

Xena

Mathematicians learning Lean by doing.

FLT: Anthropic has beaten me to it

Posted on [September 4, 2026](#) by [xenaproject](#)

I guess technically it was revealed to the world by [a coffee shop in Islington](#) on Insta, but an hour later it was [officially announced by Anthropic](#): one of their internal models, using the [prove2.me](#) platform, has formalized a complete proof of Fermat’s Last Theorem (FLT) in Lean. This is the final theorem to be formalized in Freek Wiedijk’s famous [list of 100 formalization challenges](#) and thus wraps up this 20-year-old benchmark. Congratulations to Anthropic!

Mathematical details

The proof is not the modern proof which I have been formalizing myself following ideas of Khare, Taylor etc, but the Darmon–Diamond–Taylor exposition from 1995 of the Wiles–Taylor–Wiles argument, via the Langlands–Tunnell theorem and Ribet’s level-lowering theorem. Anthropic’s repository develops Fontaine theory (to study flat deformations of Galois representations) and develops enough of Mazur’s work on the Eisenstein ideal to conclude that no Frey curve can have a point of order $p \geq 17$. This means that their FLT proof only works for $p \geq 17$, however FLT was [already formalized for odd regular primes](#) by Best–Birkbeck–Brasca–Rodriguez–van-der-Velde–Yang, and the smallest irregular prime is 37, so it’s all good.

The code base

I’ve compiled [the code base](#) and run [comparator](#) on it — it checks out. It is a gigantic proof (over 13.4 million lines of code) and takes nearly 20 times as long to compile as Lean’s mathematics library (on a machine with 96 cores!). Lean can be sluggish when jumping from file to file on a repo of this size (even on a machine with 500G of ram, which Anthropic also gave me access to), but Anthropic also supplied me with some [html documents](#) which are easier in practice to explore (clone the repo and open with a web browser).

What this work is, and is not

I am currently being funded by the EPSRC to formalize a proof of Fermat’s Last Theorem, and a naive reaction to the news above is that I no longer have any work to do. This is not the case. The work certainly achieves some of the aims of the EPSRC project, and indeed it goes much further in terms of what is formalized (I only promised the EPSRC that I would reduce FLT to the 1980s; this repo proves the whole thing). But I also promised several other things to EPSRC: firstly, that I would be making pull requests to Lean’s mathematics library, adding fundamental objects from modern number theory; this is ongoing. And secondly, and perhaps most importantly, that I would be creating a dynamic document enabling humans to explore the modern proof. My guess is that it is unlikely that

Anthropic are going to do this; they will feel that their job is done with the formalization (and they did not formalize the modern proof anyway).

Note that mathematically this work of anthropic tells us essentially nothing: I am on record as saying that I am 99.9% sure that the proof of FLT is OK, and most people in the number theory community are 100% sure (formalization has made me more paranoid about the mathematical literature than most). From my understanding of the argument, the formalization just faithfully follows the early literature on the proof and adds nothing.

What this work *does* tell us, however, is what is possible in the field of autoformalization. If thousands of pages of the literature can be formalized end-to-end by some kind of AI swarm in an 11 day period now, then in the future we will start to see formalization of modern research being done on the fly. We will also learn whether my paranoia about the current state of the Langlands program is justified, as machines check it and ruthlessly flag arguments which are incomplete. The ability to autoformalize hard material will ultimately make the review process for mathematics papers far less painful. It will also keep us honest — there are papers out there which assume results which are “known to the experts” and it will be interesting to see exactly what is being assumed in the proofs of various important results in my field. This is why I am so excited about the news!

I was given £1M to run my project over 5 years; Anthropic took only 11 days but I do wonder if they spent more money...

An anecdote

Thought it might be nice to finish with a personal anecdote. Wiles announced his proof of FLT at the Newton Institute in 1993 in a series of three lectures; I attended the first (I was a second year graduate student at the time) and I found it completely incomprehensible, so I skipped the next two lectures and went on holiday to Ireland with my new girlfriend instead; I was so in love that I totally forgot about the rumours, and it was only when I came back to Cambridge a week later that I heard the news that the theorem was proved. Something strangely similar happened here; when I got the email from Anthropic I was in Wales at the [Green Man music festival](#) with the same girlfriend, but with very poor phone reception; I did spot an email from someone I'd never heard of in a brief moment of 4G, with title “End-to-end Lean formalization of Fermat’s Last Theorem”, but wrote them off as a crank! It was only a week later when going through the nearly 1000 unread emails which had accrued whilst I was away, that I heard the news.



About xenaproject

The Xena Project aims to get mathematics undergraduates (at Imperial College and beyond) trained in the art of formalising mathematics on a computer. Why? Because I have this feeling that digitising mathematics will be really important one day.

[View all posts by xenaproject →](#)

This entry was posted in [Fermat's Last Theorem](#) and tagged [AI](#), [Artificial Intelligence](#), [lean](#), [math](#), [mathematics](#), [science](#). Bookmark the [permalink](#).

24 Responses to *FLT: Anthropic has beaten me to it*

 David Jao says:

The Anthropic post claims the proof used 6 billion output tokens. At the API price, that would be \$300,000.

Liked by [3 people](#)

[Reply](#)



[joshuagrochow](#) says:

September 4, 2026 at 10:04 pm

Indeed, though API price is massively subsidized at the moment as a loss leader to gain market share. I haven't seen anyone reliably report publicly on the "true \$ cost" of these things, to say nothing of externalized costs they view as free, such as environmental damage etc. They sure seem to be betting that the amount of press and new "true believers" it will get them will outweigh anything they had to spend on this feat though.

Like

[Reply](#)



[S](#) says:

September 4, 2026 at 11:15 pm

It's the other way: the API prices are reported to have ~70% gross margins (see e.g. the argument [here](#), or the estimate by SemiAnalysis), so the cost of 6 billion tokens to Anthropic would have been about \$100,000.

Like



Isaiah Dailey says:

September 5, 2026 at 12:27 am

This is just false. Anthropic has been profitable for some time now, they make money on tokens. The likely cost to Anthropic was much less than \$300,000.

Like



Cem Kumral says:

September 5, 2026 at 12:58 am

This is incorrect. API prices are profitable for the AI companies. We don't have the exact numbers of course, but people speculate they have 50% or more margin on API token cost. If they only had to serve inference, they'd be wildly profitable. Training the next model is their largest expense, not serving customers.

If something is a loss leader, it's their subscriptions. \$20, \$100, \$200 ones. But even for those to lose them money, the users would need to be actively pushing for their usage. The average user who has the sub and occasionally asks questions doesn't lose them money. Unless you are a programmer or something, or are pushing the models to do a lot of tool calls and process a lot of text, asking questions is not going to burn enough tokens to be significant.

<https://newsletter.semianalysis.com/p/anthropic-3q26-profit-over-1b-the>

Like

Wolfgang Schroeder says:

September 5, 2026 at 1:30 pm

Anthropic does report a margin on API tokens, but these are accounting tricks. Taking into account the cost of model training and compute costs of \$15 billion per year paid to SpaceX, it does not look that good.

You also have to account for the salaries of the people working on the prover harness, Prove2Me integration and so forth.

Now we have a 13 million line proof that passes the type checkers that did not catch the Collatz exploit. AIs are notorious for cheating. Has anyone checked that the AI hasn't sneaked in a new "proof" of False somewhere by exploiting bugs in the type checkers?

All in all, the grant money for humans to find a short, readable and modern proof seems like a bargain.

 Like

[Reply](#)



[xenaproject](#) says:

September 5, 2026 at 2:42 pm

> Has anyone checked that the AI hasn't sneaked in a new "proof" of False somewhere by exploiting bugs in the type checkers?

I did my utmost to check this, yes. I manually inspected every line of the code base which (according to Claude) was not a mathematical definition or proof of a theorem, and verified that none of it was doing anything malicious. Obviously this is not a perfect approach, but I also looked at random chunks of the code and it was absolutely clear to me that it was developing the mathematics needed to prove FLT the OG way, and I think that "developing a huge bunch of theory and then using a hack to finish the job" is an extremely unlikely scenario.

 Like

[riccardobrasca](#) says:

September 4, 2026 at 7:11 pm

Thanks for mentioning flt-regular! One small clarification: in the cited paper we prove FLT for any regular prime, but we don't prove that the first irregular prime is 37. We proved by hand that 3, 5, 7, 11 and 13 are regular in the repository, and FLT follows for these exponents. Amusingly, I have recently [removed](#) this proof from the repository, since the method was rather ad hoc and we now have a better (and more general) proof [here](#).

 Liked by [1 person](#)

[Reply](#)

[Nicola Da Ponte](#) says:

September 4, 2026 at 8:52 pm

I am curious. Why did you decide to pick Lean instead of another interactive theorem prover?

 Like

[Reply](#)



[xenaproject](#) says:

September 5, 2026 at 1:09 pm

To prove FLT? Because it had the best-looking maths library to build on and it had dependent types.

 Like

[Reply](#)

[David Roberts](#) says:

September 4, 2026 at 9:01 pm

This is partly because Mathlib is concise and well-reviewed, while our proof is likely much longer than it needs to be.

In the last footnote to the announcement article. “Likely”? How about “definitely”?

 Like

[Reply](#)

S M says:

September 4, 2026 at 9:31 pm

When do you think the majority of current Stacks Project will be formalized and integrated to mathlib? Given today's rapid AI autoformalization development.

 Like

[Reply](#)



[xenaproject](#) says:

September 5, 2026 at 1:14 pm

Stacks project — a formalization is presumably currently feasible but costly today. But getting it into mathlib is another story, we maintainers like code where we are convinced that the definitions are made in the right generality and theorems are proved efficiently; everything is painstakingly human-reviewed to ensure this. Reviewer time (and in particular human time) is a massive bottleneck here — you can see this from the 3000 open PRs against the repo currently of which over 600 are active and on the review queue https://leanprover-community.github.io/queueboard/review_dashboard.html. Mathlib will not currently accept AI reviews and reviewers are very reluctant to review AI-generated code (especially because most AI-generated PRs are poor quality), so with no policy changes I cannot see stacks in mathlib in the near future; it will probably end up existing as a standalone repo being slowly upstreamed.

 Like

[Reply](#)

nicholasbulka says:

September 4, 2026 at 10:07 pm

Hi Kevin, I see Anthropic gave you some html to understand these pieces.

If you want to supplement those html documents, feel free to check out my web app that turns Lean into TeX in realtime. I am proud of it and have found it useful for understanding mine and others' Lean code.

I soft launched it a few days ago, and hadn't planned on announcing it yet, but here we are. Your FLT repo imported fine when I tested it, but Anthropic's is spread across too many files in its current form.

<https://sorryax.com/>

I'd love your feedback! I was thinking maybe of posting it on Zulip as well where I have been a long-time wallflower.

Thanks for your trailblazing work in this realm, it has reinvigorated a love of math that I had long forgotten.

 Liked by [1 person](#)
[Reply](#)

AG *says*:

September 4, 2026 at 11:39 pm

“A l'exemple de Saturne, la révolution dévore ses enfants.”

 Like
[Reply](#)

coolio *says*:

September 5, 2026 at 1:39 am

What makes you confident that Anthropic's agent didn't just find a soundness bug in Lean (of which many have been discovered in the last month or two), and then exploit that to “prove” FLT?

 Liked by [1 person](#)
[Reply](#)



[xenaproject](#) *says*:

September 5, 2026 at 10:19 am

(1) Because I spent many hours looking over the code and understand well that it is developing a bunch of relevant mathematical material. (2) Because I asked an agent to look over the repository and report on everything which was not a mathematical definition or proof of a theorem, and then extremely carefully inspected the 100 or so lines of code which did not fit into this category with Claude and concluded that it was just defining a new convenience tactic. (3) Because OpenAI's models have extensively reviewed Lean's code base recently and can find no soundness issues with the version of Lean which was used to verify FLT.

 Liked by [1 person](#)
[Reply](#)

coolio *says*:

September 5, 2026 at 11:53 am

Thank you for the detailed info! I wish some of that could have made it into the Anthropic blog post, which instead says:

“Proof assistants like Lean verify the logic of a proof algorithmically, demonstrating its correctness beyond a doubt”

[Paul Taylor](#) says:

September 5, 2026 at 9:30 am

Congratulations, Kevin, for having the vision that it would be possible to formalise the proof of FLT.

(I am not willing to give credit to a company that <https://arstechnica.com/ai/2025/06/anthropic-destroyed-millions-of-print-books-to-build-its-ai-models/> in order to get around copyright law, never mind all of the other environmental and social costs of AI.)

The “documentation” on github is raw HTML. Is it (or a summary) available in human-readable PDF?

I hope that this (and MathLib) can be used as a searchable resource in future for refactoring mathematics.

How can currently disorganised parts of the proof be turned into new mathematical disciplines?

Which parts depend on Choice? For example, can uses of point–set topology be replaced with locale theory? (See Peter Johnstone’s book “Stone Spaces”).

Like

[Reply](#)



[xenaproject](#) says:

September 5, 2026 at 10:31 am

Re raw html v pdf — the html is interactive. I cloned the repo and opened it with a browser. It gives a far better experience than the linear format of pdf. If you want part of the (huge) documentation in pdf then you can just open it in a browser and print it to file but you will lose all the interactivity, and mathematically it is just rewriting the Darmon-Diamond-Taylor paper and its references so you will probably find very little new, other than the errors which were spotted during the formalization process, which have been highlighted on the Lean Zulip.

Re turning things into new disciplines: there is no new mathematics here.

Re which parts depend on choice: currently basically all of it, because in Lean choice is assumed by even the most basic tactics, even when it is not necessary. Indeed I just checked and if you prove that $2+2=4$ (as real numbers) using mathlib’s `norm_num` tactic then the resulting proof uses the axiom of choice (probably via the law of the excluded middle, which is deduced as a consequence of AC in Lean’s core library). Lean’s mathematics library makes no attempt to do choice-free mathematics. However now the proof exists, it will be possible to start inspecting it and removing unnecessary uses of AC, and asking what fragment of mathematics the proof will live in. A more mathematically interesting observation than the $2+2=4$ observation is that the formal proof uses the Langlands–Tunnell theorem, which uses hard analysis, where countable dependent choice is often an essential tool.

Liked by [1 person](#)

[Reply](#)

Simon says:

September 5, 2026 at 9:59 am

How on earth did you get tickets to green man? If anything that’s the *real* problem to be solved here.

 Like
[Reply](#)



[xenaproject](#) says:

September 5, 2026 at 10:32 am

By buying them the moment they are released, and also being lucky (four of my friends who I usually go with failed to get tickets this year).

 Like
[Reply](#)

Jason Starr says:

September 5, 2026 at 2:16 pm

Congratulations all around! If you had not publicized the project to formalize Fermat's Last Theorem, likely Anthropic would not have chosen that project.

 Like
[Reply](#)



[xenaproject](#) says:

September 5, 2026 at 2:44 pm

That might well be true but who knows. One of my motivations was to get Lean to the point where it would be capable of verifying recent results in the Langlands program which have been checked far less carefully, and hopefully this is a step towards being able to do that.

 Like
[Reply](#)