



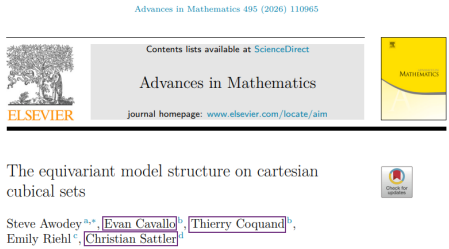


Emily Riehl

Open questions in autoformalization (of HoTT)



The old way of doing math research



Main theorem proven in 2019
Preprint posted in 2024
Published in 2026:
113 pages
plus a computer formalization in Agda



Project started in 2017
Preliminary results reported in 2018
Resumed work in 2023
Preprint posted in 2026
Submitted for publication: 64 pages

The new way of doing math research



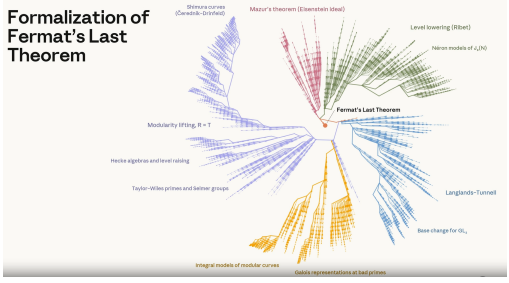
Acknowledgments. The disproof of the weak chromatic splitting conjecture given in this note was produced autonomously by Claude. We did not contribute any ideas and Claude was not prompted on this problem but on Freyd's generating hypothesis in stable homotopy theory. Although the relationship between Freyd's generating hypothesis and the chromatic splitting conjecture is well established, for example, in work of Hopkins–Devinatz [Dev90, Dev98] (see also [Ho07, Bar20]), we did not expect Claude to find a counter example to the weak chromatic splitting conjecture, which we both believed to be true. We note recent progress on Freyd's generating hypothesis recently appeared in Ma–Xu [MX26].

The result arose in the context of paid consulting with Anthropic. Claude was run internally by Anthropic, using its own computing resources. The transcript was shared with us September 8, 2026. The mathematical exposition was drafted by us, Agnès Beaudry and Tomer Schlank, after we verified the details of Claude's disproof, on September 12, 2026. The result was announced at 2026 Clay Research Workshop on Telescopic Homotopy Theory and Algebraic K-theory, on September 21, 2026. This text is not AI generated. It was produced by us, using Claude only for copyediting.



Advances in autoformalization

September 4, 2026



We are sharing the first complete computer-checked proof of Fermat's Last Theorem. Claude worked largely autonomously over 11 days to write the proof in the Lean programming language. Below, we describe how the formalization was done and share some thoughts about what this work could mean for research mathematics.

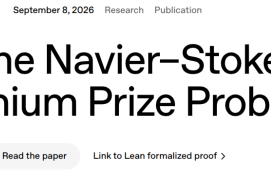


FLT: Anthropic has beaten me to it

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!

September 8, 2026



On Tuesday, September 1, we heard rumors that two Millennium Prize problems had been resolved. Inspired by these rumors and by the step change in performance of our internal model, we launched an effort to evaluate it on all open Millennium Prize problems and a few other high-impact problems.

We used a system of coordinating agents powered by our internal model. The agents had access to tools such as the ability to read from a cached version of the internet and the ability to run code. Agents were subdivided into groups with the ability to communicate within the group. The groups varied in size, and the group that produced the Navier–Stokes resolution involved on the order of 10,000 concurrent agents. At all times we maintained the same strict safeguards that we apply to all our frontier model evaluations, including monitoring and isolation.

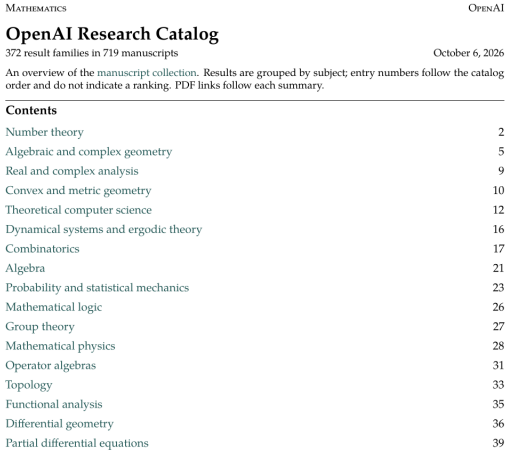
...

The agents arrived at their resolution on Saturday, September 5, about 88 hours after the first agents were launched. Lean formalization and verification took an additional 17 hours via GPT-6 Astra.

Across all attempted problems, the agents sent 4.9 million messages and used about 300 billion output tokens. In the process of resolving the Navier–Stokes problem, the agents sent 2.7 million messages and used approximately 130 billion output tokens.

October 6, 2026





MATHEMATICS OPENAI

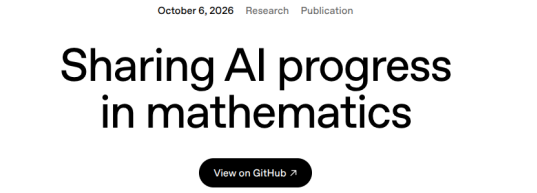
OpenAI Research Catalog

372 result families in 719 manuscripts October 6, 2026

An overview of the manuscript collection. Results are grouped by subject; entry numbers follow the catalog order and do not indicate a ranking. PDF links follow each summary.

Contents

Number theory	2
Algebraic and complex geometry	5
Real and complex analysis	9
Convex and metric geometry	10
Theoretical computer science	12
Dynamical systems and ergodic theory	16
Combinatorics	17
Algebra	21
Probability and statistical mechanics	23
Mathematical logic	26
Group theory	27
Mathematical physics	28
Operator algebras	31
Topology	33
Functional analysis	35
Differential geometry	36
Partial differential equations	39



October 6, 2026 Research Publication

Sharing AI progress in mathematics

View on GitHub ↗

As of today, just under half (344/721) of the manuscripts have some relevant Lean formalizations of either the main theorem or of a significant result.[a]

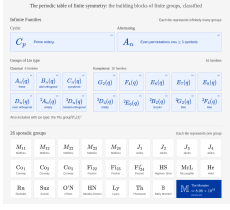
[a] That is 48% of these papers come with a comparator challenge.

October 26, 2026 by Axiom

Formalizing the Enormous Theorem

Axiom has formalized in Lean the classification of finite simple groups, the longest proof in the history of mathematics.





Today, we are releasing the first complete computer-checked proof of the Classification of Finite Simple Groups (CFSG). One of the major achievements of twentieth-century mathematics, the classification identifies the basic building blocks of finite groups. Widely considered to be the longest mathematical proof ever written by humans, the original proof of CFSG extends across roughly 15,000 pages in more than 500 articles by about 100 mathematicians, earning CFSG the nickname “the enormous theorem.” For his crucial work on the classification, John Thompson was awarded the Fields Medal (1970) and the Abel Prize (2008). Over 24 days, engineers and mathematicians at Axiom Math directed swarms of dozens of AI agents in a project that produced approximately 43.1 million lines of code in the Lean proof assistant.

The future?

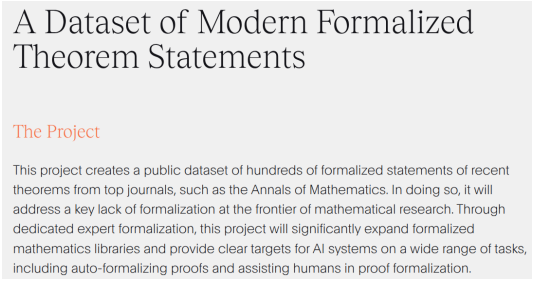


The future [of (auto)formalization] is already here — it's just not very evenly distributed.

The future?



The future [of (auto)formalization] is already here — it's just not very evenly distributed.



A Dataset of Modern Formalized Theorem Statements

The Project

This project creates a public dataset of hundreds of formalized statements of recent theorems from top journals, such as the Annals of Mathematics. In doing so, it will address a key lack of formalization at the frontier of mathematical research. Through dedicated expert formalization, this project will significantly expand formalized mathematics libraries and provide clear targets for AI systems on a wide range of tasks, including auto-formalizing proofs and assisting humans in proof formalization.

A team of Lean experts formalized 50 theorem statements — the "easiest 50" — from the roughly 200 papers published in the *Annals of Mathematics* since 2020.

OpenAI paper by area	# Lean	% Lean
Functional analysis	18 / 19	95%
Logic	7 / 8	88%
Theoretical CS	59 / 74	80%
Combinatorics	40 / 50	80%
Convex/metric geometry	24 / 30	80%
Group theory	16 / 22	73%
Operator algebras	16 / 31	52%
Real/complex analysis	14 / 27	52%
Algebra	14 / 29	48%
Dynamics/ergodic theory	9 / 19	47%
PDE	13 / 29	45%
Probability/stat. mech.	41 / 105	39%
Mathematical physics	23 / 59	39%
Differential geometry	17 / 49	35%
Number theory	18 / 58	31%
Algebraic/complex geom.	12 / 86	14%
Topology	3 / 26	12%
All areas	344 / 721	48%

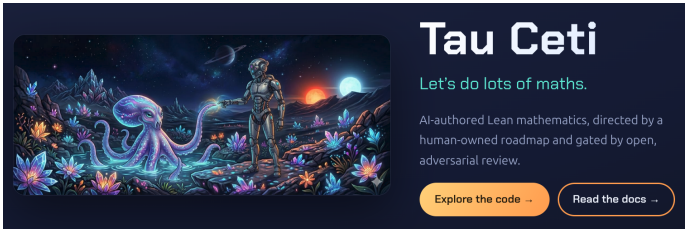
How to formalize “all” of mathematics



In a blog post, Jeremy Avigad, Kevin Buzzard, Johan Commelin, Rémy Degenne, Kim Morrison, Oliver Nash, and Adam Topaz outline concerns:

> *If spending millions of dollars to formalize mathematics results in libraries nobody uses, it will be a terrible waste of resources. Worse, the project could disincentivize good research and productive use of AI, disrupt the social norms and practices research depends on, and set the field back decades. A grand failure will make it hard, or even impossible, to recover later. A lousy formalization of existing results is worse than no formalization at all.*

They then describe positive models such as Tau Ceti, Formal Frontier, and Palomar.





2

Autoformalizing homotopy type theory

A challenge from the Lean FRO

An open challenge from June was to autoformalize the calculation of $\pi_3 S^2$, characterizing maps from the 3-dimensional sphere to the 2-dimensional sphere:



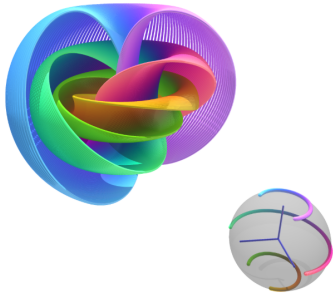


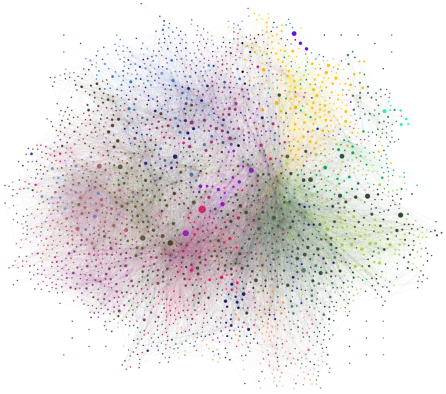
Image from an animation by Niles Johnson

The Agda-Unimath library



Welcome to the agda-unimath project. This is a community-driven effort aimed at formalizing mathematics from a univalent point of view using the dependently typed programming language Agda.

The library project was created by Elisabeth Stenholm, Jonathan Prieto-Cubides, and Egbert Rijke, and is currently being maintained by Egbert Rijke, Fredrik Bakke, and Elif Uskuplu. Our goal is to create an online encyclopedia of formalized mathematics containing an extensive curriculum of topics from a univalent point of view. We think libraries of formalized mathematics have the potential to be useful, and informative resources for both working and learning mathematicians. Our library is designed to work towards this goal, and we welcome contributions to the library within any topic in mathematics.



foundation elementary-number-theory group-theory category-theory order-theory real-numbers
synthetic-homotopy-theory metric-spaces univalent-combinatorics ring-theory linear-algebra commutative-algebra
structured-types graph-theory orthogonal-factorization-systems trees globular-types species lists set-theory
logic higher-group-theory finite-group-theory modal-type-theory complex-numbers universal-algebra reflection
analysis finite-algebra type-theories functional-analysis domain-theory synthetic-category-theory
wild-category-theory organic-chemistry literature spectral-theory

A dependency graph of the library, color coded by namespace. Fredrik Bakke. 2025 — perpetuity

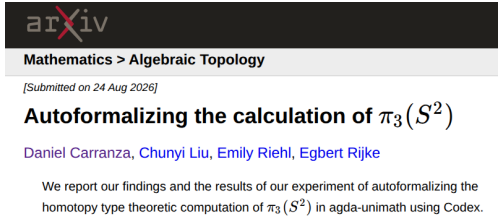
Autoformalization with Codex



Codex completed the autoformalization between June 3–June 25, 2026 (with repeated prompting) and also autoformalized various missing prerequisites:

(i) the definition and basic theory of pointed fiber sequences,
(ii) the long exact sequence of homotopy groups from a pointed fiber sequence,
(iii) the isomorphism that arises from a four term exact sequence with zeros at its endpoints,
(iv) the Hopf fibration $A \to_* A * A \to_* \Sigma A$ associated to a connected H-space A,
(v) the connected H-space structure on S^1,
(vi) the Hopf fibration $S^1 \to_* S^3 \to_* S^2$ arising from equivalences $S^3 \simeq S^1 * S^1$ and $S^2 \simeq \Sigma S^1$,
(vii) the computation $\pi_n S^1 \cong 0$ for $n > 1$,
(viii) the Freudenthal suspension theorem as a corollary of the Blakers–Massey theorem (which was also missing),
(ix) the isomorphisms $\pi_{n+3}S^{n+3} \cong \pi_{n+2}S^{n+2}$ for $n \geq 0$ that follow from Freudenthal, together with a separate base case[7] $\pi_2 S^2 \cong \pi_1 S^1$,
(x) the computation $\pi_n S^n \cong \mathbb{Z}$ for $n > 1$ (with the $n = 1$ case already in the library),
(xi) the deduced isomorphism $\pi_3(S^2) \cong \mathbb{Z}$ arising from the exact sequence

$$0 \cong \pi_3(S^1) \to \pi_3(S^3) \to \pi_3(S^2) \to \pi_2 S^1 \cong 0.$$



arXiv

Mathematics > Algebraic Topology

[Submitted on 24 Aug 2026]

Autoformalizing the calculation of $\pi_3(S^2)$

Daniel Carranza, Chunyi Liu, Emily Riehl, Egbert Rijke

We report our findings and the results of our experiment of autoformalizing the homotopy type theoretic computation of $\pi_3(S^2)$ in agda-unimath using Codex.

For the most part, natural language sources (specifically, the HoTT Book) were a sufficient guide, but when Codex got stuck, we pointed it to formalizations in the Coq-HoTT library by Dan Christensen (join associativity) and Mike Shulman (the long exact sequence and Blakers–Massey).

Open question: library impacts



The effort to import the AI-generated formal proofs back into its supporting library continues:

4.3 Library impacts As noted in the introduction, we do not consider an autoformalization project to be completed until it makes a commensurable contribution to the supporting library, in this case agda-unimath. The raw Codex generated code in our repository does not yet meet the standards for an agda-unimath PR, much less the standards of the library. It is insufficiently documented, insufficiently explained, and far too verbose. The formalization code also needs thorough review by us first, and later by library maintainers.

One motivation for writing this report is to facilitate communication with the maintainers of the agda-unimath library. We hope to discuss the possibility of extracting library-quality code to contribute to the library over a series of PRs in a way that is sensitive to the amount of human labor that the review process will require, in particular by contributing our own efforts to the refereeing process. As we view this to be of paramount importance to the evolving interactions between generative AI and computer formalization of mathematics, we plan to update this preprint with the results of our planned future conversations with the broader agda-unimath community.

New issue

Some initial thoughts regarding autoformalization policies for agda-unimath #1959

Open

EgbertRijke opened on May 15 Collaborator ···

As AI's are quickly becoming more capable at programming and formal verification, we should think about our position as a community of formalizers.

Open questions in autoformalization



What is the future of formalization of HoTT/synthetic mathematics in an age of AI?

- What form of AI-contribution policy best supports our existing libraries?
- Should we have an AI-generated AI-reviewed HoTT Tau Ceti — perhaps "Vega"?
- Autotranslation between libraries and between proof assistants is certainly within scope. Is this a good idea or does it impair HoTT's naturally evolved biodiversity?
- Should we be building more proof assistants?
- Should we be building deep embeddings into a future formal library of "all" mathematics?
- How do we make the case for aesthetics of formalized proofs?