Mathematics in the age of AI

Public lecture, International Congress of Mathematicians 2026

Terence Tao

University of California, Los Angeles

July 24, 2026

Historical prologue: the crisis in foundations

For centuries, mathematics successfully operated using “naive” foundations, with some fundamental questions largely delegated to philosophers, such as:

- What is a set?
- What is a number?
- What is infinity?
- What are the axioms of mathematics?

But in the early twentieth century, discoveries such as Russell's paradox (1901) or the Gödel incompleteness theorems (1931) forced practicing mathematicians to critically re-examine their implicit assumptions about the foundations of mathematics.

- This crisis in foundations ($\sim$ 1900–1930) was a turbulent period for mathematics.
- But the end product was extremely valuable: an explicit, rigorous, and standardized foundational framework.
- There is scope for further improvement.
- But our current foundations have survived strenuous testing and are a trusted environment for mathematics.

- I believe we are entering a similarly turbulent period —[1] a crisis in the foundations of mathematical values and practices.
- But once we thoroughly examine and codify these foundations, our community will emerge stronger and more resilient than before.

[1] All em-dashes in these slides were human-generated.

The motivating question

This talk is focused on the following question:

Community Response Question

How should the mathematical[2] community respond to the advent of modern AI technologies, and their real and/or claimed capabilities to perform mathematical tasks?

This is a question for the entire community; I do not presume to have all the answers. Nevertheless, I have some things to say.

[2] The impacts of AI of course extend far beyond mathematics, but that is too vast a topic to cover in this talk.

Community Response Question

How should the mathematical community respond to the advent of modern AI technologies, and their real and/or claimed capabilities to perform mathematical tasks?

- This question is **not** a mathematical question; it is a metamathematical (and political, ethical, and cultural) one.
- However, in this talk, I will **borrow** from the precise, familiar language of mathematics in order to clarify the points I wish to make today.

The first subquestion: AI capability

The answer to this question depends crucially on the following subquestion, which I will formulate pseudomathematically as a conjecture (or more precisely, a family of conjectures):

AI Capability Conjecture (template)

At some point in the near future, some AI tools will, at some expense, and with some level of human supervision, be able to accomplish some research-level mathematical tasks in some fields of mathematics, with some non-trivial success rate, and at some level of correctness and quality.

The words "some" here should be viewed as placeholders (or "variables").

AI Capability Conjecture

At some point in the near future, some AI tools will, at some expense, and with some level of human supervision, be able to correctly accomplish some research-level mathematical tasks in some fields of mathematics, with some non-trivial success rate, and at some level of correctness and quality.

- One can create many formulations of this conjecture, depending on how one fills in the placeholders.
- The finer distinctions between these formulations is not the point of my talk today.
- I will only make only a coarse distinction between "weak" and "strong" forms of the conjecture.

- If even weak forms of the AI Capability Conjecture are false, then we could safely choose to dismiss AI tools as being of no long term significance, and largely continue “business as usual”.
- But if the strongest forms of the conjecture are true, then it becomes challenging to maintain our current culture and practices, **particularly if we prioritize such goals as obtaining as many solutions of unsolved problems as possible**.

- It is difficult to have a constructive discussion on the Community Response Question when the status of the AI Capability Conjecture remains under dispute.
- Thus, much of the debate about AI for mathematics has focused on which versions of the AI Capability Conjecture are true.
- I myself have devoted many lectures, writings, and posts on social media to such topics.

- There are now many data points for or against various forms of the AI Capability Conjecture.
- But **most have not been gathered under controlled scientific conditions**.
- In particular, much of the publicly available evidence is highly subject to reporting bias and non-scientific incentives, with some important costs and variables remaining undisclosed.
- Furthermore, it is important to distinguish the truth value of a given form of the conjecture from its desirability.

- Despite the central relevance of the AI Capability Conjecture to the Community Response Question, my talk will not be about that conjecture.
- In this direction, I will only mention the recent results of the First Proof challenge `1stproof.org`.

First Proof

- First Proof is an independent assessment of the state of frontier AI models and harnesses.
- Each “batch” consists of ten novel research-level problems in various fields.
- The second batch was tested under controlled scientific conditions against four AI harnesses on May 28, 2026.
- Results were refereed by experts for both correctness and exposition.
- Seven of the ten problems were solved at a publication-level quality by at least one team.
- Compute costs ranged from 10 to 1000 USD per problem.
- Further “batches” are planned in the near future.

The complement to the capability conjecture

- This talk will instead be about the "orthogonal complement" to the AI Capability Conjecture within the Community Response Question.
- The rest of this talk will therefore be conditional on the following (imprecisely stated) hypothesis:

Working Hypothesis

A reasonably strong version of the AI Capability Conjecture is true: AI tools will, reasonably soon, become capable of performing a reasonable fraction of research-level mathematical tasks, with reasonable levels of success, quality, supervision, and cost.

The precise definition of "reasonable" is not critical for my talk.

Working Hypothesis

AI tools will, reasonably soon, become capable of performing a reasonable fraction of research-level mathematical tasks, with reasonable levels of success, quality, supervision, and cost.

- For the rest of this talk, I will ask you to assume that the Working Hypothesis holds.
- I am **not** asking you to want, believe, or accept that this hypothesis is true — this will be a conditional analysis.
- Evidence for or against the Working Hypothesis is orthogonal to the rest of my talk.

The orthogonal subquestion: our goals and values

Once we condition on the Working Hypothesis, another fundamental subquestion comes into view.

Goals and Values Question

What are the precise goals, objectives, and values of our mathematical community, and the enterprise of mathematical research?
Not just the explicit goals that we communicate to the public (or to funding agencies), but also the implicit goals that we actually seek in practice?

Goals and Values Question

What are the precise goals, objectives, and values of our mathematical community, and the enterprise of mathematical research?

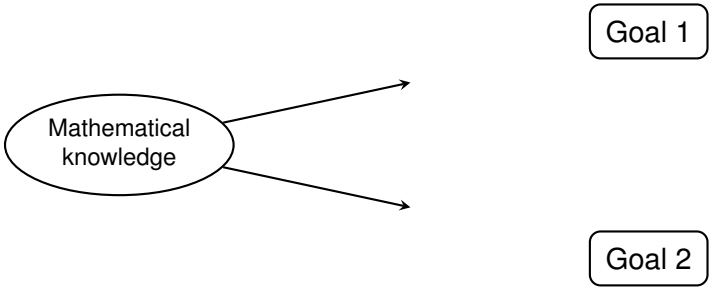
- In the past, we largely delegated this question to the humanities, and focused instead on the technical aspects of our profession.
- Assuming the Working Hypothesis, we will no longer have this luxury.
- However, I argue that a critical examination of this question will ultimately be highly valuable, **regardless of the status of the Working Hypothesis**.

What are our goals?

There are many reasons to justify mathematical research. To list just a few:

- To solve unsolved problems (both pure and applied).
- To develop new theories and techniques.
- To understand the world around us.
- To build a community of mathematicians.
- To train the next generation of mathematicians to guide its future directions.
- To contribute to the shared network of mathematical knowledge.
- To create enduring works of aesthetic value.
- etc.

- In the past, these goals have been largely positively correlated with each other: progress in one goal typically also led to progress on the others.
- As such, one could use one or two of these goals as proxies for the others.
- Many of the goals could be left implicitly stated only.

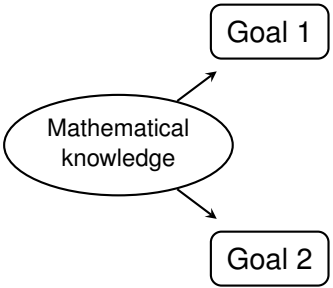


However, all metrics, when excessively optimized for, are at risk of being subjected to Goodhart's law:

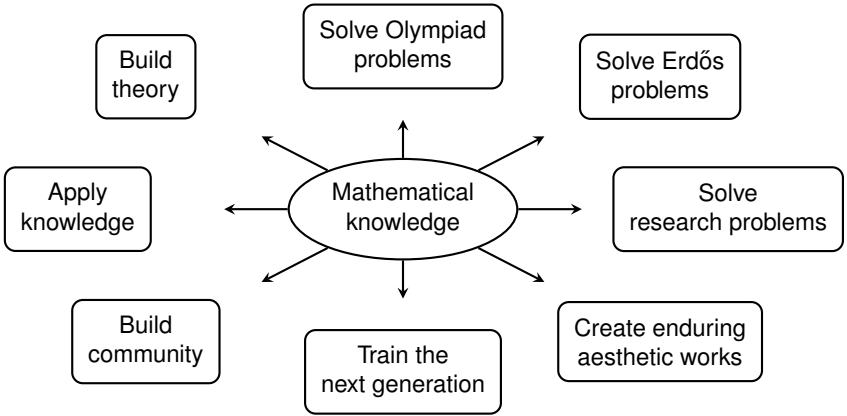
Goodhart's law (1975)

When a measure becomes a target, it ceases to be a good measure.

The inherently ungrounded nature of generative AI, as well as the financial incentives of AI companies, make the use of AI tools particularly vulnerable to this law.



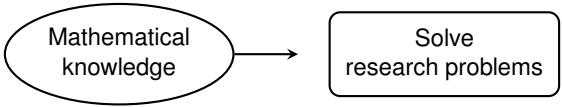
Excessive AI optimization may in fact cause the many (previously largely aligned) goals of mathematics to diverge[3] from each other.



[3]The diagram is extremely oversimplified; in particular, it should be much higher dimensional.

Case study: problem solving

- In this talk, I will illustrate this point with the problem-solving component of mathematical research.
- I will stress that this is not the only aspect of our profession.
- For instance, theory building is an important complementary aspect to problem-solving, which requires its own separate analysis.
- But problem solving seems particularly susceptible to being impacted by the Working Hypothesis.

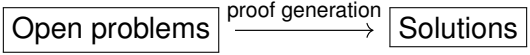


Suppose we specify the following goal:

Goal (first attempt)

Solve as many unsolved problems as possible.

We are now trying to optimize the flow in the following network:



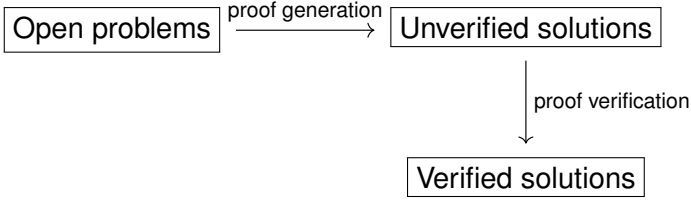
Even before the advent of AI, we already knew of a significant risk in optimizing this metric:

- One would see a large number of incorrect solutions to major open problems (e.g., the Riemann hypothesis).

So, we can update our goal:

Goal (second attempt)

Solve as many unsolved problems as possible, and verify them to be correct.



- Advances in AI and autoformalization (using proof assistant languages[4] such as Rocq, HOL, or Lean) have already significantly accelerated both proof generation and proof verification in many cases.
- Assuming the Working Hypothesis, this acceleration will continue to increase.
- But what if an AI tool generates a lengthy proof that nobody — not even the original humans prompting the AI — understands?

[4] The topic of formalization would require an entire lecture in itself, which we will not give here.

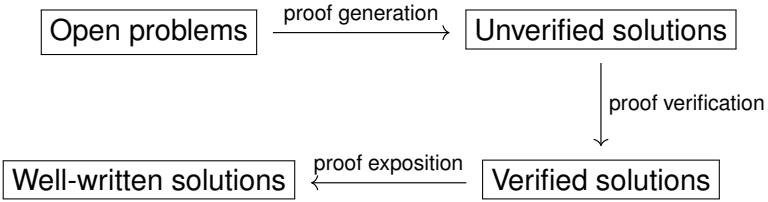
- Already, sites devoted to mathematical problems, such as[5] `https://www.erdosproblems.com`, contain dozens of AI-generated proof submissions.
- Many are likely to be correct, but no human expert has yet volunteered to verify and vouch for them.
- In several cases, even the human submitters have declared themselves unqualified to do so.
- Could we have a verified proof of a major result that **no** human understands enough to explain it?

[5] The emergence of flourishing online mathematical communities is another important topic that we will not be able to cover in this talk.

We can update our goal to elevate the role of exposition:

Goal (third attempt)

Solve as many unsolved problems as possible, verify them to be correct, and ensure that the results can be clearly communicated and understood by the mathematical community.

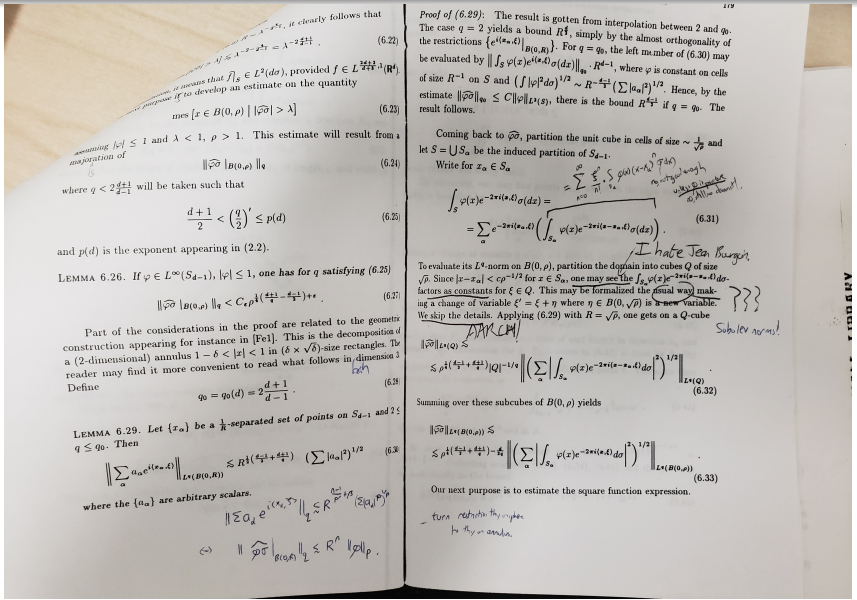


- Current AI tools have a very mixed record with proof exposition.
- On the one hand, the spelling, grammar, and formatting is close to flawless[6].
- However, the writing often dwells at length on trivialities, while passing very briefly through (or even obscuring) the most interesting and novel portions of the argument.
- AI-generated texts also often fail to note connections with prior literature, or provide high level overviews of the result.

[6] One can argue that they are *too* flawless.

- Proof exposition is a “fuzzier” optimization target than, say, proof verification.
- But the Working Hypothesis predicts that AI tools will soon be able to improve their proof exposition skills from current levels.
- However, even exposition can be over-optimized.
- A proof may end up being **too slickly written** — with both the routine and difficult parts of the argument being presented as equally easy to digest.

- In a human-written proof, the parts of the argument that the author found difficult will typically retain some natural friction that prompts the reader to slow down and pay more attention.
- An excessively AI-polished proof may remove both “artificial” and “natural” friction, without actually encouraging the reader to learn and understand the key ideas of the argument.
- Paradoxically, “mistakes” in human exposition can ultimately be helpful to the reader!



A 1991 paper of Bourgain, annotated by my much younger self.

William Thurston, “On proof and progress in mathematics” (1994)

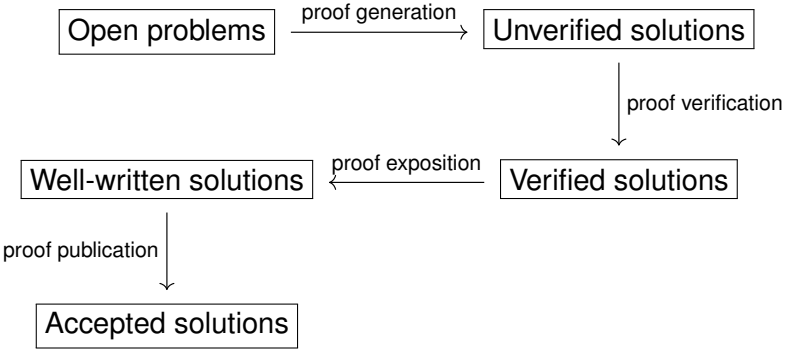
"We are not trying to meet some abstract production quota of definitions, theorems and proofs. The measure of our success is whether what we do enables people to understand and think more clearly and effectively about math."

- For a proof to be actually contribute to the broader field, it is **not enough** for it to be correct and easy to read.
- It also needs to be accepted and valued by the community.
- Other mathematicians need to digest the result and incorporate it into their own work.

- Authors can assist in the digestion process by describing their own insights and stories from when they were working on the problem.
- However, current AI tools are quite opaque about their problem-solving process.
- This is particularly true for proprietary models whose inner workings remain a corporate secret.

Goal (fourth attempt)

Solve unsolved problems, verify them to be correct, ensure they are clearly communicated, and have them digested and accepted by the mathematical community.



- Community acceptance of a result, by its nature, is slow and human.
- It can be encouraged with good exposition and careful writing.
- But it is ultimately an external process that **cannot be optimized purely by the authors and their AI tools**.

- Our current publication infrastructure relies on human editors and referees to voluntarily provide this community acceptance as a service.
- This work is often regarded as less prestigious than that of generating proofs in the first place.
- But it is an essential component of our profession.
- It is also how we convert the individual achievements of mathematicians into collective progress and understanding.

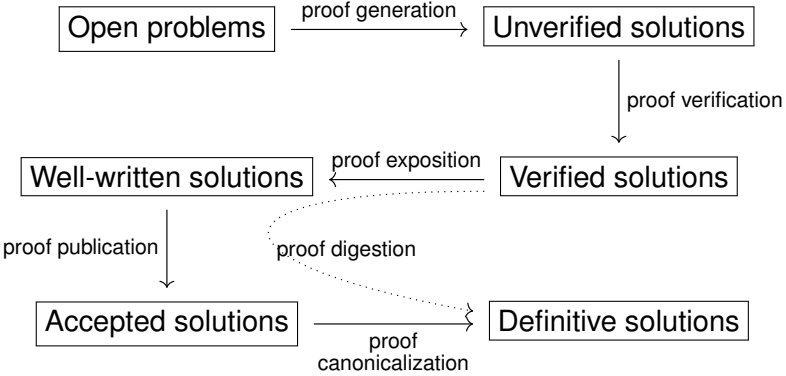
- AI evaluation tools may serve as helpful filters for community acceptance.
- For instance, journals could automatically reject papers that are flagged for inadequate verification or exposition.
- But such tools cannot substitute for community acceptance completely.

- In fact, even publication of a proof is **not always the final state of a solution**.
- Key results should ultimately become part of the definitive textbooks and reference material for the subject — taught to the next generation of students.

- Such canonicalization of a proof is the **slowest stage** of all.
- It requires broad, deliberative consensus by the community.
- It is the stage least amenable to optimization by AI tools.
- But it is the most valuable part of the entire process.
- Many applications only become feasible once the underlying mathematics has been fully digested.
- For instance, the success of AI tools in mathematics crucially relies upon the canonical theories that human mathematicians have painstakingly built over the centuries.

Goal (final attempt?)

Solve unsolved problems, verify them to be correct, ensure they are clearly communicated, and have them digested, accepted, and incorporated into the definitive theory of the field.



If the Working Hypothesis holds, then without suitable policy and cultural changes, significant "impedance mismatches" (or "proof indigestion") will emerge all throughout this process:

- AI-generated proofs will accumulate waiting to be verified.
- Many verified AI-generated proofs will await a readable writeup.
- AI-generated proofs, even when required to be correct and well-written, will overwhelm our traditional peer-review system.
- And even the published proofs will be too numerous for the community to work into a definitive form.

In short, we will transition from an era of proof scarcity to an era of proof abundance.
Some signs of this indigestion are already occurring; in fact, strains were emerging even before the advent of modern AI.

From goals to recommendations

Identifying the goals of problem solving helps realize how to respond to these emerging impedance mismatches.
The Leiden declaration `https://leidendeclaration.ai` – a topic of a lecture by Jim Portegies on July 26 – is an excellent starting point in this regard.

Leiden Declaration on Artificial Intelligence and Mathematics

2 June 2026 · DOI: 10.5281/zenodo.20302944

Preamble

Technological developments have repeatedly transformed the practice of mathematics. Recent artificial intelligence technologies, including symbolic and neural methods for the generation and formalization of mathematics, may already have initiated a significant chapter in this long history. Among researchers, artificial intelligence has produced a wide range of reactions: enthusiasm for its potential to yield new discoveries; intimidation by the pace of developments; indifference to these rapid changes; and concern for the implications, both for mathematics and in wider society.

Recommendations for individual mathematicians

1 Disclose tool use

Transparently disclose the use of automated tools, including large language models, machine learning systems, proof assistants, and other mathematical software. Include a "Tool and computational resource disclosure" section in your papers; many journals, publishers, and professional organizations have already developed guidelines for this, and though the precise form of such a section will necessarily evolve, we encourage authors to live up to the spirit reflected in the UNESCO Recommendation on Open Science and the FAIR principles. When acting as a reviewer, abide by publisher guidelines. If the use of artificial intelligence is allowed, be transparent about how you used it, and take responsibility for any significant recommendations you make.

Commentary: one needs to avoid the worst case scenario in which authors use AI tools covertly to aid their work, but conceal that usage to avoid criticism from peers. Normalize the responsible disclosure[7] of AI assistance.

[7] AI tools were used to autocomplete text and to generate diagrams in these slides.

2 **Support the needs of reviewing**

The use of artificial intelligence in preparing papers can introduce material that makes reviewing more demanding. Make it easier for your peers to review your work by disclosing tool use, giving precise and complete references to previous results, and providing formal proofs where feasible and appropriate.

Commentary: we need to decrease the emphasis on proof generation, and of being the "first" to solve a problem; and increase the emphasis on "proof digestion": exposition, publication, and canonicalization.

4 **Retain the responsibility for correctness**

When automated techniques are employed in published mathematical research, the responsibility for the correctness and adequacy of the arguments and results, as well as for the completeness and accuracy of citations to relevant prior work, remains exclusively with the human authors.

6 **Put effort into proper attribution**

The known limitations of automated tools in properly attributing ideas create a corresponding obligation for proactive effort to find and credit the sources that made a new result possible. Where a satisfactory attribution is not possible, state this explicitly in the publication.

Commentary: my suggested rule of thumb: if the authors cannot convincingly demonstrate that they can give a clear, expert-level talk on their results, that is correct and properly attributed, then the result should not be published.

Closing thoughts

- I presented problem solving as one aspect of mathematics where the Working Hypothesis forces us to inspect our goals and values.
- But the Working Hypothesis potentially impacts many other aspects of our work: teaching, mentoring, hiring, grant applications, public outreach, ...
- We should perform similar analyses for these aspects as well.

- In many areas (particularly in education and training), it will be crucial to emphasize the human aspect of our work, and tightly restrict the use of AI tools.
- In others areas, we will need to take the initiative on AI usage, and define best practices to incorporate these tools into our workflows on our own terms.
- We will also need new workflows and new infrastructures to complement our traditional ones, but this is an entirely different talk which I will not give here.
- Our community needs to come together to have open and honest discussions about both AI capability and our goals and values.

7 **Participate in public discourse**

Mathematicians have a responsibility to support serious science journalism and to engage in public discourse to explain and contextualize artificial intelligence-assisted methods and results. This is particularly important for work within our own subfields, where specialized knowledge is required to assess claims about the depth, difficulty, and significance of results. Moreover, we encourage mathematicians to seek opportunities to cooperate with and support other researchers and creative professionals facing similar challenges.

Some new workflows and infrastructures

- Mathlib: `https://mathlib.org/`
- Mathematical Discourse: `TBA`
- Erdős problems: `https://www.erdosproblems.com`
- Optimization constants database: `https://github.com/teorth/optimizationproblems`
- SAIR Competitions: `https://competition.sair.foundation/competitions`

Thanks to Bryna Kra, Jeremy Avigad, Martin Hairer, Akshay Venkatesh, and Emily Riehl for feedback on early versions of this talk.