Will mathematics exist in 2035?

W. T. Gowers

Collège de France and University of Cambridge

September 19, 2025

What I wrote in 1999

An apparently successful prediction.

... let me ... present an imagined dialogue between a mathematician and a computer in two or three decades' time. The idea of the dialogue is that the computer is very helpful to the mathematician, while not doing anything particularly clever. This represents an unthreatening intermediate stage between what we have now, computers that act as slaves doing unbelievably boring calculations for us, and full automation of mathematics. I have written the dialogue in English, but this is supposed to be a translation of a more formal language which has not yet been invented.

Rough structure and classification, in Visions in Mathematics, GAFA 2000 Special Volume, Part I.

A more general prediction of the impact of ATP

Finally, then, here is my guess about the impact of computers on mathematics over the next century. Although some of the difficulties involved in automatic theorem proving are formidable, a century is a very long time in computing, so I expect computers to be better than humans at proving theorems in 2099. I do not feel particularly happy about this, but I expect the impact on my own mathematical life to be entirely positive, because a semi-intelligent database of the kind I have described would be a wonderful resource and would take a great deal of the drudgery out of research. There might even be a golden age when computers were good at exercises but not yet good at having deep insights. Then, instead of wasting a week not noticing that a hoped-for lemma had a simple counterexample, one could get the computer to check it. In other words, computers would still do the boring bits for us, but these would not be quite as boring as they are now.

However, such a golden age, if it occurs, is unlikely to last for long. The next stage might be one where only a very few outstanding mathematicians could discover proofs that were inaccessible to computers. Perhaps others would concentrate on teaching, but computers would probably become better than us at this as well, at least if they themselves had been taught in anything like the way I have outlined. In the end, the work of the mathematician would be simply to learn how to use theorem-proving machines effectively and to find interesting applications for them. This would be a valuable skill, but it would hardly be pure mathematics as we know it today.

Some possibilities for the next ten years

Scenario	Likelihood
The external disaster	low but non-zero
The general singularity	hard to say
The mathematical singularity	hard to say
The steamroller	low to medium
The black box	low
The gradual takeover	(conditionally) high
The collapse of incentives	medium to high
The plateau	medium
The glorified calculator	low

Why might there be a plateau?

Proof finding is a computational problem that we do not understand very well.

What we lack (WD):

Theorem

Let Π be a proof with properties $P_1, \ldots, P_k$. Then it is feasible for human-like search methods to find Π.

But such a characterization (or more precisely, an objective measure of difficulty) appears to exist.

cf Sudoku puzzles

A hypothesis that would predict a plateau

Let $\mathcal{P}$ be following computational problem.

Problem. *The natural mathematical statement S has a "nice" proof. Find it.*

Hypothesis.

- The typical education of a research-level mathematician imparts in them a probabilistic algorithm that runs in expected time $f(n)$.
- The best AI training methods impart in computers a probabilistic algorithm that runs in expected time $g(n)$.
- g grows much more rapidly than f.

Weak evidence for the hypothesis

AlphaProof and its successors take *much* longer to solve IMO problems than human mathematicians (measured in steps).

What matters more is scaling. It would be interesting to attempt to measure this experimentally!

The limitations of the human brain put huge constraints on how we learn: the resource constraints on a computer are very different. (JA)

The way computers are currently trained to do mathematics is extremely different from the way humans are trained to do mathematics.

So maybe the chances of a steamroller (if no singularity) are lowish.

A hypothesis that would predict that the black box will not occur

Let $\mathcal{S}$ be the set of "nice", "interesting" mathematical statements.

Hypothesis.

- Finding "nice" proofs of statements in $\mathcal{S}$ is a *much* easier computational problem than finding completely opaque proofs that magically work.

Fairly strong evidence: it is quite clearly much easier for us!

So computers would tend to find nice proofs (when they exist).

Why is the glorified calculator unlikely?

The glorified calculator is very good at finding proofs, but needs human guidance about what is worth proving.

Hypothesis.

- It is not possible to be very good at finding proofs without also being good at formulating good questions.

Strong evidence for hypothesis.

- Watch yourself next time you try to find a proof!
- Think about how you ever manage to come up with good questions.

If correct, then a glorified calculator would only ever exist if you artificially restrained its ability to ask good questions.

What might happen in the next ten years?

Assume no singularity and that scaling up current approaches isn't enough.

If new approaches not found.

- Computers continue to improve but at diminishing rates, so that with enough training (and help from computers!) humans can still solve problems that computers can't.
- Many tools are built that make processes of formalizing proofs, finding proofs, testing hypotheses, performing clever AI-directed searches, etc. very much easier and more convenient, so we can work much faster.
- By 2035 the golden age is in full swing.

And if new approaches are found?

Assume additionally that our understanding of proof search improves to the point where we can teach computers to look for proofs with something much closer to human efficiency.

- As we teach computers how to solve various classes of problems, they become better than us at more and more mathematical tasks, including finding proofs, but this happens more quickly in some domains than others. (Gradual takeover.)
- The dream that many young people have had of solving a famous unsolved problem ceases to be realistic, so a large part of the romance of a mathematical career disappears. (Collapse of incentives.)
- Computers are not just good at solving (and posing) problems, but they are also good at diagnosing mistakes and misconceptions, because now they are actually thinking mathematically. So they become very good private tutors in mathematics.

Would this be good or bad?

Clearly problematic for many current professional mathematicians.

Less obvious whether in the long term it would be bad for society as a whole.

The uncertainties on which the outcome depends

1. Will there be a singularity (or some other AI breakthrough that changes everything)? Subquestion: will there be a mathematical singularity before a general singularity?

2. Are the proof-finding algorithms that computers naturally learn fundamentally less efficient than the ones that humans naturally learn?

3. Will attempts to understand better how to search for proofs efficiently (i) be successful and (ii) lead to much faster and more powerful algorithms?

What my group is working on

- A more precise idea of what “proof-finding moves” are.
- *A database of “motivated”, or “move-generated” proofs. (HM)
- *A platform that implements proof-finding moves.
- *Creating advanced Lean tactics to make Lean more helpful for finding proofs.
- Possible future work: training computers to choose moves.

*Funded by Renaissance Philanthropy’s AI for Math Fund.

More excerpts from the 1999 article

Many of the best proofs, or at least the best-written proofs, come with an indication of how they were discovered. To put that the other way round, proofs that depend on magic tricks, while they may be impressive and elegant, are not completely satisfactory as explanations.

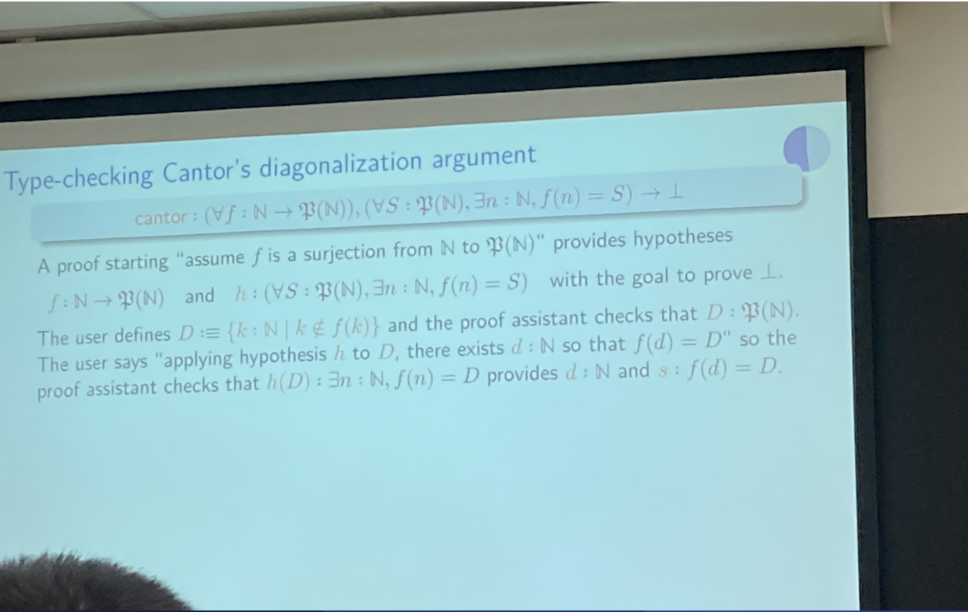
⋮ ⋮ ⋮ ⋮

I would like to ... adopt [this] as a working definition of "explanation" as opposed to "guarantee of truth". That is, I shall think of an explanation as a proof together with an account of how the proof was discovered. This account should not be something like "I was sitting in the bath and suddenly had the following brilliant idea", but rather a demystification of the proof: that is, a convincing demonstration of how a mere (mathematically trained) human being could search for a proof and find this one, without relying on excessive amounts of luck. (SDD)

One way of explaining a proof, in the above sense, is to show that it can be generated using what one might think of as standard mathematical reflexes. Some of these might be very general methods of approach, and others might be trying to apply well known techniques from the area of mathematics in question.

Examples mentioned: generalization, modification of known argument, spotting patterns in special cases, getting bored.

The Cantor rabbit



Where does the set in Cantor's theorem come from?

After a standard reduction (unfolding definitions etc.) we get to the following proof state.

$f : X \to \mathbb{P}(X)$

$\vdash \exists A \subset X \ \forall x \in X \quad f(x) \neq A$

Coming up with a suitable A

$\vdash \exists A \subset X \ \forall x \in X \ \ f(x) \neq A$

A very standard way to define sets is by using properties. So we could try to find a property P and let $A = \{y \in X : P(y)\}$.

(The ChatGPT prompt “I have a set X and I need to find a subset A of X with a certain property. Can you suggest any general methods for dealing with this kind of situation?” elicits this as its second suggestion.)

New goal.

$\vdash \exists P : X \to \texttt{Bool} \ \forall x \in X \ \ f(x) \neq \{y \in X : P(y)\}$

$$\vdash \exists P : X \to \texttt{Bool}\ \forall x \in X \quad f(x) \neq \{y \in X : P(y)\}$$

After standard manipulations, this becomes

$$\vdash \exists P : X \to \texttt{Bool}\ \ \forall x \in X\ \ \exists z \in X$$
$$(z \in f(x)\ \wedge\ \neg P(z))\ \vee\ (z \notin f(x)\ \wedge\ P(z))$$

or even

$$\vdash \exists P : X \to \texttt{Bool}\ \ \forall x \in X\ \ \exists z \in X \quad z \notin f(x) \iff P(z)$$

By far the most obvious choice for z is x. So let's try it.

$$\vdash \exists P : X \to \texttt{Bool}\ \ \forall x \in X\ \ x \notin f(x) \iff P(x)$$