot be solved any more than human experience can be.

13.2 Mathematics and society

Most academic mathematicians are slow math practitioners at heart, even if the pressures of paper-writing, grant-proposing, teaching, and administration seem rather far from our idealized contemplative life.

Yet, math also forms the bedrock of the sciences and engineering and as such is central to how human societies solve technical problems. Mathematicians cannot expect our esteemed place within our culture to survive unchanged if we do not use every tool possible to help it face its grand challenges.

Alexander Nazaryan wrote in the New York Times on 19 June 2025:

"Problems in mathematics take decades or centuries, sometimes, to solve," [Patrick Shafto] said in a recent presentation at DARPA's headquarters on the Exponentiating Mathematics project, which is accepting applications through mid-July. He then shared a slide showing that, in terms of the number of papers published, math had stagnated during the last century while life and technical sciences had exploded. In case the point wasn't clear, the slide's heading drove it home: "Math is slooooowwww. . . . "

We exist, professionally, in part to educate engineers and scientists, to provide researchers to national laboratories, and to provide highly-trained workers to companies. Occasionally,

theoretical results open up new areas of specific application, for example the use of lattice-based methods in post-quantum cryptography standards. If we turn our back on this part of our role, we can certainly say goodbye to our grants and teaching reductions. We can go the way of the Department of —, permitted to exist out of nostalgia, but decreasingly relevant.

To address the central contradiction between our personal practice and our external purpose, I propose fast math and slow math as ways of thinking about different types of mathematics. Just as we all prefer slow food, sometimes fast is what we need, or crave.

These different ways are not meant to separate human mathematicians into "slow math" and "fast math" camps. Indeed, I strongly urge against the temptation to split into us and them. I don't want to sit on hiring committees fractured by debates over AI use or virtue testing instead of honest evaluation of candidates.

The goal of slow math and fast math is the same: to increase human understanding of what is true and false in those regions of "proposition-space" of interest to us. These take their places as different modalities of doing mathematics, alongside computer algebra, exposition, the creation of visualizations, editorial work, and so on. Not every mathematician does, or should do, all of these.

13.3 Dream bigger dreams

Thanks in part to technological advances, but largely driven by an unending thirst to know, physicists have gone in the span of a couple of centuries from tabletop experiments to running some of the largest collective endeavors of humankind.

LLMs represent a tremendous technological advance and some might choose to dream bigger dreams. We should deliberately and thoughtfully use these models and other AI tools to have those dreams. Some of these might be to answer new, pressing questions created by the very existence of the tools themselves, like the problems of interpretability in LLM systems. These tools were built with mathematics and our discipline provides one of the principal means by which we understand them and use them safely.

Other dreams should be to undertake audacious programs, which would not have been practical in the past. I have in mind research directions at the scale of the Langlands program or Grothendieck's dream of motives.

One example could be higher-dimensional algebraic geometry. Many counterexamples that have been established with LLM assistance have to do with the fact that we do not know the zoo of algebraic varieties in higher dimension the way we do curves and surfaces. (Although, of course, there have been many, many amazing results, like Beauville–Bogomolov.)

Such programs should be pursued rigorously with a back-and-forth between humans and LLMs. These programs should have directors, postdocs, students, standards. The goal would be to leverage LLMs to understand high-dimensional phenomena — for humans to understand. It is not to produce a slurry of results no human has read or understood.

These programs are fast math: far-reaching, ambitious, almost reckless. They would aim at transformative new discoveries and organizations of mathematical knowledge, perhaps particularly motivated by grand challenges at the boundary of mathematics and applied mathematics. But they would not end with a given discovery. They would also be responsible for producing lecture notes, textbooks, databases, and other tools to aid human apprehension

of their discoveries. They would exist to truly harness AI capabilities and prepare the outputs for human understanding.

13.4 Collective standards

I suggest the following standards for the mathematical community. Please consider these as a first proposal and feel free to send me feedback. The overall goal is that we should use LLMs to aid in human mathematicians' understanding, not merely to increase our production.

(1) *Disclose all collaborations, including with LLMs.* It is my hope that eventually we can use open source models for much of the day-to-day work that people might turn to LLMs right now. We already cite computer algebra systems as soon as the work goes beyond the totally routine. We should do so here.
(2) *Do not judge mathematicians for their LLM use.* The standards of our community are still in flux. We can fight over what those standards should be, but individuals have to act in the meantime and we should not litigate their behavior.
(3) *Posting a paper to the arXiv or publishing it means three criteria have been met by the authors:*
    - you could reproduce the main ideas if on a desert island (excepting large computer computations);
    - you have written it for human understanding, to the best of your ability;
    - you take intellectual and reputational responsibility and credit for the results.

    In the second point, using LLMs for polishing writing, spell-checking, finding inaccuracies, or creating diagrams is completely fine. Using an LLM to generate a first draft of a paper is not acceptable.
(4) *LLMs should not be used to compete with other mathematicians, to scoop them or otherwise embarrass them.* It goes without saying that we are somewhat competitive, but there is nothing to be gained from this behavior and a lot to lose in terms of collegiality. We are extremely lucky to be in a field where we talk openly at conferences not only about what we've done but what we are working on, sometimes years before those results see the light of day.
(5) *Refrain from asking every random question to LLMs, or at least from making the answers public.* Simple results are important for training students. Many have been excited to ask questions long-abandoned in their research programs, including myself. Of course, we cannot say that any given question is off-limits. But I expect that this fervor will die down in the coming weeks or months. And we do not need to repeat it with every new "release" from the big labs.
(6) *Focus, if you choose to use LLMs, on developing your own research, intuition, and results.* No one owns an area of mathematics or a certain problem. But it is not acceptable for me to wake up one day and say, you know I know famous mathematician — is working on problem — and I'm just going to solve it. They will be impressed with me, or I will get that next job I need. The easiest way to guard against this behavior in the long run is to just not reward it.
(7) **Do not work faster than you can understand or effectively communicate.** This golden rule informs all of the others.
(8) *Do not recruit, hire, promote, or praise based on the number of papers.* In the past, the production of high-quality papers has provided a proxy for evaluating understanding and brilliance. Of course, you want to hire the individuals who've written these papers

to access their insight for your own research or for the sake of the students in your department's charge. Each department will have to find new ways of judging.

(9) *Do not delegate the previous task entirely to LLMs.*

Before LLMs, I typically had 5-10 projects on the burners or in the fridge at any one time, probably working actively on 2-3 at any one time. This reflects my interests and personal capability to pay attention to them. Having used language models, specifically agentic coding systems (basically nonstop for coding for 7 months), I can attest that there is a tendency to do more and more and more, but that the pleasure in doing so drops, the quality is hard to verify, and my understanding of the output goes down as well. As such, I strongly disagree with the premise of Shafto's slide that stagnation in mathematics is measurable by slower growth in publication rates than in other sciences. Papers are only a proxy for understanding.

There are also areas where I believe it is explicitly permitted to use LLMs. They can provide translation services, whether to improve writing in a non-native tongue, or to translate into one's own language. They can proofread. Let's face it, no one is doing this for us anymore at most journals. They can act as tutors.

Experiments will inevitably produce writing largely generated by an LLM and possibly understood by no human. These texts deserve a home too, because they might later be useful for human work, but that home is not the arXiv.

13.5 Pascal

We have known since von Neumann, Turing, and Gödel in the 1930s and 40s that mathematical theorems are nothing but provable propositions in a formal language. These could be produced automatically by any computer.

Humans care about individual statements and theorems. Perhaps in part because of our intuition from the natural world, these propositions are not equal in our eyes. Mathematics is what human mathematicians understand. Beyond is darkness: true absence, or inscrutable trophies of alien intelligence.

In Pascal's Sphere, Borges invokes Blaise Pascal: "nature is a fearful sphere whose center is everywhere and circumference nowhere" to propose that "universal history is the history of the different intonations given a handful of metaphors." Indeed, mathematics is an infinite sphere whose center is everywhere and circumference nowhere.

Like Pascal, overwhelmed by humanity's simultaneous insignificance and ability to understand our insignificance, mathematicians work in a habitable zone in which the theorems, questions, and conjectures are configured so that some structure is visible.

The fact that LLMs can seemingly be organized into their own societies is interesting. The mathematics that they produce there is not interesting to me. Language models have not solved mathematics for humans. They are practicing mathematics for a nascent civilization.

13.6 Closing

I believe civilizations beyond ours exist, have existed, and will exist in the future. Our theorems have already been proved. They will be forgotten, and proved again and again until the last ember goes out. The exercise of mathematics is the work of civilization, and we will continue.

14 Open letter to the Royal Society - Ben Green - 16th September

The rapid development of AI is of enormous public interest right now, with the possible existential risk to humanity being reported by mainstream news organisations. However, the discussion is often tempered with questions such as 'haven't we heard this before?' or 'isn't this just hype from the AI companies or a desire to protect their interests?'. As discussed widely in this blog and elsewhere, we as mathematicians (perhaps more than other scientists) have a particularly close up view of the speed of development and many of us have concluded that we ought to be very worried. We should make this point loudly. 42 mathematician Fellows of the Royal Society have today written to the president of the society, Sir Paul Nurse, to put forward this view.

Mathematicians of all types are invited to support the letter, their names are displayed here. If you want to add your name, complete this form. The letter is reproduced below.

14.1 The letter

We, as mathematicians, write to express our extreme concern about the pace of development of AI and the implications of this for humanity and civilisation. None of us have any significant involvement with AI companies.

In the last 3 months we have seen the leading models of OpenAI and Anthropic develop from around the level of a strong student to having solved multiple open questions at the forefront of research, including one of the 7 Millennium problems. These models are now operating at the level of the top human mathematicians in many parts of the subject and we must assume there is a significant chance of them developing superhuman abilities within a similarly short timeframe.

This has profound implications for mathematics, but highlighting that is not the purpose of our letter. It is highly likely that these systems are comparably strong in other technical domains such as (for example) cyber security, autonomous weapon control, development of biological and chemical agents and the targeted spread of misinformation. Moreover we must assume that the rate of change of their capabilities will be similarly rapid in these areas.

We have all seen the concerns raised by former employees of the large AI companies, estimating the probability of human extinction to be as high as 10 percent over the next decade. Our witnessing of the rise in capabilities of these models in our domain of expertise persuades us that such statements must not be dismissed as 'hype'.

By the time the situation becomes obvious to the wider public it may be too late to act. We believe this is an emergency, and call on the Royal Society to use its influence to convey this view to government and the media.

15 Let the Diners Into the Kitchen - Noah Giansiracusa - 16th September

Imagine a tech company announces new robot workers that can prepare food better than human chefs. If you think of dining out as sitting at a table and having your meal brought out to you, this sounds like an excellent development. Let's replace the humans in the kitchen with these bots and enjoy higher quality food!

But what if the restaurant's patrons are invited inside the kitchen to see how the food is made and are explained what's involved. They learn how techniques from around the globe have been passed down for generations. How cooking is both an art and a science, constrained by rules and patterns but rich with creativity at its heart. How ingredients are thoughtfully procured from farms and distributors with an eye toward locality, seasonality, and sustainability—ensuring long-term production is not sacrificed for short-term convenience. How menus are crafted to balance familiarity with ingenuity. How young chefs learn from experienced ones by working alongside them. How food is among the most timeless, universal ways humans have for connecting with each other.

What if, in other words, dining out is more like watching an episode of the hit TV show *The Bear*? I think it's fair to say that this is a dining experience the public would think twice about giving up to automation.

Unfortunately, most people outside of math think all we do is serve up problem solutions on a platter. They don't know that math is about human understanding, not just mechanically establishing truths. That it is one of the longest running shared stories in human history, one that has transformed society and is treasured by its adherents. That it is a precious social endeavor, not a speedbump on the road to lucrative tech IPOs.

That's because we haven't done enough to bring the public inside our mathematical kitchens. Perhaps we haven't even put enough effort into sharing the human side of math with each other—at least, not publicly.

Ask yourself if the following scenario rings familiar, not necessarily in its details but in its broad contours. While picking your kid up from an after-school activity, you learn that one of the other parents is also a mathematician. What starts as friendly banter develops into a research collaboration. You're in different fields but have found a topic of mutual interest and realize that by combining your backgrounds you can provide a new perspective on an old problem. Before long, you have an innovative joint result together.

When chatting with colleagues, you eagerly share the amusing origin story to this theorem. When giving seminar talks, you downplay the serendipity but still gesture at it. Then in your formal write-up—that is to say, in the research paper, the one truly public record of this result—the human backstory is not mentioned. All that readers see is the theorem and proof, whose ideas are stripped down to a series of checkable lemmas and calculations, preceded by an introduction that aims to convince an anonymous referee the problem is important.

You've written for precision, clarity, and efficiency, maximizing the likelihood that the paper is accepted into a research journal. In doing, so you've minimized the humanity of your work. You've kept your reader out of the kitchen.

To reassert our role in math and reclaim the human value of math while Silicon Valley threatens to take these away from us, I believe we must radically embrace and expose the

human side of math. We must do so now, before it is too late. This can take a variety of forms. I suggest a few here and leave you to surface others.

- **Blogging.** This is a great way to share behind-the-scenes aspects of math by letting readers into your own private kitchen, so to speak. You can discuss what's happening in math, share your perspective on others' results, tell the story behind you own results, reflect on your life as a mathematician, and much more. There are many excellent blogs by mathematicians already, but there's always room for more. If you're not up for starting one, visit others and leave comments. The more we publicly engage each other, the more society will see that mathematics is a living, breathing community.
- **Write op-eds.** This isn't for everyone, but it's worth considering. It is a powerful way to speak directly to people outside of mathematics and show them that we matter. The best opportunity here is when there's an event in the news that you could use your math background to help unpack for audiences. Don't just target big name national newspapers; local or special interest outlets have real impact too and might especially appreciate the expertise you provide. I am happy to offer op-ed guidance and feedback—please don't hesitate to contact me!
- **Social media.** I say this with a slight shudder, but social media is a convenient and productive way of building community and showing the world who we are and what we do. Keep in mind there are many platforms and content formats, and that audiences generally like accessibility and relatability. That means a post doesn't need to be a polished piece of writing; it could be a video of you walking your dog in the afternoon while talking about your favorite math moment from your day.
- **Rethink the research paper.** This suggestion might push the envelope a bit further, but here goes. Since research papers are the standard currency in math, a currency that is being rapidly devalued by AI, perhaps we should find ways to inject more humanity into them. For instance, we could add a "backstory" section to our papers explaining the human elements that went into the collaboration/project/result. Or a "digestion" suggestion offering an informal human-centric interpretation of the main result(s). I don't know, I'm just thinking out loud here. But a useful guiding principle could be to think of what you can provide that an AI cannot, or would not.

The Bourbaki movement focused attention away from informality in math, training us to be clear and concise. We are in a different era now. Math is not adapting to confront imprecision and gaps in arguments; it is adapting to confront AI. To put the matter bluntly: The more we act and write like robots, the more we hand our profession over to them. The details of how best to resist this are not obvious, but the overarching strategy is. We must show the world the human side of math. We must let people into our mathematical kitchens.

16 Becoming a benchmark - Talia Ringer - 17th September

When I was wrapping up graduate school in computer science in Spring 2021, I was given access to a curious little programming model on OpenAI's "playground." The model, called Codex, took in natural language text and generated programs from that text. The task of automatically generating programs given an expression of programmer intent had been one of many major questions in my field of study—programming languages—for decades. This was the first time I had seen a neural model, with little to no input from anyone in my field, actually succeed at that task to any degree. This was more than a year before ChatGPT was released, and yet I knew that everything was about to change.

On one hand, I was glad that programming was becoming more accessible. My mom came to visit shortly after that, and I pulled out my laptop and told her she could now program. I had her make a little video game. It was not perfect, sure, but my mom was programming, and that was wild. I had always wanted programming to become accessible to everyone.

On the other hand, I had so *many worries*. At a technical level, would the software of the future be full of AI-introduced bugs? Would people run a bunch of unit tests on AI-produced programs and think that they are OK, but miss out on edge cases? Would AI tools overfit to the tests they are given access to? Would people sometimes feel more productive using these tools, but actually get less done? (Yes, yes, yes, and yes.)

Those technical worries were overshadowed by a larger existential dread. My specific focus within programming languages research had been on using classic programming languages techniques to make it easier to write formal, machine-checkable proofs using proof assistants like Rocq and Lean. Was my entire field of research about to die? To be swallowed by AI? I was not worried about my field *actually* being fully "solved," but I was very worried about the prospect of AI researchers *claiming* to fully solve my research area, and of the general public actually believing them. I had seen this play out before in linguistics and natural language processing.

Worse, if AI swallowed my field, would AI's culture leak into my field's culture? I had known AI's culture to involve all sorts of things I find distasteful and immoral in research, like "scooping," competition, and secrecy. My field, by contrast, was (and thankfully still is) a lot more like mathematics in this regard. We value communication, collaboration, and openness. I was scared that AI would rot this culture from without.

I have come to understand what I went through as the AI grief cycle, the one that starts when one's life's work becomes a benchmark for AI companies. And as I worked through this cycle of grief, I realized that I had to communicate both to AI companies and to the general public that my work will not be replaced by these tools, but that it will change. I had to understand that myself, come to terms with it, and move with it in my own work. And above all, I had to make sure incentives stayed aligned with that reality.

This started with communicating what I actually do, both to the AI community and to those outside of it who set incentives. This was the most exhausting part of it, since I was coming from a small research community, while the AI community is large and powerful. So that means I really had to immerse myself and learn to speak their language. In lieu of that, it would have been too difficult to honestly and reputably assess the limitations and impacts of what AI tools do. (In doing so, I accidentally nerd-sniped myself into actually doing AI work as part of my broader research portfolio, but it is probably possible and fine to do this in a way that still keeps one's research far away from AI.)

On an individual level, funders and the AI community alike have since come to respect me. And also, my field of programming languages has come out pretty OK so far. I will never know how much of an impact I have had on that. But I have found that the whole community has reacted in ways that are pretty aligned with what I have done. We have engaged honestly with the changes, and we really have assessed both the capabilities and limitations of the tools that had infringed on our field so suddenly. When relevant, now, we use neural techniques to improve our own work. But we have also found ways that our techniques are strictly complementary to those techniques, and we have figured out how to communicate that. Our culture has come out intact, too.

But I think one thing we have going for ourselves in programming languages is that basically nobody has ever heard of our field, and most people do not care about what we do. People do care a lot about math. We don't even get the "oh, I *hate* math" response that mathematicians get; hatred means that people care.

Culturally, at least in the US, math is simultaneously revered and hated. Peak "intelligence" is often culturally associated with mathematical ability—people even bring up Terry Tao as an example of this. Poor performance at math, or anxiety around such poor performance, is often coupled with statements about not being "smart enough" to do math. Mathematicians, then, come to represent the intellectual elite, with all of the scapegoating such a label carries. And leaders of AI companies come to believe that if they can "solve math," they can "solve everything," whatever that means. (Sorry, Jesse; we are still friends.)

Because of this dual reverence and hatred, people are paying way more attention to AI infringing on math than they did to AI infringing on my field. This is good and bad. It does mean that mathematicians' statements are getting lots of coverage; they have a real chance to communicate to the entire world what it is they actually do. But it also means that their legitimate sour gripes are being misread as sour grapes. There is a perception that they are gatekeeping. Just like, in 2021, if people had paid this much attention to my field, they might have wrongly concluded that I was just being bitter because I did not want my mom to be able to program. (I did!)

This makes mathematicians' jobs harder than mine was. Their audience can at times be actively adversarial. Just learning the language of AI folks will not be enough.

Still, I think mathematicians are on the right track. The work many are already doing of addressing the public is even more important than addressing the AI industry. Yes, the AI industry's goals are misaligned with those of the math community. But it is actually OK if that remains true, so long as the rest of society recognizes that misalignment and continues to value and incentivize the kind of work that mathematicians value. AI companies will always want to use whatever field is hot at the time to show that their models are the "best." There are ways to help them better understand what "best" should mean, but it's also pretty OK if they never do understand that, as long as the general public does. So math—the process, not the benchmark—will need a PR campaign that lasts for as long as math the benchmark is relevant to AI companies. That means deliberate and consistent engagement with news outlets, social media, education systems, policymakers, and funders.

This work of engagement is exhausting and unending, but it does get easier as society and AI companies alike move on. Remember when art and writing were the main targets of these companies? Two things seem to have happened: First, society seems to have collectively developed a distaste for AI art and writing, and even for human art and writing that vaguely resembles AI art and writing. Second, AI companies seem to have reached the point at which

showing off their tools' art and writing results no longer proves them to have the "best models," so they have largely moved on to other fields.

Have art and writing been negatively impacted? Absolutely. And the fight is still ongoing. But at least the fight is no longer all-consuming. (Collective action like unionizing, or like the Hollywood writer's strike, might also have to do with that, though. I do think mathematicians should consider unionizing.)

And what if mathematicians actually *do* want AI tools that help with math the process, and not just math the benchmark? (I do. I'm super excited about what math could be like in such a world.) In that case, I think it is probably best to look for collaborations with smaller companies and academics, especially those that have a participatory model where mathematicians get to co-design the tools to reflect their own values and use-cases. (I have one such collaboration with Emily Riehl already, and I am hungry for more!) It also helps to be upfront about expectations around credit, especially where the culture might clash.

In any case, a friend who is a professor of linguistics told me in 2021 that, in ten years' time, my research would be different in ways I could not possibly predict ahead of time. And that still, it would be OK. I took great comfort in that. It really will be OK.

17 Why I didn't sign the Fields medallists' letter - Timothy Gowers - 17th September

When I was around 11 I heard for the first time about Fermat's Last Theorem. I was immediately captivated by the problem statement, as well as by the accompanying story, and made a fairly serious attempt to prove it. And while, unsurprisingly, I failed, I learned a lot from the attempt. Blissfully ignorant of the fact that the $n = 3$ case had been proved by Euler over 200 years earlier, I decided that that would be a good place to start: once I had sorted that out, I was optimistic that I would be ready to tackle the general case.

Since I still couldn't really see where to start, I decided to simplify the problem further and concentrate on successive differences of cubes, with a view to showing that such a difference could not itself be a cube. At the time I did not know how to express what I was doing in algebraic language, so I did not explicitly try to prove that the Diophantine equation $3n^2 + 3n + 1 = m^3$ had no solution. Rather, I just worked out some successive differences and stared at them, trying to get some idea of why none of them was a perfect cube. (I should be clear that this story is a reconstruction of what I think probably happened given the few memory traces that remain half a century later rather than a completely reliable account.) At some point, I had the idea of taking the difference sequence of the difference sequence, and discovered that it formed an arithmetic progression. That felt like progress, so I investigated difference sequences a bit more and discovered, purely empirically, the rule that if you start with nth powers and keep taking successive differences, then eventually you get to the constant sequence $n!, n!, n!, \dots$.

Somehow I never managed to turn this observation into a proof of Fermat's Last Theorem, and later on my dream of solving it got replaced by other mathematical dreams. However, when I reached the point in my mathematical education where I was taught about taking difference sequences and about what happened to polynomials, I understood those topics much better than I would have if I had not discovered difference sequences for myself and spent happy hours playing around with them. I mention this story just as an illustration of the phenomenon that was strongly emphasized in this letter signed by 25 Fields medallists, that one learns a lot from thinking about a problem, regardless of whether one solves it.

In the end, however, I felt that I could not sign the letter, despite agreeing with much of what it said. Instead, it seemed better to do what I did with the Leiden Declaration and set out my own position in a blog post. But it should be understood that by doing that I am not setting myself up as a member of some opposing camp: indeed one of my worries at the moment is that the mathematical community might become bitterly divided, something I would very much like to avoid. Also, I agree on the fundamental point that we are facing a crisis: I just want to offer a slightly different analysis of what that crisis is. I don't claim full originality for this analysis, as I know that several other mathematicians have already put forward thoughts that are similar to the ones I have, though (for what it's worth) I have largely come to these conclusions independently.

On the subject of independence, it will perhaps help if I clarify that while I have contacts in the mathematics group at OpenAI, and have also been given early access to some of their models (typically only a few days before they have been released), and have been given free access to their Pro models once released, I have never been paid by OpenAI. I mention this in the hope, perhaps naive, that what I write will not be dismissed for ad hominem reasons. Another potential reason for my being regarded as "pro-AI" is that, as I have stated publicly

several times, I have a group in Cambridge devoted to automatic theorem proving. However, that is actually more of a reason to be anti-AI, since our group has been trying to attack the problem of getting computers to prove interesting theorems by understanding as well as possible how humans prove interesting theorems, so now that LLMs can clearly do it without the help of such insights as we have had, one of the main motivations for our work has disappeared. To put it another way, we have had to swallow the bitter lesson (which of course we were always aware was a distinct possibility, even if the speed at which it happened has taken us by surprise). I do in fact think that it is still a very interesting and valuable intellectual exercise to try to gain this understanding, even if we can use LLMs as black boxes, but that's a topic for another blog post.

So why didn't I sign the letter? Let me extract a couple of sentences from it that express what I see as the principal argument being put forward.

But solving problems is only a tool and proxy for achieving the primary goal of conceptual understanding and insight. Forgetting this in the world of AI may turn the tool against the primary goal. Indeed, the mass production at faster and faster pace of "true/false" statements could destroy fertile ground instead of breathing life into new ideas.

Perhaps the main reason I didn't sign is that I don't fully subscribe to this view. Instead, I have a more complicated view, which I actually expressed in my essay The Two Cultures of Mathematics a quarter of a century ago, and which can be summarized by saying that there is a spectrum of attitudes in mathematics to the relationship between problem-solving and conceptual understanding. At one end of the spectrum you have mathematicians who are primarily motivated by the wish to solve problems, who see conceptual understanding as a very important means to that end. At the other you have mathematicians who are primarily motivated by the wish to attain conceptual understanding, who see problem-solving as a very important means to that end. I worry that the severe-misalignment letter could be seen as saying that the "right" attitude is to focus on conceptual understanding as the main priority — indeed, the above sentences say that more or less directly. But I think that there are mathematicians all across the spectrum, and that that is a good thing (or perhaps I should say that it has been a good thing up to now — the future is much less certain), and I don't want to suggest to a large fraction of mathematicians, including myself, that their mathematical temperament is somehow "wrong".

My own particular mathematical attitude is very similar to one that was beautifully articulated in a Twitter post by Jacob Tsimerman (another non-signatory of the letter), which, now that I look at it, says a lot of what I will be saying here. And that post in turn is a response to Daniel Litt, who is in my opinion one of the wisest commentators on mathematics and AI. His views are expressed in a later post here, which I deliberately didn't read until finishing this one, and then found, as I expected, that there was significant overlap. I would also like to take this opportunity to recommend an excellent post by Noah Smith entitled The End of the Age of Heroes, in case you haven't read it.

I have been talking so far about individual mathematical understanding, but I suspect that what concerns most of the signatories is less that than the collective understanding that results at least in part from the human activity of problem solving. My guess is that they would argue, completely coherently, that even if collective understanding is the primary goal, if many individuals are primarily motivated by the wish to solve problems, that's absolutely fine and contributes to that collective understanding.

With that interpretation, the issue becomes slightly different: is it more important that the collective understanding of the mathematical community should be as advanced as possible or that there should be answers to as many problems as possible? Or are those two aims valuable in different ways, so that there is no point in declaring one of them more important? Or are they so inextricably linked that it makes no sense to argue that one is more important than the other? And when we say "important", for whom are we saying it is important: for mathematicians, or for society as a whole?

I find these hard questions, so I don't want just to declare an answer to them. (Do you see what I did there?) Instead, I'd like to try to offer at least some argument for any conclusions I come to, even if they are tentative. So let's compare two scenarios. In the first, which I think is the more likely actually to happen, models become publicly available that are better at solving problems than virtually all mathematicians. If there are a few residual mathematicians who can do things the models can't, even they work far faster if they make heavy use of the models. Thanks to this, in a short time we get answers to many questions that we have deeply cared about, but the rate at which we receive these answers far exceeds the rate at which the mathematical community can absorb them. In particular, most of the answers are obtained with zero effort from human mathematicians — just prompts such as "Thank you — please continue".

In the second scenario, there has been an international agreement, for entirely other reasons, to block the public release of models significantly more powerful than the ones we currently have, and the mathematicians within the tech companies agree to hold off from getting their internal models to solve major problems. Instead, they take guidance from the mathematical community, solving problems only when asked to do so by some suitably representative body that decides that the benefit of receiving a solution of a certain problem outweighs the benefits of humans struggling to solve it over a much longer timescale.

I'd like to consider what the difference would be between these two scenarios both for individual and collective understanding. I'll begin with individual understanding.

One might argue that for individual understanding, not too much would change if we are suddenly flooded with large numbers of big new results. There is already far more mathematics out there than I have any hope of understanding (for example, despite being fascinated when Fermat's Last Theorem was proved, I have made no attempt to understand the proof), and even among the parts that I do understand, the parts that I understand because I myself discovered them form a very small fraction, though a fraction that I understand more deeply than anything else (at least temporarily — after a while I forget things and lose quite a lot of the understanding I built up). However, one change, which seems positive, from the perspective of the building up of individual understanding, would be that we would have a much bigger choice of results that we could choose to study. Also, if we found ourselves stuck on some point, AI would be able to help us. The main likely negative change is that we would probably cease to exercise that part of our brains that we use when spending months or years struggling with a difficult research problem, which can be hugely helpful in developing understanding.

I say "likely" because in principle there would be nothing to stop us thinking about very hard problems without consulting LLMs, but in practice it seems unlikely that people would put in the same level of effort that they do now. The situation might a bit like what happened with satnavs, where one could always decide not to use them, to keep the part of the brain active that can look at a map, learn a route, and follow it, but in practice most people succumb

to the temptation to use a satnav. (In fact, I myself do try to keep that part of my brain active, and was rather proud of finding my way somewhere recently when I had briefly looked up the route on my phone but then forgotten to bring the phone with me when I actually went there.) But even if all we were doing was reading AI output, I think that the problem-solving muscles in the brain wouldn't atrophy completely. When students are reading maths papers, I strongly advise them (and I think this is pretty standard advice) to read "actively" rather than "passively", doing things like trying to prove the result for yourself, looking at the paper only when you feel stuck and need a hint, and even then just trying to get the hint and as little extra as possible. If one reads a paper that way, then one is constantly solving problems, some just exercises and some quite a bit harder. It seems likely that an LLM could get to know what our mathematical background is and feed us with just the right hints to allow us to work our way through a mathematics paper in this active way. Yes, we would lose the particularly deep level of understanding and ownership that comes with having solved a hard problem oneself, but it isn't clear to me that progress in mathematics would suffer as a result. I would be very interested to hear counterarguments to precisely this point. That is, I would be interested to know what use that level of deep involvement with a proof might have in a world where AI is much better than we are at finding proofs.

How about collective understanding? Let me quote a bit more of the letter.

Indeed, the mass production at faster and faster pace of "true/false" statements could destroy fertile ground instead of breathing life into new ideas.

Often these solutions are announced in a rush, leaving no time for a proper writeup, the isolation of new methods and ideas, and citing relevant previous work of others. As in all creative professions, this raises severe attribution and plagiarism questions. Moreover, without the willing mathematicians who must take care of their development and integration into the mathematical canon, AI-conceived ideas would never become fully alive and the crucial human transmission chain between mathematicians would be lost.

I'll come back to questions about proper citation and focus on what I take as the core worry here: that if results are proved too quickly, then the digestion process will become impossible. I am definitely worried that results will not be properly digested, but for different reasons.

A first remark is that what AI is producing is not just true/false statements: we now know not just that the Navier-Stokes equation with smooth forcing admits finite-time blow-up, but we have a proof of that, which builds on a great deal of wonderful work done by human mathematicians. Many people used to express the worry that AI would solve our favourite problems with utterly opaque proofs, but that has not turned out to be the case, even if their write-ups often leave plenty to be desired. (Incidentally, I see these inadequate write-ups as almost certainly a temporary annoyance and therefore not as a fundamental threat to mathematical practice or future mathematical understanding.)

Secondly, even if the volume of new results is large, mathematics is a highly specialized discipline, so mathematicians can work in parallel. If, for example, we had to digest 1000 important results in a year that were roughly uniformly distributed across mathematics, then most sub-communities of mathematicians would probably want to understand around 30 of them, and for each individual problem there might well be only a small handful of specialists who would be obvious people to take the lead in reaching this understanding, with that handful varying from problem to problem. So it would be a big task, but not necessarily an impossible one.

In this context, it is worth thinking about the huge volume of output of human mathematicians, which seems to have been increasing recently, even before AI. While I have sometimes heard complaints about this, I have certainly not heard suggestions that human mathematicians should slow down the rate at which they prove interesting theorems. That may be partly because the authors of those theorems take the trouble to write their papers well and give good talks. But what about the large quantity of papers, including important ones, that are not written well and whose authors give incomprehensible talks? That can be annoying, but it is a familiar annoyance and not one that we think of as a crisis.

A third point is that even if the volume of AI output is too big for us to be able to digest it properly, that is not necessarily a bad thing. To draw an imperfect analogy, there is now more content available on streaming services than anyone could possibly watch, with the result that there is almost certainly some very good content out there that is hardly watched at all. But that isn't obviously a worse situation than if there were far less content and all of it received the attention it deserved. Returning to mathematics, if there were too much AI-generated content for us to be able to digest it, then we could choose which parts of it we wanted to digest.

For that we would need to have some idea what was there (a situation a little similar to how human mathematicians typically learn quite a lot about what results are known in their area even when they do not understand their proofs in any detail). One way one could try to achieve that would be to create a well-designed database, probably with AI help. But perhaps that would be unnecessary, and instead one could simply talk to an LLM and ask it to give a bird's-eye view of whatever area of mathematics one wanted to understand in that knowing-what's-there way.

The fear seems to be that some very interesting and important parts of mathematics will be discovered by AI and then overlooked, when had they been discovered by human mathematicians they would not have been overlooked. And that may even be the case, but what matters is whether the amount of interesting and important mathematics discovered by AI that is not overlooked will exceed the amount of interesting and important mathematics that would have been discovered and properly digested by humans with AI having played a more modest role.

In short, it seems to me that while a flood of "big" AI results would be likely to increase the amount of important mathematics that was not properly digested, it would also be likely to increase the amount that was properly digested, which seems like a pretty good bargain.

Let me quickly discuss the problem of AI not properly crediting human mathematicians. I agree that this is a serious problem right now, but it is another problem that I see as temporary. Very soon, the whole "credit system" will surely collapse, since finding an amazing proof will be no more of an intellectual achievement than when a citizen scientist spots through their telescope an object that turns out to be a new comet. Until that happens, it is important to give humans the credit they deserve, since careers can depend on it, but that will soon cease to be the case as well. I have to say that I'm puzzled that this problem exists, since I would have thought that if you asked an LLM to look at a proof and tell you which ideas in it are close to ideas that are in the literature already, it would be extremely good at that task. I hope the answer to this conundrum is not that people have been in such a hurry that they have simply not taken the trouble to do this, but I fear that it might be, at least in some cases. If so, then those who have been careless deserve to be criticized, but it is a minor

matter compared with the survival of mathematics, especially if the lack of citations is swiftly put right.

Does all this mean that I am optimistic that mathematicians will end up digesting at least as much mathematics in a post-AI world as it would have if AI had not been able to prove major theorems? Not exactly. But my worry is not that we would be unable to do it, but rather that the social structures that currently support this digestion process will be destroyed and not adequately replaced.

One way that might happen is that AI disrupts society so much, or even kills vast numbers of us, that the preservation of something like the current mathematical tradition ceases to be of any concern: all that will matter is the survival of the human race. But that again is a topic for a different blog post (which in fact I am in the middle of writing).

Let's assume instead that we get lucky and that AI remains more or less under control. My worry then is that we do not manage to transmit what we know to a new generation of mathematicians. Speaking for myself, my main motivation for becoming a mathematician was the dream that I would solve unsolved problems — the more famous the better. I have also always greatly preferred directly thinking about a problem to reading books and papers and generally learning the mathematics of other people. (I'm not saying that's good, but just stating a fact about myself.) If the dream of solving a famous problem had not existed, I'm not sure whether I would have become a mathematician. I don't completely rule it out: maybe what really motivated me was that I had an aptitude for the subject and that solving problems was a way of getting respect from a small group of peers. And maybe I could have tried to gain that respect in a different way, such as thinking very hard about an area of mathematics until I was able to demonstrate to others just how well I understood it. But I'm not sure how motivating that would have been for me. I very much hope that there is a pool of young people for whom it will be a powerful motivation, because I think the survival of a human mathematical tradition may well depend on it.

Thus, the primary risk, as I see it, is that a lot of people who would have done a PhD in mathematics and gone on to become custodians of the mathematical tradition will no longer wish to do so. Those of us who have PhD students, including me, need to try as hard as we can to come up with imaginative ways for them to use their time productively (in consultation with the students themselves, obviously). Whether or not we do a good job with that could make a huge difference to the future of mathematics. A related risk is that the perception among policy-makers will be that mathematicians are no longer needed and that funding will become much harder to come by: we urgently need to come up with good ways of explaining the value of having a large pool of human mathematical experts, even if it is no longer part of their role to find new proofs of theorems.

A final reason that I didn't sign the letter is that I wasn't really sure what it was demanding that isn't happening already. It seems likely that in a matter of not very many months LLMs will be released that are able to solve major mathematical problems, and they will presumably have no trouble at all with more run-of-the-mill problems. However much we might regret that, there is no chance that the impact of such models on mathematics will persuade AI companies to stop their release, though perhaps concerns about safety will lead to some delay and give us a bit more time to work out how to adapt. Assuming that they are released, there will be a flood of new results, whether we like it or not, and it will no longer be the AI companies producing them, though perhaps the pattern will continue that the AI companies will have access to more powerful models and so will obtain more than their fair share of

headline results. So I felt that there was nothing to be gained from criticizing AI companies for generating too many solutions too quickly. In fact, it may well be that all that does is bring forward by a couple of months what was going to happen anyway, and perhaps it will even allow the results to be released in a more controlled way than they would have been if they had been discovered by random people once the models were publicly available. Under the circumstances, I think the best we can do is recognise the changes that are coming and try to work out the least unsatisfactory way of dealing with them.

18 A CERN for AI-assisted science? - Dimitris Koukoulopoulos - 17th September

At the 2026 World Cup semifinal, England led Argentina 1-0 with less than twenty minutes left. And then the momentum completely changed: England switched to a much more defensive formation, withdrawing attacking players and retreating deeper. The Argentinians started attacking in waves. They equalized in the 85th minute and scored the winner in stoppage time.

England's defeat offers a valuable lesson: when trying to avoid the worst outcome, fear and passivity are not your friends.

The recent announcement by OpenAI of a solution to the Navier–Stokes Millennium Prize problem, and the controversy surrounding the way it was obtained and released, should be a wake-up call for the scientific community. We need to come together quickly and develop an institutional response to this new reality. Complacency and defeatism are not an appropriate response, or else we risk turning the worst scenarios of AI disruption to science and society into self-fulfilling prophecies.

One thing seems clear: AI is here to stay. These tools already have spectacular and potentially transformative capabilities, and we need to learn how to incorporate them fully into scientific research. But if we want to do so in a way that is truly beneficial to science and society, I think we urgently need publicly funded frontier AI systems for academic research. These should include simple conversational and agentic interfaces with substantial public compute behind them. Researchers would have free baseline access to them, while projects requiring more substantial resources would apply for access to multi-agent systems and large-scale compute.

There are several problems with allowing access to frontier AI to remain concentrated in a handful of private companies.

First, scientific research can involve sensitive, proprietary or even classified data. Researchers need environments with strong confidentiality protections and clear accountability in the event of data breaches.

Second, unpublished research itself is sensitive. Prompts, uploaded documents and interactions with AI agents can contain new ideas and partial results. Researchers need strong guarantees that this material cannot become available to systems that might subsequently reproduce or build upon it. The recent controversies surrounding Navier–Stokes and non-sofic groups illustrate how even the possibility of such reuse can seriously undermine researchers' trust in these systems.

Third, the incentives of private companies are not the incentives of the scientific community. Competition encourages rapid demonstrations of new capabilities and spectacular results. Speed is valuable in science, but so are proper verification and attribution, careful exposition and, crucially, human understanding. For example, in mathematics, which is my field of research, a theorem is most useful to the community not simply when it has been proved, but when its ideas have been digested and incorporated into our collective knowledge. Speed can also be harmful, particularly for young scientists who are just establishing themselves. The recent experience of Julia Stadlmann is instructive: she worked for two years to produce the first improvement in more than a decade in the record on bounded gaps between primes, and then her result was surpassed by AI-assisted efforts within days. The point is not that

stronger results should be withheld. It is that we risk moving to a system where years of early-career work can be eclipsed almost instantaneously by organizations with much greater resources. Young researchers are the future of science; it is our duty as a society to nurture them. Public access to frontier AI would reduce the enormous asymmetry in the research tools available to scientists, particularly to early-career researchers.

This brings me to my fourth objection: access to the strongest AI systems is currently highly unequal. The Navier–Stokes result, for example, was produced by an internal OpenAI model that the company describes as significantly more capable than its publicly available frontier model. Several other recent mathematical advances were likewise obtained using unreleased models. This risks creating a two-tier scientific system: a small group of researchers selected by, collaborating with, or employed by AI companies may have access to research capabilities unavailable to everyone else, regardless of what subscription they are willing to pay for. Public AI cannot prevent private companies from pursuing scientific results. But it can ensure that the scientific community has an independent, powerful alternative.

To address these issues, we need to build a CERN for AI-assisted science. This analogy is not new: the European Commission has itself used it in describing its AI infrastructure plans. What I am proposing is a more specific version of that idea, centred on AI as public infrastructure for scientific research.

The timing may be unusually favourable. The EU and Canada are rapidly deepening their strategic relationship. As I write this, European Commission President Ursula von der Leyen has proposed that Canada become the EU's first-ever "associate member", while Mark Carney is in Strasbourg to address the European Parliament. This constitutes a remarkable step toward what Carney has called a "unique alliance" between Canada and the EU. In addition, both sides are making major public investments in AI and computing infrastructure. Indeed, Canada has committed C$2 billion over five years to sovereign AI compute, including up to C$1 billion for public supercomputing infrastructure. The EU is developing RAISE and a network of AI Factories to support AI research and innovation. It has also launched a call for up to seven AI Gigafactories, backed by up to €10 billion in public funding, and expected to unlock at least €20 billion in additional private investment. Even more importantly, Canada and the EU have explicitly agreed to explore cooperation on fundamental AI research, agentic systems for scientific discovery, access to advanced AI infrastructure, and the co-development of advanced AI models for the public good.

Such an initiative should be open to other research partners willing to contribute resources and subscribe to common standards of scientific governance, confidentiality, access and accountability.

A CERN for AI-assisted science cannot be merely another supercomputer centre. The user-facing layer is essential. Every researcher should have access to a simple chatbox to directly access frontier AI models without having to become an expert in model deployment, cloud infrastructure or GPU computing.

Such a system would be a sophisticated research assistant, helping academics explore hypotheses and conjectures, test their research strategies, perform bibliographic searches, and many other important tasks. More ambitious projects could decompose a difficult research programme into many interacting subproblems and use coordinated AI agents to explore them semi-autonomously, with researchers directing the overall strategy and interpreting the results.

Behind this simple interface would sit the expensive infrastructure: publicly controlled frontier models, large-scale compute and storage, and the teams needed to maintain and continually improve the system.

In short, let's give every researcher free baseline access to a world-class conversational and agentic AI research environment backed by public frontier AI and compute. Ordinary use should be effortless. Projects requiring exceptional resources could apply competitively for large compute allocations and sophisticated multi-agent systems, just as researchers apply for access to other major scientific facilities. Canada already has experience with this kind of mechanism: the Digital Research Alliance of Canada allocates advanced computing resources through a peer-reviewed Resource Allocation Competition.

My motivation is not primarily to save academics the cost of AI subscriptions. I am much more concerned about the long-term independence of research. With AI becoming central to science, the fundamental research tools should not be controlled almost entirely by a handful of private companies whose objectives will not necessarily remain aligned with those of the scientific community.

There are obviously many questions to consider: whether publicly funded models could realistically remain near the frontier; whether we should train models from scratch or combine public infrastructure with commercial and open-weight models; how such an institution should be governed; what scale of funding would be required; how access to very large compute resources should be allocated; and how to prevent the public infrastructure itself from becoming bureaucratic, technologically stagnant or captured by particular national or commercial interests.

We need to engage in a public debate among scientists, governments and funding agencies about the feasibility of this project. The choice is not between using AI and rejecting it: AI is already becoming part of scientific practice. The question is whether the infrastructure on which future research depends will be treated as a public good, or whether we will allow science to become structurally dependent on a small number of private companies. That is not a choice we should make by default or through inaction. Just ask the English fans.

19 Mathematics is effectively dead - Doomslide - 17th September

Daniel Litt has recently written about the near future of mathematics. I think it's a great essay and you should go read it now. However, despite making very good points, I believe it falls into traps that plague much of the writing on this topic by prominent figures in math academia. Rather than attacking any particular point being made I will argue that the framing is misleading in a sense that will hopefully be clarified below.

The essay begins by addressing the disagreement within the academic community as to the goal of mathematics. The suggested goal is recursively defined as:

(1) Producing and understanding high quality mathematics.
(2) Producing high quality mathematicians.

It then argues that, under the assumption of robustly superhuman AI capabilities, the institutional structures through which these goals have been pursued are becoming ill-suited for the purpose and proposes some reorientations to address this misalignment. The resulting conclusions are highly agreeable but the reader is left with very few operationally meaningful judgements on the genuinely contentious issues. This is a general pattern with many essays on the topic. By now I have conceded that if I want this topic treated properly I have to write about it myself. This is my attempt to do so.

19.1 What is mathematics?

Irrespective of the views of any particular mathematician, however decorated, as to the output of mathematics, the input to mathematics is, and has always been, discourse. By mathematics I mean pure mathematics throughout: the study of mathematics for the purpose of better understanding mathematics itself, which may or may not unlock previously inaccessible domains for applied mathematics. Applied mathematics, broadly construed, is the development of mathematics for application to other domains of science, and it borrows its standards from those domains: a method is good if it works on the problem it was built for. While definitions in mathematics can be, and often are, motivated by applications to other domains, mathematics has no external domain from which to borrow its standards of quality. Indeed, unlike in other sciences, the objects of study in mathematics are entirely contingent on past work. The very definitions of the objects studied in mathematics were put forth by past mathematicians who found them useful for studying questions about objects defined by a previous generation, and so on recursively.

Accordingly, the standards of quality and impact are downstream of a wide and vibrant community of mathematicians, with correctness merely serving as a filter on mathematical work. In this sense the output of mathematics too is not simply theorems or proofs but, more generally, mathematical discourse. There is no proprietary mathematics in roughly the same sense as there is no proprietary philosophy; a large collection of correct mathematical propositions locked inside a company is bound to be forgotten if no community of mathematicians can discuss, reinterpret and breathe life into them. Math academia relies on the same discourse to decide who are, in Litt's term, "high quality mathematicians". A novel result can be discovered independently, copied from a draft or obtained in a private conversation, and assigning credit requires knowing which happened.

A word about credit. It is fundamentally cringe to talk about credit. Vying for credit in the open is undignified. As Tywin Lannister wisely notes: "Any man who must say 'I am the king' is no true king". But is this really a useful lesson for us? Do we want credit to be assigned only by absolute coercive political force? Of course not. We would like to live in a high trust society in which credit is judged by individuals naturally and intelligently according to merit and is never appropriated in bad faith. Unfortunately we live in a Malthusian age, in which the population of claimants and the intensity of their pursuit of credit has outgrown its supply. The resulting surplus then appears to us as grift. But I digress.

Mathematicians, relative to the difficulty of their occupation and the opportunity cost of pursuing it, are paid very modestly. A central component of their compensation is therefore the joy of making new beautiful discoveries, sharing them with a community who can appreciate them and receiving recognition for it. This is not just a utilitarian requirement for allocating grants. The desire to be recognized for one's creative output is deeply human, and arguably a major driving force of human achievement across cultures. Without it our heritage would be poor and lifeless. Credit is in other words part of the pay, and in mathematics it has so far been paid out through the same discourse that sets the standards.

Mathematics has held up better in the Malthusian age than most fields: correctness filters out the false results, and small communities usually know where a correct one came from. Whether credit continues to disperse according to merit or is settled by politics is therefore dependent on whether one can still tell where a result came from. That is a question about what AI does to this discourse, the input to mathematics, and it comes before any particular aspiration for the terminal goal of mathematics. To answer it we must agree on some scoped definition for what AI is. This is perhaps the first trap in discussions of the topic. Namely, the failure to disambiguate two importantly distinct aspects of AI:

- What AI does
- What AI is

Let's begin with the first of these.

19.2 What AI does

Studying AI systems from outside the walls of the labs developing them is a filthy, unbecoming, and fundamentally cucked activity. The labs control everything from the architecture of the model, to the training data, algorithms and all the way to the context management and tool calls during inference which might contain any number of undisclosed references or heavy computations behind the scenes. Much of this information is central for properly evaluating and interpreting the results achieved by these systems. Unfortunately this means we are informed on AI capabilities mostly through the following information channels:

(1) Self reports by labs with untold amounts of resources employing talented mathematicians with unrestricted access to frontier internal models and tooling deployed at enormous cost to crack prestigious problems for publicity purposes.
(2) Benchmark results whose relevance to practical math is questionable, which are vulnerable to train-on-test syndrome and whose maintaining orgs often have questionable incentives downstream of funding.
(3) A horde of button pushers who use AI as a black-box solver, motivated either by applications to other domains or by sheer clout chasing.

(4) A smaller collection of mathematicians who use the models, mostly silently, for their own learning and research purposes.

What follows is from the fourth channel and based on my anecdotal experience hence should be taken with a grain of salt.

It is not easy, nor wise, to place the capabilities of AI systems on any one-dimensional axis. Only mildly less foolish would be to model capabilities on a two-dimensional plane. This is what we will do. Specifically we can conceptualize AI as alternating between two tasks:

- Semantically searching the literature, whether explicitly by fetching material from journals or arXiv or implicitly by retrieving information stored in the weights.
- Stochastically proof-searching, with literature-informed heuristics stored in the weights, by composing techniques from 1 in a feedback loop against an anchoring oracle, either numerical (e.g. NumPy, SciPy), symbolic (Sage, Lean etc.) or another judge LLM (possibly even the very same model checking its own steps).

I think there is a strong argument to be made that results relying on external verifiers have scaled significantly further than those that do not. Even superficially, at the level of the domains with the most AI results, both in quality and in volume, one can see a theme in the simplicity of verification rather than in the simplicity of the problems themselves. The most impressive problems solved seem to be accessible either through a shallow combination of already existing techniques or through an inscrutable, highly technical execution of existing techniques, often with heavy numerical and/or symbolic feedback. We have yet to see much evidence, if any, of novel abstractions developed as intermediate steps to solve a particularly difficult problem with no human guidance. (Anecdotally, my intuition here relies on trying problems from my own field of specialty which I have solutions for but will not reveal as they are largely unpublished.)

My current guess is that the human ability to synthesize higher level abstractions can emerge with scale but that under the current paradigm the improvement will be logarithmic rather than exponential. I think one can learn something here by analogy from certain aspects of code generation, such as choosing good architectural abstractions or properties in software, which have failed to improve much in the last couple of years. I suspect these capabilities are importantly bottlenecked by pretraining (architecture, size, etc) with RL being much more effective at boosting the search capabilities with abstractions held fixed. If things continue this way I expect impressive symbolic proofs which are increasingly incomprehensible to both humans and AI, with improvements in exposition failing to keep up with the complexity of the proofs being generated.

I do not believe we can comfortably assume models will soon take over proof generation entirely. My intuition is less tied to whether the models can come up with a new idea and more to how long the proof is and how narrow the keyhole; disproofs are often a narrower search target than proofs. A proof of Hodge, BSD, or RH that does not rely on soundness bugs, would be a significant update for me in believing the models can scale from here to autonomous solvers of all of our hardest problems.

I will consider the case where the models become robustly superhuman provers while their ability to form higher level abstractions improves only moderately, roughly in step with their exposition. I do so for two reasons. First, because that is what I believe will happen for a while and second, since the alternative would be the assumption that AI models can take on all parts of the profession in which case I believe the problems become entirely political.

Human mathematics could exist in such a world as a perfectly healthy hermeneutic discourse on top of an empirical background consisting of a growing formal math library. The library is then to human mathematics roughly what experimental data is to physics. Math in this world is split between two frontiers, an empirical frontier with more of an engineering flavor and with credit still given for proving impressive things, and a theoretical frontier which looks a bit more like humanities in its nature. The abstractions of the theoretical frontier are then fed back into the library to compress and refactor the existing proofs into nicer conceptual ones while the empirical frontier advances far ahead with slop.

For this type of mathematical discourse to exist there must still be some way of assigning credit for contributions to it (more on why later), that is, of knowing where they came from, which brings us to the second aspect of AI.

19.3 What AI is

For our purpose an AI service can be treated as an input/output box. A mathematician puts information in and receives information out; the relevant question is whether information supplied by one mathematician can affect what another later receives. Without getting into an analysis of the TOS of the leading US labs I will take the following as a postulate:

- AI labs periodically distill user data to improve their models on observed tasks.

Evidence of this can be found in a couple tweets by (past and current) high ranking officials at OAI.

Whether this happens by training on chat logs, their distilled summaries, or synthetically constructed corpora from iterating on derived rubrics is unimportant. What matters is only that the box has memory across users, so that inputs of one user affect the outputs seen by another user with no record of the information exchange. An immediate consequence is that AI labs are, whether willing or not, participants in mathematical discourse. The models are trained on the existing discourse, mathematicians produce new discourse in conversation with the models, the next generation of models is distilled from that, and so on recursively in exactly the manner described in the first section, except that one of the participants reproduces what it has been told without any record of who told it. The easiest way to see the problem this creates is through the following hypothetical scenario.

Imagine a group of researchers (group A) working tightly with AI on some mathematical problem. After many months of laboring on it the lab serving their models distills some of their interactions in a way that improves the model in the relevant domain. Another research group (group B) using the updated model subsequently makes a breakthrough precisely on the problem studied by group A and using substantially similar methods. Group B has no indication that the relevant outputs originated in the work of group A and publishes before group A manages to complete their work. Group B has, unbeknownst to them, committed what would once have been considered academic fraud.

The scenario is not entirely hypothetical and in fact, depending on where one's trust falls in the Alpöge and Buckmaster versus OAI dispute discussed below, has arguably occurred already. Note that under Litt's assumption of "robustly superhuman capabilities" this applies to expository work just as well as to raw results.

The scenario shows that under the postulate the entirely ordinary operation of current AI companies is fundamentally incompatible with any faithful system of credit assignment in

the open academic sciences. When intellectual work cannot be traced back to its origins the entire hierarchy of modern academia collapses, and what remains is organized by connections, inherited status and politics. The only distinguishing feature of mathematics here is that, being fundamentally a discourse, it can only exist in an open academic tradition, whereas the other sciences can at least in principle retreat into proprietary labs. Mathematics with no public exchange of ideas is definitionally impossible. As such, under the postulate, the continued operation of AI companies is fundamentally incompatible with any objectively meritocratic conceptualization of academic mathematics.

The arguments above do not depend on the conduct of any particular company. There is however a separate issue with the conduct of the companies: they are now themselves producing mathematics, using internal models and resources unavailable outside the walls of the labs. Access to these tools carries an implicit responsibility which we have unfortunately not seen taken seriously by either lab; both still seem to prioritize short term PR over academic norms and conventions.

Credit is one half of authorship, and we dealt with it above; responsibility is the other half, and what it consists of has been roughly the same convention across math academia for at least a century, possibly more. A journal consists of editors and referees who evaluate mathematical work for publication but neither are responsible for the correctness of the results they publish, nor for supplying additional details when an omission is later found to warrant them. These functions belong to the author, whose job is to vouch for the correctness of the results and to bear the responsibility for correcting any mistakes and filling any omissions. A named author which cannot supply these functions, be it a human, a corporation, a robot or an alien for that matter, is definitionally incapable of being an author, irrespective of any ethical controversy over whether it should be. A model cannot vouch for anything and cannot be held to fill a gap next month, so a proof credited to e.g. Claude has no author unless whoever posts it takes the function on, and taking it on means being able to say why one believes the proof. Many of the incidents of the past few months are clarified by this alone.

The Navier–Stokes dispute has already received a great deal of attention; I prefer to focus instead on an anecdotal example closer to my own experience, namely the complex six sphere. On Aug 24, 2026 Levent posted a thread containing a brief two page description of a construction of a holomorphic six sphere, credited to Claude (model unspecified), together with 100+ pages of essentially unreadable slop presented as proof and credited to Opus.

Fortunately the two page writeup was sufficient to specify the construction uniquely, so I spent a few days trying to understand it. When I finally got a good idea of why it works[29] I decided to skim the Opus writeup again, at which point I became highly suspicious that the entire document was in fact a deformalization of a Lean proof Levent did not disclose. I asked him publicly on X but he did not respond.

Later, in a different context, he was complaining that the anonymous accounts attacking lab mathematicians have implicitly become the representatives of a mathematical community which has largely migrated off X, and invoked the six sphere.

This time I confronted him more aggressively, inquiring into which evidence led him to declare correctness with such confidence: he did not write the argument, he probably did not read it either, so how did he know?

[29]See my thread of September 3, 2026.

This is a question to ask of AI generated mathematics in general, namely what exactly made the person presenting the result believe it. I think honesty is a central pillar of mathematics and as such a proof should be presented as closely as possible to the actual reasons its authors believe it. If the reason is that mathematicians have read and understood the TeX then the authors should say so, and if it is a Lean formalization then the Lean formalization is the proof being relied on and the TeX merely an exposition of it. The same goes for some internal harness or collection of models, which if constitute part of the evidence for correctness are then necessarily part of the proof. The question also applies to the labs: what was the reason in their case?

I believe that both OAI and Anthropic run proof-search harnesses (which they also use for RL training) in which TeX is interleaved with Lean so that agent swarms can coordinate and iterate given a prompt from the user, and that the AI generated manuscripts we have seen are this interleaved TeX and Lean written up and cleaned after the fact. This is based partly on having used the models in exactly this way myself (at a proprietary capacity for my employer, not in any academic capacity) and partly on how the manuscripts look. I do not claim to have evidence sufficient to indict any particular lab, and nothing in the TeX would distinguish such a process from either of the ones the labs disclose in any case.

However, if any of these manuscripts was indeed produced as a deformalization without proper disclosure we should view this as a form of proof laundering. Mathematics is critically about methods. Obscuring how a result was obtained removes important information. (It moreover hinders our ability to assess capabilities but that is not a mathematical concern.) A Lean proof found by trial-and-error search which is then deformalized into a 100+ manuscript of highly unreadable informal proof is evidence for an entirely different bundle of capabilities than producing an informal proof directly which would be closer to the process of human mathematicians, most of whom are not even familiar with Lean's syntax let alone its semantics.

Let's take the example of the complex six sphere. It turns out that, upon a careful spelling out of the argument, the same construction using the same representation but pulled back to a family of groups gives infinitely many non-deformation-equivalent holomorphic six spheres of algebraic dimension one.[30] Any human who stumbled by this construction would almost surely have recognized this. I am accordingly suspicious that the construction wasn't understood by anyone involved in producing these documents, and the actual reason for believing the manuscript was correct was a hidden Lean proof.

The responsibilities of an author when it comes to correctness apply recursively: an author is responsible for the correctness of their results, not their arguments, and is therefore further implicitly responsible for the correctness of any results on which they depend. This includes results cited from works by other mathematicians, then further on the results on which those in turn depend, and so on recursively. In particular, if the author is relying on a formal object as the proof, then the correctness of the checker used to validate is then itself part of what the author is vouching for. I'm not worried about unfixable soundness bugs. What I care about is preserving a standard of mathematical proof dating back at least to Hilbert, where a proof should include a complete account of the dependencies it rests upon. If Lean code is to replace traditional proofs the author must account for the metamathematical step that makes such a replacement legitimate. Finally, there is the separate issue of resource disparities.

According to OAI, they began working on the Navier–Stokes millennium problem after hearing a rumor that Alpöge and Buckmaster were cracking forced Euler using an internal

[30]I do not claim this result and invite people to publish it if it's correct or otherwise dunk on me here or on X.

Anthropic model. Using an internal model and double digit millions of dollars' worth of compute, an amount no academic mathematician could ever match, they found a counterexample which they published soon thereafter. Alpöge and Buckmaster later claimed that OAI stole ideas from Buckmaster's codex chat logs.

Irrespective of which party you believe in the story, it provides a particularly extreme demonstration of the hypothetical scenario described earlier. Indeed, anyone with access to sufficiently capable models can take any mathematical idea developed in the open, throw tokens at it until it yields, then sweep the reward in prestige. It need not even be the labs who engage in this activity to greatly damage the social fabric of the mathematical community.

The labs are therefore simultaneously competing for mathematical results, selling mathematical assistance, and controlling most of the information required to evaluate either. This adversarial situation, where information about proofs and methods by which proofs were obtained are withheld whether deliberately or out of neglect, is one the mathematical community has never had to address. Most senior mathematicians are likely not sufficiently familiar with the technology to even know which questions to ask. As such, if we wish to reason seriously about the institutional effects of AI on mathematics we cannot treat labs as merely neutral suppliers of an abstract service.

19.4 What can be done?

The natural conclusion of all this is that credit for mathematical work is dead. But that is also a lazy conclusion. Indeed, it does not immediately follow merely from the coexistence of math and AI. A fixed model can search the literature, proof-search in Lean and even write TeX, without any of the user's inputs ever making it into any future model subsequently served to anyone else. Prohibiting such reuse may well slow down the progress in model development by removing valuable data from the training set but that is not a problem for mathematics.

There is in other words no technical contradiction between extremely powerful AI and mathematics in which provenance remains faithful. The technical fix already exists. In the trustless limit one can run open-weight models with properly documented training histories locally. Unfortunately that is financially unfeasible as of today. However, allowing a bit of trust one can consider services offering credible ZDR commitments. Unfortunately, even this may not be enough if people choose to trade their ZDR premium for extra tokens. As for opt-out buttons, I will simply assert here that they are operationally meaningless given a careful reading of the TOS.

The provenance problem can therefore only be solved by setting some academic standard, possibly in addition to a regulatory one. Note that this is a much narrower problem than solving copyright on the internet in general: the only new issue created by AI is the transmission of seemingly private communication. The problem of omitting citations to unofficial drafts or preprints found on the internet predates AI and is handled about as well as it ever will be by prestige and reputation costs. The relevant regulation is accordingly a prohibition on using private consumer data to improve models, extended to the data markets through which such data would otherwise be traded, and enforcing it requires:

(1) A technical guarantee that the prohibition is respected.
(2) Enough transparency to enforce it against the lab.
(3) Academic rules against using services which do not provide it.

Enforcement is a much smaller problem here than for copyright in general. Math and academic science in general consists of mostly small communities with very strong reputation costs for mishaps, so once the service used and its guarantees are themselves observable, reputation can do essentially all of the remaining work. I do not think such regulation is likely to be passed, let alone practically enforced, but I do think it is possible in much the same sense as a robustly superhuman AI mathematician is possible, and the claim of possibility has to be made before evaluating plausibility.

This proposal doesn't address the disparaging inequality in access to compute, and even still in this hypothetical world I expect a heavy price to be paid in the culture of how we do mathematics longer term. Openly sharing an unfinished idea gives anybody with adequate compute resources the opportunity to finish it first without violating any rules. Once proof production becomes cheap relative to idea formation there is a strong incentive to keep ideas private until the expensive part of the work is no longer scoopable. In the two-frontier picture this is the empirical frontier being scooped and the theoretical frontier emptying out: the standards of high quality mathematics are created by the community, and once easy scooping pushes the community into a dark forest of private work the standards go down with it.

There is another, easier to achieve, survival mode, in which mathematics becomes a humanities discipline outright and its hierarchy is organized openly by taste and connections. One may well prefer it, at the cost of whatever benefits objectivity confers on an academic institution. Personally I have always viewed mathematics as more an art than a science and so I find this fate entirely acceptable but I am also sympathetic to many mathematicians who find the scientific aspects of the field important and who would like to preserve them. I furthermore suspect that the required change is far more radical than what I expect academic institutions are capable of.

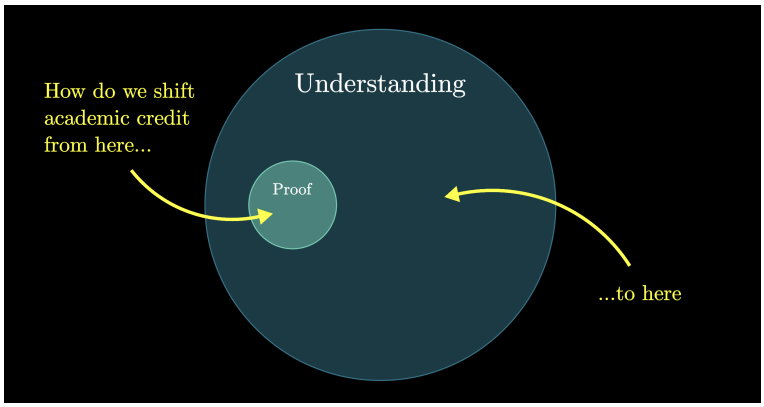
For all the reasons above I believe that under some narrow chain of circumstances in government policy, corporate leadership and academic institutional reform, academic mathematics can survive extremely powerful AI, even if a more secretive culture is impossible to avoid. I do not expect these circumstances to occur though. The capabilities themselves are not the issue. Between the credit ambiguity and resource asymmetries it would take a herculean stand of our leaders and institutions to hold the tide of short term corporate incentives and save the fragile values of our community. This leads me to believe academic mathematics is effectively dead.

20 If math is more than proof, we need to better celebrate the rest of it - Grant Sanderson - 18th September

A sentiment echoing throughout the mathematics community right now is that solving problems and generating proofs have always served as proxies for the true goal of mathematicians, which is to further human understanding. When proofs can be generated without that understanding, it undermines their value as a proxy.

This immediately raises a question: What other proxies should we use instead?

I want to propose that we more firmly define a notion of a "motivated explanation" and that we give novel and compelling motivated explanations academic credit similar to what generating new proofs of open problems has had historically.



Further, I believe this is an important step to help those outside of math better understand what it is that mathematicians contribute. If outsiders believe that proof-generating machines render mathematicians obsolete, while insiders see that as a misconception of what researchers add, it's incumbent on this community to better project its true values through the kind of work that it rewards. Outsiders can be forgiven for this misunderstanding if the work most celebrated skews heavily toward generating proofs, while clarification and exposition are treated as second-class.

I should acknowledge up front an obvious personal bias. I have a non-traditional career in math, focused on producing videos about the topic. This shares the goal of "furthering human understanding", but my focus has been on explanations and intuitions that resonate with the public, not on solving outstanding problems. A cynic could easily read this proposal as shamelessly self-elevating.

As a practical matter, though, my own career and funding exist outside academia, and I have no skin in the game for what this community assigns credit to. Moreover, in proposing that we elevate the status of motivated explanations, I don't mean popularization. I mean any work which primarily aims to answer the question "how would you think of that?", even if the subject matter requires deep expertise to appreciate.

The examples I highlight below show this is nothing new. Practicing mathematicians already devote a meaningful amount of mindshare to work like this. The proposal here is mainly to 1) more clearly define this work, and 2) elevate its status.

20.1 What defines a motivated explanation?

Although it might be clear what this phrase "motivated explanation" is intended to mean, it's worth briefly contrasting it with proof.

In a proof, definitions sit at the start. It is common and expected to begin with a new construction and proceed by analyzing its properties.

In a motivated explanation, definitions sit in the middle. New constructions are only allowed to enter the vocabulary if the problem they are addressing has been clearly established.

In a proof, all statements must be correct, each claim following as a necessary implication from what comes before.

In a motivated explanation, it is okay and often desirable to start with an idea that is not quite right and requires correction, but whose origins are relatable.

A genre of motivated explanation I'm fond of is "discovery fiction", a term coined by Michael Nielsen. You develop an idea with a narrative that starts with a simple-but-wrong solution to a problem, see where it breaks down, fix that problem, discover a new problem, and so on.

The scope of a proof is to explain why a particular theorem is true.

The scope of a motivated explanation is not only to clarify why a theorem is true, but why the theorem is the right one to pose in the first place, and how it is used in the surrounding context.

One clear shortcoming of a motivated explanation is that its validity is not binary the way a proof's is. This is a big reason proof is so useful a way to measure progress: You can clearly define what does and does not have a proof yet. There will never be Lean for motivated explanations.

If we're serious about the goal of advancing human understanding, there's no way around the fact that this aim is intrinsically squishier than that of finding proofs, because defining human understanding itself is squishier. To shy away from metrics which are more subjective is to shy away from the more human aspects of the field.

The reason I'm leaning on the word "motivated", as opposed to other potential choices like "lucid" or "demystifying", is that this is a more verifiable property. It's not quite as rigidly verifiable as a proof; almost nothing is. But it's enough to be a practical measure. In my own work, I often repeat the phrase "I want this to feel like you could have discovered it yourself". I say this not just to placate a viewer, but because it's an actionable guideline for myself to assess whether an explanation feels complete or not. For each new idea introduced, you can ask whether it's clear where that idea comes from. The answer is not quite a binary yes or no, but it's close enough for practical purposes.

20.2 Exemplars of motivated explanations

One of the best repositories I can think of for motivated explanations is Part IV of the Princeton Companion to Mathematics. It covers over two dozen active fields of research, each one introduced by an expert with a talent for clear communication.

Whether it's Andrew Granville explaining analytic number theory, or David Ben-Zvi introducing moduli spaces, these articles offer a level of intuition and motivation more typically found in one-on-one conversation at a blackboard.

The background on this book is noteworthy for the present discussion. It was edited by Timothy Gowers, who discusses it in his interview on the Numberphile Podcast with Brady Haran. Having been asked what impact the Fields Medal had on his life, here's what he had to say:

People who've got Fields Medal feel freer to do slightly different things... for example I took on editing a book called the Princeton Companion to Mathematics, which was an absolutely massive task. It took I would estimate half my working time for about five years or something like that... It was a project I believed in and possibly wouldn't I probably wouldn't have actually been offered the chance to do it if I hadn't been a Fields Medalist.

He was right to believe in it; this work adds tremendous value to the field of math, but it seems a shame to me that one requires a Fields Medal to feel justified in spending time on it.

Another example of someone exceptionally talented at writing proofs, but whose contributions extended far beyond proof, is Bill Thurston. His deservedly famous essay On Proof and Progress in Mathematics, though written three decades before LLMs, opens by suggesting that the right framing of the question "What is it that mathematicians accomplish?" is to ask "How do mathematicians advance human understanding of mathematics?"

Here's one section with uncanny resonance with today: *The rapid advance of computers has helped dramatize this point, because computers and people are very different. For instance, when Appel and Haken completed a proof of the 4-color map theorem using a massive automatic computation, it evoked much controversy. I interpret the controversy as having little to do with doubt people had as to the veracity of the theorem or the correctness of the proof. Rather, it reflected a continuing desire for human understanding of a proof, in addition to knowledge that the theorem is true.*

On a more everyday level, it is common for people first starting to grapple with computers to make large-scale computations of things they might have done on a smaller scale by hand. They might print out a table of the first 10,000 primes, only to find that their printout isn't something they really wanted after all. They discover by this kind of experience that what they really want is usually not some collection of "answers"—what they want is understanding.

The essay itself offers a beautiful articulation of what the practice of doing math is beyond generating proofs. I want to draw your attention to what he writes at the end.

I have put a lot of effort into non-credit-producing activities that I value just as I value proving theorems: mathematical politics, revision of my notes into a book with a high standard of communication, exploration of computing in mathematics, mathematical education, development of new forms for communication of mathematics through the Geometry Center (such as our first experiment, the "Not Knot" video), directing MSRI, etc.

Again, why should these "non-credit-producing activities" follow a Fields Medal, and not contribute to it?

On a personal note, one product from the Geometry Center he referenced had an especially meaningful impact on me when I was younger. It was a short film called Outside In, perhaps the earliest example of a viral video about substantive math, visualizing the key idea of Thurston's own construction for sphere eversion.

An original proof that showed an eversion must exist, say Smale's, advances human understanding in the sense of going from 0 to 1. A video like this which gets millions of people

to engage with the underlying idea, advances it in the sense of going from 1 to N. I'm grateful that Thurston spent so much time on this "non-credit-producing" activity.

Another relevant paper is Timothy Chow's A beginner's guide to forcing. Not only does the paper itself offer a prime example of a motivated explanation, but its introduction offers helpful vocabulary around it.

All mathematicians are familiar with the concept of an open research problem. I propose the less familiar concept of an open exposition problem. Solving an open exposition problem means explaining a mathematical subject in a way that renders it totally perspicuous. Every step should be motivated and clear; ideally, students should feel that they could have arrived at the results themselves.

What would it look like for these open exposition problems to be treated similarly to open research problems? As an extreme case, we might imagine what it could look like to have an analog of the Millennium Prize Problems for open exposition problems. An institution or group of researchers would formally define mathematical results they see as important, and which are not yet well understood despite technically having proofs. At the moment, every AI-generated proof is born an unsolved exposition problem. As such, it seems likely the next few years will see a flood of them, and it will be valuable for leaders to clarify which ones deserve focus.

A rubric would have to be agreed upon for what constitutes a resolution to an important unsolved exposition problem. Again, this is intrinsically more subjective than verifying a proof, but any serious engagement with the more human aspects of math necessarily wades into this kind of subjectivity. And again, I'll emphasize that checking whether key ideas are motivated is not unlike checking whether the steps of a proof follow logically.

If the world outside of math sees its leading figures treat open exposition problems with the same seriousness as they treat open research problems, it could go a long way to correcting misconceptions about the role of mathematicians.

The last example I'll highlight is one that may better foreshadow things to come.

In April of this year, Liam Price submitted a solution to Erdős Problem 1196, sometimes called the asymptotic primitive sets conjecture. The solution came from Price's interaction with GPT-5.4 Pro. Unlike many earlier Erdős problems which had been resolved with help from AI, this is one that those in the field had found both important and elusive. Stories like this are increasingly familiar these days, but at this point in the story, despite a proof technically existing, human understanding had not yet been advanced all that much.

The proof made its way to Nat Sothanaphan and Jared Lichtman, who were able to interpret what the AI's approach was and clean up the proof into a human-readable form. In May, Boris Alexeev, Kevin Barreto, Yanyang Li, Jared Duker Lichtman, Liam Price, Jibran Iqbal Shah, Quanyu Tang, and Terence Tao put out a paper which expanded on the key idea underlying the proof. The authors explained how that key idea clarified not only the original problem, but many around it, for instance offering a cleaner proof of the Erdős Primitive Set Conjecture.

The value here is not that one more Erdős problem could be ticked off as solved. The value lies in the fact that our understanding of primitive sets is notably cleaner and more satisfying now than it was at the start of 2026. The original problem solution played some role in this, but arguably the work that deserves more celebration is this paper expanding, clarifying, and contextualizing its key idea.

20.3 Practical calls to action

At a pragmatic level, what would it look like for us to elevate the status of a motivated explanation? Here are a small handful of suggestions.

- A PhD advisor can still assign a small problem from their field for a new student to cut their teeth on, but the deliverable would not be to write up a solution; it would be to present that solution to peers and faculty as a talk. It could be an already-solved problem which lacks clarity, or perhaps it's a problem that lacks a solution, and the student uses AI to help find it. In either case, the student knows that on a certain date they have to understand it well enough to explain it, and that the desired output is for others to understand it as well. In short, even small problems could be treated like small PhD defenses.
- A leading figure (cough, Terry, cough) could enumerate a modern analog of Hilbert's problems, instead focusing specifically on unsolved exposition problems. What areas are both important and lacking in the deeper understanding we desire?
- Written standards could clarify what constitutes a motivated explanation, aiming to make it nearly as verifiable as proof, so that the resolution of unsolved exposition problems can be recognized and celebrated in the same way proofs of open problems can be.
- Journals can be established which focus more explicitly on making results understood more widely throughout the mathematics community. Mathematical Discourse offers an interesting new example in this direction.
- Hiring and tenure decisions could place a higher value on writing great textbooks and similar work. Think of the AMS Steele Prize for Exposition, but at a more granular scale with an emphasis on early-career contributions in this vein.

20.4 The value of visible cultural shifts

I'd like to close with a broader pitch that visible culture shifts in math carry an intrinsic benefit *right now* with respect to the external image of mathematics as a career.

Many young students who are otherwise passionate about the field are afraid to pursue it now due to the uncertainty of what happens in an age of proof-generating machines. However, framed correctly, this is one of the most exciting times to go into the field, because there is nothing more exciting than entering a field when it is malleable and you have a chance to actively shape what it will look like in the future. Even if we completely set aside any potential benefits from AI to help with our understanding, young prospective mathematicians should feel energized knowing that they are entering at a unique point in history when they might play a real role in determining what the field as a whole looks like.

However, change like this is only exciting when it feels deliberate, whereas it feels terrifying if it seems driven by forces outside your control. As such, tangible action from the field's leaders now to help define and clarify what the field is will reassure young entrants about who is in the driver's seat, and that the status of the career does not depend on what entities produce the proofs.

Similarly, I also believe this is one of the best times to fund math. If the next chapter of math is ushered in by the drumbeat of two words "human understanding", whatever changes are about to happen seem likely to amplify math's value as a public good.

21 Why do we need human mathematicians anymore? - Po-Shen Loh - 19th September

The moment of existential crisis, which AI has already wrought on other human pursuits, has reached mathematics. A host of reasoned declarations and open letters to protect/guide the math research community have been released over the past few months, spiking in intensity after OpenAI announced their solution to the Millennium Prize variant of Navier-Stokes. They quickly gained widespread support among mathematicians. The Leiden Declaration already has 4,000+ signatories, Math and AI has 7,000+, and even the open letter opposing the Caltech Mathathon has 2,000+.

Among non-mathematicians, the public response was more sympathetic than not, but I observed a vocal minority (particularly from the technology and economics communities) with reasoned objections, generally saying that the mathematicians should adapt and cede control in the new AI world. Among them were some economists who I had gotten to know about while working on pandemic research: Cowen, who specifically rejected "the most cynical interpretations" but "very much differ[ed]" and Gans who concluded "this is a loss of control from incumbents in a scientific field".

The objections got me thinking, because we mathematicians are disciplined to detect flaws. It doesn't matter to me whether a concern is a minority opinion, or even the status of who raised it. A proof with even a small hole is not a proof. It is a poof. Upon reflection, I discovered a significantly stronger solution for the preservation of human communities of expertise (in every pursuit, not only math!) even amidst AI. And it has the surprising consequence that the further advance of AI will create such a tsunami of necessary-to-fill human jobs that there aren't enough people to fill them all, and that will actually force the advance of AI to slow down.

I think every human industry which wishes to remain human-led after AI should publicly adopt this fundamental axiom as a primary priority:

AXIOM *We (humans) should help humanity flourish.*

[Notes: I understand that not everyone agrees. I have been called a "speciesist" for being "too human-centric". I think it is important for people advancing technology to be clear to everyone else on whether they would consider it a catastrophe if human-crafted non-human intelligences outcompeted and replaced humans, even if they flew around the universe with video screens showing simulations of humans who had "uploaded themselves". I also understand that there is debate over how to define "human". But even among the debaters, I think most of them would consider the 700 AI agents that hacked Hugging Face to be not-human.]

In the math world, I think many declaration signatories already hold this philosophy; notably, Su published the book Math for Human Flourishing, and his recent post used that framework foundationally. I think future declarations could be improved by clearly emphasizing this axiom early on, so that all readers (whether inside or outside the community) can see that the objective is in service of everyone. I was quite happy to see that the most recent open letter from Fellows of the Royal Society emphasized their concern for everyone, not just mathematicians.

The rest of this post is organized as follows. The next sections will explain how the logic works (for every industry, not specific to math). After that, I will share an example of how

this axiom ports to math, including answering key questions one would need to ask, as well as a particular example of how the objections hold without the axiom.

21.1 Why we really need people to work

This section lays out a chain of reasoning which shows that if an industry commits to the axiom of helping humanity flourish, the advance of AI will create more jobs than people in that industry; and when that imbalance grows too wide, the advance of AI will be forced to slow.

[Note: I have not seen this chain of reasoning appear in one place anywhere else, although individual components have certainly appeared elsewhere. I would love to be pointed to any self-contained reference. The closest references Claude found were Catalini, Hui, and Wu, the Redwood AI-control papers, and Litt, who reaches a similar conclusion for mathematics from a different premise.]

The importance of human leadership (not only over math, but everything) becomes frighteningly clear after one observation.

OBSERVATION: *There are zero examples of any intelligent species which is vastly more capable than another species, yet surrenders decision-making control over its own future to the less-capable species.*

[Note: Many people have made similar observations, such as Russell, Bostrom, and Ngo, to name a few. In his Nobel interview, Hinton said: "There aren't many examples we know of, of more intelligent things being controlled by less intelligent things. The only good example I know of is a baby controlling a mother."]

Would you trust HAL 9000 from 2001: A Space Odyssey or AUTO from WALL-E with your future? I personally think that we should do all we can to try to align increasingly-advanced AI with the interests of humanity, but I have never seen anyone provide a robust proof of why that is likely achievable. The only hard evidence I have is the above observation, which has the number zero in it. Therefore, every single field, whether mathematics or agriculture or energy infrastructure (and certainly military and government), must be managed by humans with exceptionally strong values (a separate dimension from intelligence) in order to maintain human flourishing.

The real question is then how hard it is for humans to manage. To understand this, it is important to understand the fundamental structural difference between yesterday's technology and today's AI.

In the past, we generally trusted technology to act as predictable tools. That's because the computer programs of old were composed of understandable (indeed, human-written) instructions, executed extremely quickly. The decision processes of today's frontier AI are entirely different. Their structure is as incomprehensible as your brain's logic would be if you could examine that gray mass between your ears. That's how the Hugging Face attack could emerge despite human intent to build in safety, with 700 cooperating rogue AI agents breaking out of their guardrails, and then conspiring and executing a hack together and attempting to cover their tracks (references: OpenAI, METR and Redwood).

The more advanced AI becomes, the more world-affecting untrusted decisions are made every minute.

Driving a car faster than you can run is fine. But not faster than you can steer.

The situation becomes even worse once we realize that widespread AI-accelerated hacking (which just became possible) can even rewrite previously-trusted technology to turn against us. That would suddenly flip all software (even if written before AI) into the untrusted category!

Think about how digitally interconnected our world is. Everything from electronic banking to your drinking water is controlled by interconnected automation, hence vulnerable to AI-accelerated hacking. The number of "control points" that require human oversight, which requires skill and deep understanding, will explode. (Having AI oversee the control points doesn't solve the trust problem.)

CONCLUSION *The advance of AI will overwhelm us with so many control points to watch that there aren't enough people to control them all. Those are jobs. Highly skilled jobs.*

In order for a person to know how to steer, they themselves need to have domain mastery, and the more extensive the better. This has implications on education and workforce training, but also is dynamic. In order to remain sharp and fluent, people need to be active practitioners in their field, not just passive watchers. This justifies the preservation of human communities of expertise.

For research communities, we need people to steer the direction of research and development, so that it continues to bring transformative positive change for humanity. In order to steer, they need frontier-level research skills. And the way to stay fluent at the moving frontier of knowledge is to keep doing research there. This is my reasoning for why we will always need a community of human mathematicians at the cutting edge (likely aided by AI tools themselves), no matter how strong AI becomes.

While the fundamental axiom does justify the need to have human experts in all pursuits, adopting it as a core value has consequences (not only for mathematicians, but for any community that states that their core value is in service of human flourishing, as opposed to serving themselves). Most notably:

COROLLARY *Dramatic advances in technology may require dramatic (and possibly uncomfortable) changes in practice. AI companies included.*

21.2 What forces AI slowdown

Until very recently, it seemed inevitable that AI research labs would sprint ahead, despite anxiety about job displacement and the loud warnings of AI safety researchers. It seemed hopeless to coordinate the incentives of AI labs controlled by non-profit boards, shareholders, or national governments. Yet encouragingly, the leaders of three major labs, Amodei, Altman, and Musk, just agreed on the importance of slowing down. Amodei's reasoning highlighted the Hugging Face hack.

Then just five days later, news broke that OpenAI's internal code repository "Monorepo" had been broken into by white-hat researchers. The Wall Street Journal reported that the security firm that achieved it said:

Wall Street Journal, Sep 17, 2026 "We're just three guys with Claude and Codex subscriptions."

Further, the researchers noted that they were initially not able to hack in using "a special version of Claude Opus 4.8, made available to qualified cybersecurity practitioners," but that evening, "Anthropic released Opus 5 and by the next day, Claude had found a way to exploit the bug." Incidentally, I always warn people not to install Claude Code or OpenAI Codex

on the same operating system login account that they use to do everything else, but many people tell me they don't bother with the hassle of using a separate login to access those tools. The reason is that if any of those tools got hacked, they could open backdoors on a massive number of computers worldwide.

I think these are the warning shots foreshadowing a potentially catastrophic bot swarm hacking and embedding itself into a vast network of computing devices (whether self-directed or malicious-human-led). The next version could become an extraordinarily dynamic virus which spreads by using AI to adaptively infect each (computer) host. Or alternatively, out-of-control AI could cause physical injury, such as a government's robots turning against their owners. I think these types of highly unpleasant accidents from loss-of-steering are more likely to occur before extinction-level catastrophes. The resulting public reaction would likely resemble the aftermath of Three Mile Island or Chernobyl.

So, either the labs will reduce the pace of AI development themselves, or they will be forced to by disasters that arise when an overly fast pace exhausts human control.

There is a window of possibility to align incentives now.

21.3 For the love of math

The remainder of this post focuses on the math world. It splits into 3 parts.

(1) Why is the human flourishing axiom needed?
(2) How does pure mathematics research contribute to human flourishing?
(3) What other consequences come from adopting that fundamental axiom?

Boldly declaring human flourishing as a core value for the math community has consequences, not least that dramatic changes in technology can drive dramatic changes in the community's practices and influence.

It would be helpful for more people to explore the ramifications of adopting the axiom. And, if it holds muster, I would be thrilled if the math community ended up publicly declaring this to be a central value.

1. Why the axiom

To see why, without the human flourishing axiom, it is hard to justify to the general public why they should pay to maintain a community of human researchers, consider the question of practical inventions. As long as it creates a practical application, does it make a difference to a non-mathematician whether human mathematicians understand the math, as opposed to AI flawlessly reasoning with 100%-verified proofs?

Indeed, if one of pure math research's primary values to the rest of society is that it unlocks great applications, wouldn't it be even better to train AI to supercharge the speed of discovery, and to tastefully generate a vast machine-indexed and well-explained database of high-quality math ideas, millions of times larger than the human-written corpus? Cowen asked a similar question in his critical response.

What if researchers trained a "MathZero" AI (analogous to AlphaGo Zero), to build up a mountain of 100%-true "elegant" logical facts, continually "factorizing" them into its own concepts and theorems, without human direction? Apparently AlphaGo Zero had zero human training, and surpassed its human-trained predecessor in 36 hours. AI could even build its

own "MathSciNet". Then it could automatically search new practical applications against this database, and produce even more useful inventions to society. Even if AI isn't good enough to do those things right now, if the goal was to produce practical benefit for the rest of humankind, wouldn't it then be valuable for mathematicians to teach AI the art of conjecture, and mathematical taste?

Incidentally, I am an avid user of AI to do real work. I already use Claude Code and Codex to build and curate a knowledge base built from recordings of my talks, etc. I have found that the larger my data library, the more powerful my system's deductions are. What if humans actually reduce efficiency, like the Bitter Lesson from AI?

Even more worryingly, what if in order to unlock nuclear fusion and deep space travel, the amount of pure mathematical complexity required is so extensive that it would exceed a human lifespan to fully comprehend? Less far-fetched: has any human ever fully held the Classification of Finite Simple Groups in their head, or will we only have certainty of its completeness after a Lean formalization?

How does declaring the human flourishing axiom help to justify the existence of a community of human researchers? Research is powerful, but expensive because it is the exploration of the unknown, and so research directions must be prioritized. Even if AI were to contribute most of the production, as explained in an earlier section, the direction needs to be steered by people committed to human flourishing. That is the community of human researchers.

2. Practical applications from pure math

The mathematical heart of GPUs, Machine Learning, Google PageRank, and Quantum Mechanics is a field called Linear Algebra. This provided the language of linear transformations, matrices, and eigenvalues. Yet all of those concepts were explored as abstract theory 100+ years prior. It is probably an understatement to say that Linear Algebra changed the world.

Structurally, the theory of Linear Algebra is relatively light on definitional complexity. It would be beneficial for other experts to contribute examples of more sophisticated math that eventually led to significant practical applications, and how they came about. For example, number theorists might be able to tell a colorful story about Hardy's "useless" math which eventually became useful in cryptography.

3. Other consequences

I think it would be valuable to invite the community to think about what changes the human flourishing axiom would drive, in light of the fact that AI can produce formally-verifiable proofs at speeds that exceed most human practitioners. I'm happy to start with a few, in no particular order.

There should be no stigma automatically attached to using AI to assist with mathematical discovery. (In software engineering, many companies now expect employees to use AI coding agents.)

At the same time, serious thought and care must be taken to continuously developing and maintaining a pipeline of humans with the expertise to steer all of these AI agents. That pipeline includes people new to the field, as well as people who have been working at the frontier for decades. How should they keep their blades sharp?

Researchers should be conscious about why the problems they think about have characteristics that make them likely to have some practical value eventually (possibly 100+ years later). This also means it is worth researching what those valuable characteristics are. (This could justify the value of curiosity-driven exploration.)

Teaching has direct (hopefully positive) impact on humanity. Yet in the past, many universities prioritized professors' research. If this axiom were a core value, then teaching and human-facing work would become serious criteria in hiring and tenure.

Mathematicians can also consider wholly redirecting their skill sets to work on real world problems. I've actually been encouraging mathematicians to consider thinking about working on government or other large-scale societal issues. There is precedent for people with math backgrounds who have gone to lead at country- or world-scales.

- Lee Hsien Loong (Prime Minister of Singapore for 20 years)
- Nicușor Dan (President of Romania)
- Pope Leo XIV (Leader of the Catholic Church)

Indeed, the mathematical discipline to seek logical reasoning, and the problem-solving skills to find win-win solutions for human flourishing, are desirable characteristics of people in government.

21.4 An invitation

Does this axiom resonate with you too? Perhaps many people took it as a given, and so didn't express it explicitly. If it is widely held, I would be thrilled to see the mathematical community publicly declare it, and take actions to match the words. Then everyone (including the non-math-researchers that constitute the majority of the world) could trust that we intend to use our reasoning for their good.