Mathematics, AI, and Formalization

Seewoo Lee, UC Berkeley & Axiom Math

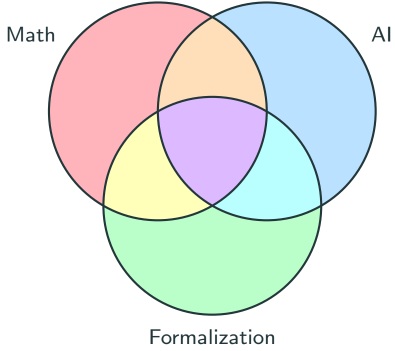
Feb 6, 2026. MAT280 at UC Davis

Goal

Many people start to talk about AI for mathematics, formalization, Lean, `ChatGPT` doing mathematics, etc. But I found that the distinction between these is often unclear.

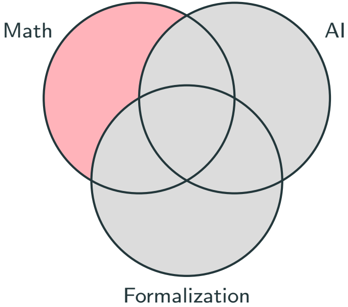
Goal

Today, I will introduce many examples of recent progress that fit into the below Venn diagram (mostly on intersections):



Disclaimer: Only a few of the works to be introduced are done by myself, and there could be incorrect explanations of others' works.

Mathematics



What is mathematics?

From Wikipedia

Mathematics is a field of study that discovers and organizes methods, theories, and theorems that are developed and proved for the needs of empirical sciences and mathematics itself.

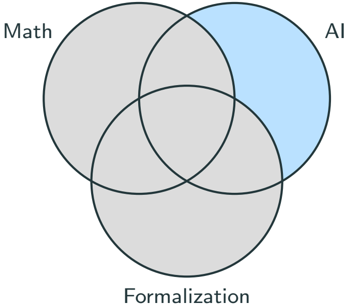
What is mathematics?

From Wikipedia

Mathematics is a field of study that discovers and organizes methods, theories, and theorems that are developed and proved for the needs of empirical sciences and mathematics itself.

We all do math. We all have fun.

Artificial Intelligence



What is Artificial Intelligence?

From Wikipedia

Artificial intelligence (AI) is the capability of computational systems to perform tasks typically associated with human intelligence, such as learning, reasoning, problem-solving, perception, and decision-making.

What is Artificial Intelligence?

From Wikipedia

Artificial intelligence (AI) is the capability of computational systems to perform tasks typically associated with human intelligence, such as learning, reasoning, problem-solving, perception, and decision-making.

These are all AI:

- `ChatGPT`, `Gemini`, `Claude`, `GitHub Copilot`
- `Moltbot`, `AlphaGo`

What is Artificial Intelligence?

From Wikipedia

Artificial intelligence (AI) is the capability of computational systems to perform tasks typically associated with human intelligence, such as learning, reasoning, problem-solving, perception, and decision-making.

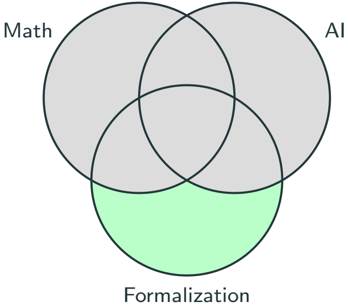
These are all AI:

- `ChatGPT`, `Gemini`, `Claude`, `GitHub Copilot`
- `Moltbot`, `AlphaGo`

but also these too:

- Logistic regression, Decision tree, SVM, . . .
- ResNet, YOLO, BERT, JEPA, . . .

Formalization



What is Formal Verification?

From Wikipedia

In the context of hardware and software systems, formal verification is the act of proving or disproving the correctness of a system with respect to a certain formal specification or property, using formal methods of mathematics.

Use **machine-checkable proofs** to guarantee correctness.

Formal Verification in Industry

Formal verification is widely used in critical systems:

- **Hardware**: Intel CPU verification, AMD, ARM
- **Aerospace**: NASA, Airbus flight control systems
- **Cryptography**: Amazon s2n (TLS), EverCrypt
- **Compilers**: CompCert (verified C compiler)
- **Operating Systems**: seL4 (verified microkernel)

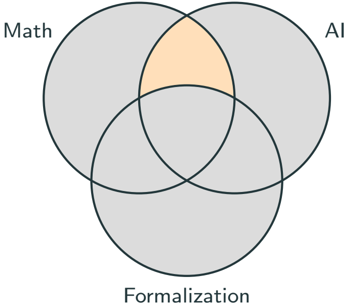
These are areas where bugs can be catastrophic (safety, security, cost).

Proof Assistants Landscape

- **Lean** — developed at Microsoft Research, now community-driven. Popular for mathematics (`mathlib`).
- **Coq / Rocq** — one of the oldest, used for CompCert, Four Color Theorem.
- **Isabelle/HOL** — used for seL4, Flyspeck project.
- **Agda** — dependently typed, popular in PL research.
- **HOL Light** — simple and trustworthy, used in Flyspeck.

Each has different strengths: automation level, library size, learning curve, community.

Mathematics $\cap$ AI



AI for Mathematics

Here we focus on **AI for Mathematics**, not the other way around.[1]

There are many ways to use AI in mathematics, e.g.,

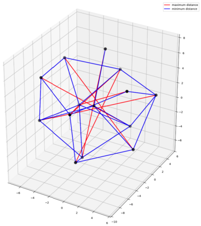
- Solving contest problems: IMO, Putnam, etc.
- Discovery: Find new conjectures, patterns, examples, counterexamples, etc.
- Proof: Assist in proving theorems. Generate ideas, prove lemmas, literature search, etc.

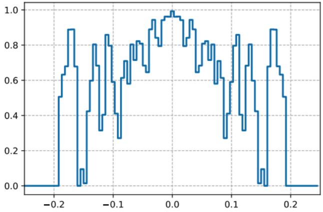
[1] Mathematics for AI includes deep learning theory, optimization, etc.

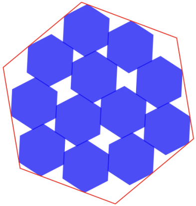
AI for Mathematics - Discovery

- Use AI to find rare examples.
  - AlphaTensor [13], FunSearch [27], AlphaEvolve [26], PatternBoost [7], FlowBoost [4], etc.
- Use AI to predict mathematical objects. Discover conjectures via interpreting the models.
  - Classify invariants in number theory [20]
  - New description of zeta map for (q, t)-Catalan numbers using Transformer model [21]
  - Use decision tree to study Galois groups [19, 23]

AlphaEvolve







DeepMind's `AlphaEvolve` [26] is an evolutionary coding agent powered by `Gemini` for general-purpose algorithm discovery and optimization.[2]

[2] Official blog post

Definition

A *Nikodym set* in $\mathbb{F}_q^d$ is a subset $N \subset \mathbb{F}_q^d$ with the property that for every $x \in \mathbb{F}_q^d$ there exists a line $\ell \ni x$ such that $\ell \backslash \{x\} \subset N$.

Bukh and Chao [6] proved the following lower bound for any Nikodym set N:

$$|N| \geq \frac{q^d}{2^{d-1}} + O(q^{d-1})$$

Conjecturally, such sets should have asymptotically full density:

$$|N| \geq q^d - o(q^d)$$

Tao considered the opposite problem: *constructing* Nikodym sets of size as small as possible.

When $d = 2$, Blokhuis et al. [5] constructed N with

$$|N| = q^2 - q^{3/2} + O(q \log q)$$

when q is a perfect square. For general d (and still q a perfect square), one has

$$|N| \leq q^d - \left\lfloor \frac{d}{2} \right\rfloor q^{d-1/2} + O(q^{d-1} \log q)$$

In [30], Tao experimented with the case of $d = 3$, where AlphaEvolve ended up with a construction by removing low-degree algebraic surfaces from $\mathbb{F}_q^3$. Motivated from the construction and conversation with Gemini Deep Think, he proved the following new upper bound:

Theorem (Gemini Deep Think, Tao [30])

For $d \geq 3$ and q an odd prime power, we have

$$|N| \leq q^d - \left(\frac{d-2}{\log 2} + 1 + o(1)\right) q^{d-1} \log q$$

as $q \to \infty$.

ML for Galois groups

Kyu-Hwan Lee and myself studied Galois groups of (Galois) number fields using Decision Tree [23].

For a number field $K/\mathbb{Q}$, let $a_n(K)$ be the number of ideals of $\mathcal{O}_K$ of index n (Dedekind zeta coefficients). Consider degree 9 Galois extensions $K/\mathbb{Q}$. By undergraduate algebra, we know that $\mathrm{Gal}(K/\mathbb{Q})$ is isomorphic to one of the following groups:

$$C_9 \text{ or } C_3 \times C_3.$$

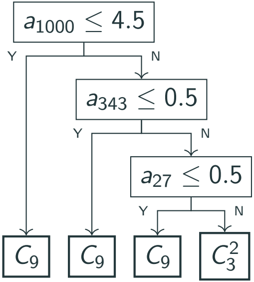
The task is to distinguish these two groups using only the data $\{a_n(K)\}_{n \leq N}$ for some N (we choose $N = 1000$).

ML for Galois groups

A decision tree is nothing but if-else statements, trained on a training dataset. We can achieve 100% accuracy on the test set using the following tree:

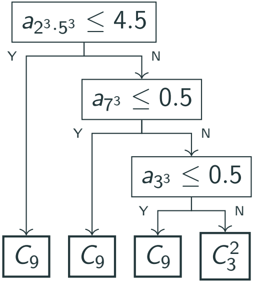
ML for Galois groups

A decision tree is nothing but if-else statements, trained on a training dataset. We can achieve 100% accuracy on the test set using the following tree:



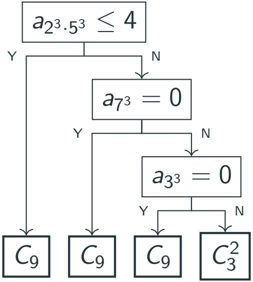
ML for Galois groups

A decision tree is nothing but if-else statements, trained on a training dataset. We can achieve 100% accuracy on the test set using the following tree:



ML for Galois groups

A decision tree is nothing but if-else statements, trained on a training dataset. We can achieve 100% accuracy on the test set using the following tree:



ML in Number Theory

By inspecting the tree carefully, we conjecture and actually prove

Proposition (Lee-L. [23])

Let K be a degree 9 Galois extension of $\mathbb{Q}$. Then $\mathrm{Gal}(K/\mathbb{Q})$ is cyclic if and only if there exists $m \geq 1$ such that $a_{m^2}(K) = 0$, where $a_n(K)$ is the number of ideals of $\mathcal{O}_K$ of index n.

which can be generalized to degree ℓ^2 Galois extensions for prime ℓ, and we also have more interesting results with degrees $6, 8, 10$. The important point here is that *proving* the conjecture is not hard (and done by humans), but *discovering* the conjecture is motivated by ML experiments.

Recently, there are several works where LLMs helped mathematicians to solve research-level problems in mathematics.

`Aletheia`[3] is a math research agent built upon `Gemini Deep Think`. It solved several Erdős problems [3, 15], but also other research problems in representation theory and number theory [16, 14], complexity theory [2], and combinatorics [22].

[3] GitHub

In [14], Feng used `Aletheia` to generalize *eigenweight* computations in their previous work [16] on Arithmetic Hirzebruch Proportionality, from Type A to other classical types.

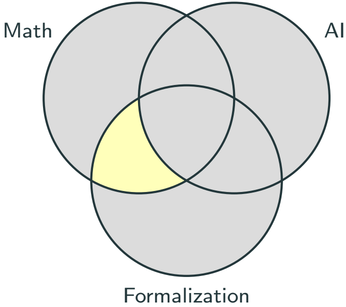
Theorem (`Aletheia`, Feng [14, Theorem 1.3.6])

Let $G = \mathrm{PSp}_{2n}$ *with* $n \geq 2$*, and* $\rho_n = (n, n-1, \ldots, 1)$ *be a partition. Let* μ *be the minuscule (spin) coweight of* G *and* $N = \binom{n+1}{2} + 1$ *be the arithmetic dimension of* G/P_μ*. For* $\Omega = \frac{1}{2}\sum_{i=1}^{n} x_i^2$*, the eigenweights are*

$$\epsilon_k(\Omega, \mu) = (-1)^{N-1} 2^{-N} \sum_{j=0}^{k-1} (-1)^j \chi^{2\pi_j(k)+\rho_n}(\nu_k) \quad \text{for } k = 1, 2, \ldots, n$$

where $\pi_j(k) = (k-j, 1^j)$ *and* $\nu_k = (2k-1, 1^N)$

Mathematics $\cap$ Formalization



Formalization of Mathematics

Recently, there is huge interest in formalizing mathematical proofs in Lean or other languages.

Formalization of Mathematics

Recently, there is huge interest in formalizing mathematical proofs in Lean or other languages.

But why?

Formalization of Mathematics

Recently, there is huge interest in formalizing mathematical proofs in Lean or other languages.

But why?

- Widely believed to be true $\neq$ we know how to prove
- As a digitized library

Check out Kevin Buzzard's ICM2022 talk.

Mathematics is “rigorous”

Sometimes, you can find these words in mathematical papers:

- “Private communication”
- “In preparation” (for 10 years)
- “Methods in [...] apply here.” (and sometimes don’t)
- “It is well known that” (but where can I find the statement?)

An annals paper

ANNALS OF MATHEMATICS

Princeton University & Institute for Advanced Study

About | Editorial Board | Submission Guidelines | Subscriptions | Contact

Quasi-projectivity of moduli spaces of polarized varieties

Pages 597-639 from Volume 159 (2004), Issue 2 *by Georg Schumacher, Hajime Tsuji*

Abstract

By means of analytic methods the quasi-projectivity of the moduli space of algebraically polarized varieties with a not necessarily reduced complex structure is proven including the case of nonuniruled polarized varieties.

Figure 1: Quasi-projectivity of moduli spaces of polarized varieties

Another annals paper

ANNALS OF MATHEMATICS

Princeton University & Institute for Advanced Study

About   Editorial Board   Submission Guidelines   Subscriptions   Contact

Non-quasi-projective moduli spaces

Pages 1077-1096 from Volume 164 (2006), Issue 3 *by János Kollár*

Abstract

We show that every smooth toric variety (and many other algebraic spaces as well) can be realized as a moduli space for smooth, projective, polarized varieties. Some of these are not quasi-projective. This contradicts a recent paper (Quasi-projectivity of moduli spaces of polarized varieties, *Ann. of Math.* **159** (2004) 597–639.).

Figure 2: Non-quasi-projective moduli spaces

Kepler's conjecture and Flyspeck project

Figure 3: Thomas Hales

Kepler's conjecture and Flyspeck project

Thomas Hales announced a proof of Kepler's conjecture (3-dimensional sphere packing problem) in 2003. However, the proof is heavily computer-assisted — including 23000 inequalities checked with computers — and some people were skeptical about the proof. It finally got accepted to Annals, "with 99% certain of the correctness of the proof" [18].

Hence, Hales decided to *formalize* the proof, which is called the **Flyspeck Project**. Using Isabelle + HOL Light, with 22 more people, he finally announced a completed formal proof in 2014 [17].

Lean is an interactive theorem prover developed by Leonardo de Moura.

Lean became popular in mathematics community, because

- Strong automation (tactics like `simp`, `ring`, `linarith`, `polyrith`, `grind`)
- Active community with mathematicians involved
- `mathlib`: comprehensive mathematics library

`mathlib`

`mathlib` is the mathematics library for Lean, community-driven and open source.

- Started in 2017, now $\sim$1.8 million lines of code
- Almost 600 contributors, including professional mathematicians
- Covers undergraduate to research-level mathematics

Examples of formalized mathematics in `mathlib`:

- Algebraic geometry: Schemes, sheaves, morphisms
- Number theory: Dirichlet L-functions, class field theory foundations
- Category theory: Fibered categories, limits/colimits, adjunctions
- Algebra: Group cohomology, representation theory, Lie algebras
- Analysis: Measure theory, Fourier analysis, complex analysis

(Also, there is a new `CSLib`![4])

[4] GitHub

Lean code example

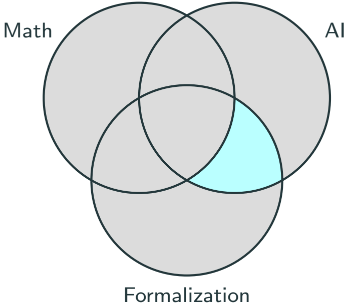
```
theorem center_eq_bot_of_odd_ne_one (hodd : Odd n) (hne1 : n ≠ 1) :
    Subgroup.center (DihedralGroup n) = ⊥ := by
  simp only [Subgroup.eq_bot_iff_forall, Subgroup.mem_center_iff]
  rintro (i | i) h
  · have heq := sr.inj (h (sr i))
    simp_all
  · have heq := sr.inj (h (r 1))
    have : Fact (1 < n) := ⟨by grind⟩
    simp [sub_eq_iff_eq_add, add_assoc,
     ZMod.add_self_eq_zero_iff_eq_zero hodd] at heq
```

Famous Lean formalization projects

- Liquid tensor experiment
- Sphere eversion
- Carleson project
- Equational theories project
- Fermat's last theorem
- ∞-cosmos project

Part of these projects are upstreamed / will be upstreamed to `mathlib`.

AI $\cap$ Formalization



Autoformalization

Formalization often requires a significant amount of work. Why not ask AI to do this?

Autoformalization

Formalization often requires a significant amount of work. Why not ask AI to do this?

But not just for mathematics?

Aristotle - Verina benchmark

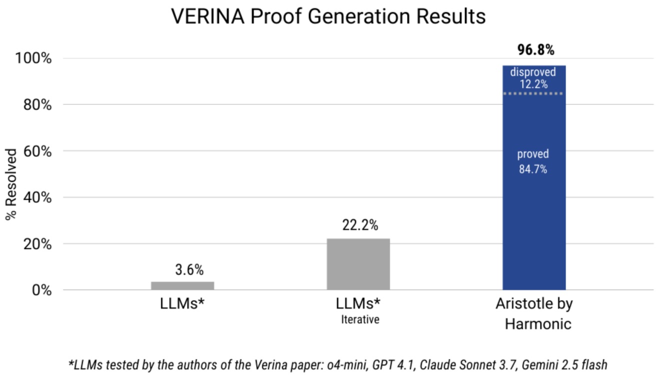


Figure 4: Aristotle on Verina benchmark

Verina is a benchmark for verifiable code generation with 189 Lean programming challenges, and Aristotle achieved 96.8% on it.

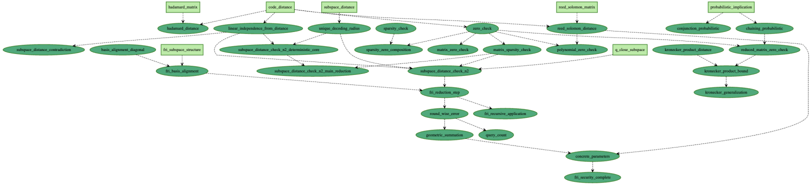
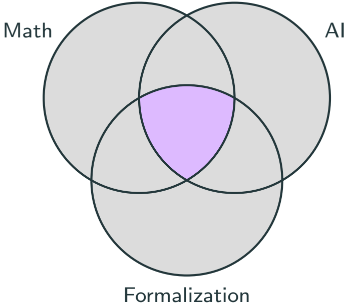


Figure 5: Gauss on certifying FRI protocol

Gauss gave a Lean formalization of the Fast Reed-Solomon Interactive Oracle Proof (FRI) protocol, a core component of modern transparent, STARK-style zero-knowledge proofs. The formalization was guided by a LaTeX blueprint with human scaffolding.

Mathematics ∩ AI ∩ Formalization



Mathematics ∩ AI ∩ Formalization

This can have several meanings, such as

1. Solve contest problems in Lean (e.g. IMO, Putnam, etc.)
2. Formalization of already existing mathematical proofs with help of AI (autoformalization)
3. AI solve an open problem, and the proof is (separately) formalized by humans/AI

IMO Grand Challenge

IMO Grand Challenge

The International Mathematical Olympiad (IMO) is perhaps the most celebrated mental competition in the world and as such is among the ultimate grand challenges for Artificial Intelligence (AI).

The challenge: build an AI that can win a gold medal in the competition.

ABSTRACTIONS BLOG

At the Math Olympiad, Computers Prepare to Go for the Gold

15

Computer scientists are trying to build an AI system that can win a gold medal at the world's premier math competition.

IMO Grand Challenge

- 2024: DeepMind's `AlphaProof` and `AlphaGeometry2` achieved Silver medal level performance (28/42, gold medal threshold was 29/42), with formal proof in Lean. [5]
- 2025: More AI models entered the game:
  - `Gemini Deep Think` [6] and OpenAI's internal version of `ChatGPT` achieved Gold medal level performance (35/42).
  - `Aristotle` and `SeedProver` achieved the same performance in Lean [1, 12].
  - None of the models solved Problem 6 (the hardest problem) correctly. [7]

[5] Official blog post by DeepMind
[6] Official blog post by DeepMind
[7] `AlphaEvolve` found *answer* later (without proof).

Putnam 2025

PUTNAM2025	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	B6
ARISTOTLE	**30**	60	**30**	180	–	**60**	150	**25**	40	–	420	**180**
SEED-PROVER 1.5	60	**30**	120	240	–	240	540	360	**30**	120	240	**180**
AXIOM	110	180	165	**107**	**518**	259	270	65	43	**112**	254	494
NUMINA-LEAN-AGENT	97	**30**	44	169	2040	89	**55**	142	**30**	308	**88**	797

Figure 6: Time spent comparison (Unit: minutes) [24]

`Aristotle`, `Seed-Prover 1.5`, `AxiomProver`, and `Numina-Lean-Agent` participated in Putnam 2025, and solved the problems in Lean. The above table shows the number of problems solved by each model [24].

`AxiomProver` solved 8/12 problems in Putnam 2025 during the contest time, and solve the rest afterward.[8] Some of the solutions are different from human solutions.

Putnam 2025 A4

Find minimal value of k such that there exists k-by-k real matrices $A_1, \ldots, A_{2025}$ with the property that $A_i A_j = A_j A_i$ if and only if $|i - j| \in \{0, 1, 2024\}$

[8] Official blog post

`AxiomProver` solved 8/12 problems in Putnam 2025 during the contest time, and solve the rest afterward.[8] Some of the solutions are different from human solutions.

Putnam 2025 A4

Find minimal value of k such that there exists k-by-k real matrices $A_1, \ldots, A_{2025}$ with the property that $A_i A_j = A_j A_i$ if and only if $|i - j| \in \{0, 1, 2024\}$

The answer is $k = 3$. There was a debate among human mathematicians, and `AxiomProver` provided a solution by setting A_i as rank-1 projection matrices onto certain vectors in $\mathbb{R}^3$.

[8] Official blog post

Putnam 2025

Putnam 2025 A5

Let n be an integer with $n \geq 2$. For a sequence $s = (s_1, \ldots, s_{n-1})$ where each $s_i = \pm 1$, let $f(s)$ be the number of permutations $(a_1, \ldots, a_n)$ of $\{1, 2, \ldots, n\}$ such that $s_i(a_{i+1} - a_i) > 0$ for all i. For each n, determine the sequences s which $f(s)$ is maximal.

`Numina-Lean-Agent` adopts a novel subagent mechanism that decomposes the proof into several subgoals and solves them independently, effectively mitigating the issue of excessively long contexts.

Sphere packing project

Goal

Formalize Viazovska's proof of optimality of E_8 sphere packing in dimension 8.

The project was kicked-off by Sidharth Hariharan and Maryna Viazovska in 2024 Spring, and currently maintained by Sidharth Hariharan, Chris Birkbeck, Bhavik Mehta, and myself.

It became public in Big Proof conference in 2025, and people started to contribute.

The goal of the project is not just a `sorry`-free Lean proof, but also a maintainable codebase where one can upstream part of them to `mathlib` for reusability. And we already did some of them - e.g. the weight 2 Eisenstein series E_2 is in `mathlib` now.

Sphere packing project

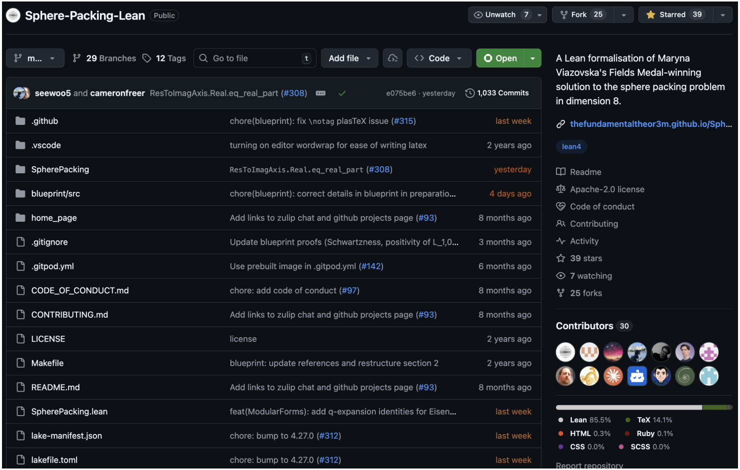


Figure 7: github.com/thefundamentaltheor3m/Sphere-Packing-Lean

Human and non-human contributors

There are 22 human contributors for the project:

Human and non-human contributors

There are 22 human contributors for the project:

But, there are also non-human contributors:

- Harmonic, `Aristotle`
- Anthropic, `Claude Opus 4.5` (with )
- Math Inc., `Gauss`
- `GitHub Copilot`
- and some other bots

These AI models certainly accelerated the formalization process. But they often write “messy” code (which never meets the high standard of `mathlib`), and further human refinement is necessary.

OPEN

Let $f \in \mathbb{Z}[x]$ be an irreducible non-constant polynomial such that $f(n) \geq 1$ for all large $n \in \mathbb{N}$. Does there exist a constant $c = c(f) > 0$ such that

$$\sum_{n \leq X} \tau(f(n)) \sim cX \log X,$$

where τ is the divisor function?

#975: [Er65b]

number theory | divisors | polynomials

Disclaimer: The open status of this problem reflects the current belief of the owner of this website. There may be literature on this problem that I am unaware of, which may partially or completely solve the stated problem. Please do your own literature search before expending significant effort on solving this problem. If you find any relevant literature not mentioned here, please add this in a comment.

Van der Corput [Va39] proved that

$$\sum_{n \leq X} \tau(f(n)) \gg_f X \log X.$$

Erdős [Er52b] proved using elementary methods that

$$\sum_{n \leq X} \tau(f(n)) \ll_f X \log X.$$

Such an asymptotic formula is known whenever f is an irreducible quadratic, as proved by Hooley [Ho63]. The form of c depends on f in a complicated fashion (see the work of McKee [Mc95], [Mc97], and [Mc99] for expressions for various types of quadratic f). For example,

$$\sum_{n \leq x} \tau(n^2 + 1) = \frac{3}{\pi} x \log x + O(x).$$

Tao has a blog post on this topic.

Figure 8: Erdős problem #975

Erdős problems

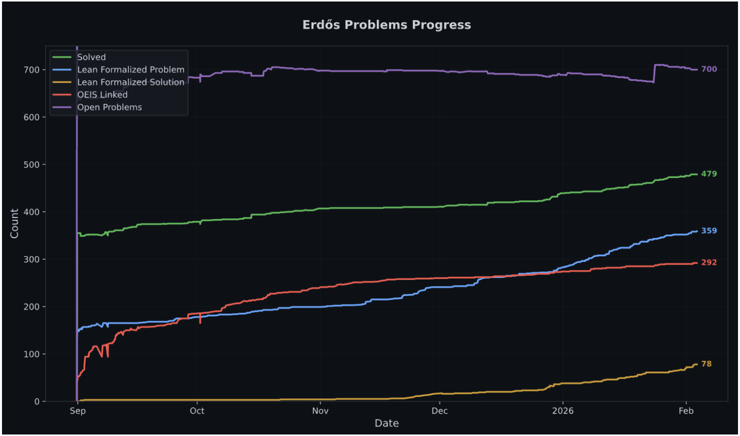


Figure 9: github.com/teorth/erdosproblems

Erdős problem 728

Theorem ✓ (ChatGPT5.2-Pro, **Barreto–Sothanaphan [29]**)

Let $C > 0$ and $\epsilon > 0$ be sufficiently small. Then there are infinitely many integers a, b, n with $a \geq \epsilon n$ and $b \geq \epsilon n$ such that

$$a!b! \mid n!(a+b-n)! \quad \text{and} \quad a+b > n + C\log n$$

Kevin Barreto tested ChatGPT5.2-Pro on several analytic number theory problems, and in particular, he found that the model gave a reasonable proof for the above problem. Then he fed the proof into Aristotle, which returned a Lean proof of it.

Check out a writeup [29] and a blog post for details.

Extremal descendant integrals on moduli spaces of curves

In [28], Johannes Schmitt studied optimization problem on the *descendant integrals* (or intersection numbers) on the moduli spaces of curves:

$$\langle \tau_{e_1} \cdots \tau_{e_n} \rangle_g := \int_{\overline{\mathcal{M}}_{g,n}} \psi_1^{e_1} \cdots \psi_n^{e_n}$$

where $\psi_i = c_1(\mathbb{L}_i) \in \mathrm{H}^2(\overline{\mathcal{M}}_{g,n}, \mathbb{Q})$ is the ψ-class for the marked point i. It is a rational number that vanishes unless $\sum_{i=1}^n e_i = 3g - 3 + n$. We write the integral as $D(\mathbf{e})$ for

$$\mathbf{e} \in E(g,n) := \left\{ \mathbf{e} = (e_1, \ldots, e_n) \in \mathbb{Z}_{\geq 0}^n : \sum_{i=1}^n e_i = 3g - 3 + n \right\}$$

We call a vector $\mathbf{e}$ is *balanced* if $|e_i - e_j| \leq 1$ for all i, j.

Extremal descendant integrals on moduli spaces of curves

Theorem (Schmitt [28, Theorem 1.2])

Let $g \in \mathbb{Z}_{\geq 0}$ and $n \in \mathbb{Z}_{\geq 0}$ with $2g - 2 + n > 0$.

(a) *D achieves its minimum at the concentrated vector $(3g - 3 + n, 0, \ldots, 0)$ (or any of its permutations), with value*

$$\langle \tau_{3g-3+n} \cdot \tau_0^{n-1} \rangle_g = \frac{1}{24^g \cdot g!}.$$

(b) *D achieves its maximum on a balanced vector.*

Extremal descendant integrals on moduli spaces of curves

The question itself occurred when Schmitt was trying to find a toy problem for `OpenEvolve`, an independent open-source version of `AlphaEvolve`. `OpenEvolve` found that maximums are often obtained by the balanced vectors.

Then he formulated it as a conjecture and submit to his own IMProofBench project[9] (research-level benchmark problems for AI), and several versions of `GPT-5` give similar proofs.

After that, part of the argument (the following theorem) is formalized in Lean by `Claude Opus` and `ChatGPT-5.2`.

[9] improofbench.math.ethz.ch

For integers $n \geq 1$ and $d \geq 0$, let

$$E(n,d) := \left\{ \mathbf{e} = (e_1, \ldots, e_n) \in \mathbb{Z}^n_{\geq 0} : \sum_{j=1}^{n} e_j = d \right\}.$$

For $\mathbf{e} \in E(n,d)$, we write $\mathbf{e} - \delta_i + \delta_j$ for the vector obtained by subtracting 1 from the i-th coordinate and adding 1 to the j-th coordinate (when $e_i \geq 1$).

Then the above theorem reduces to the following theorem on optimization problem.

Theorem ✓ (`GPT-5`, `Claude Opus`, Schmitt [28, Theorem 3.1])

Let $D : E(n,d) \to \mathbb{Q}$ be a function satisfying

- *(Symmetry)* $D(\mathbf{e} \circ \sigma) = D(\mathbf{e})$ *for all* $\sigma \in S_n$,
- *(Log-Concavity) For all* $\mathbf{e} \in E(n,d)$ *and distinct* i, j *with* $e_i, e_j \geq 1$,

$$D(\mathbf{e})^2 \geq D(\mathbf{e} - \delta_i + \delta_j) \cdot D(\mathbf{e} + \delta_i - \delta_j)$$

- *(Strict Positivity)* $D(\mathbf{e}) > 0$ *for all* $\mathbf{e} \in E(n,d)$.

Then

1. *D achieves its maximum on a balanced vector (where* $|e_i - e_j| \leq 1$ *for all* i, j*).*
2. *D achieves its minimum on a concentrated vector (where* $\mathbf{e} = d \cdot \delta_k$ *for some* k*).*

A positive integer n is called *dead end* in base b if (1) n is square-free and (2) all the numbers obtained by concatenating a digit to the right of n are not square-free.

In [25], Miller et al. conjectured that the asymptotic density of dead ends for base $b = 10$ is about

$$\frac{6}{\pi^2} \cdot \left(1 - \frac{6}{\pi^2}\right)^{10} \approx 5.21818 \times 10^{-5},$$

based on the fact that the density of square-free numbers is $6/\pi^2$.

AxiomProver *found and proved* the explicit formula of the asymptotic density of dead ends in a general base b:

Theorem ✓ (AxiomProver, Chen–Lau–L.–Ono–Zhang [10])

Let $D_b(X)$ be the number of dead ends in base b up to X. Then

$$D_b(X) = \sum_{S \subseteq \mathcal{D}_b} (-1)^{|S|} \prod_p \left(1 - \frac{\nu_{p,b}(S)}{p^2}\right) \cdot X + O_b\left(\frac{X}{\sqrt{\log X}}\right)$$

where $\mathcal{D}_b = \{0, 1, \ldots, b-1\}$ and

$$\nu_{p,b}(S) := \#\left\{ n \bmod p^2 : \ p^2 \mid n \text{ or } p^2 \mid (bn + d) \text{ for some } d \in S \right\}.$$

In particular, the correct asymptotic density of dead ends in base 10 is about 1.3170×10^{-9}, disproving the previous conjecture.

An odd prime p is *regular* if $p \nmid \mathrm{num}(B_{2k})$ for all $2 \leq 2k \leq p-3$, where B_{2k} is the Bernoulli number. Kummer proved that FLT holds for regular exponent p. It is conjectured that there are infinitely many regular primes, but it is still open.[10]

A weaker notion: p is *m-regular* if $p \nmid \mathrm{num}(B_{2k})$ for $2 \leq 2k \leq m$ for some $m = m(p) < p-3$.

[10] It is known that there are infinitely many *irregular* primes.

Theorem ✓ (AxiomProver, Chen–Lau–L.–Ono–Zhang [9])

Fix $\alpha > 1/2$ and let $M_\alpha(p) = \left\lfloor \frac{\sqrt{p}}{(\log p)^\alpha} \right\rfloor$. Then there exists a constant $C_\alpha > 0$ such that

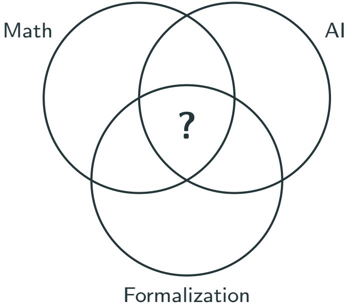
$$\#\{p \le X \text{ prime} : p \text{ is not } M_\alpha(p)\text{-regular}\} \le C_\alpha \frac{X}{(\log X)^{2\alpha}}$$

In particular, almost all primes are $M_\alpha(p)$-regular.

One of the formalized proofs used von Staudt–Clausen theorem, which isn't in `mathlib` and also formalized during the process (which will be upstreamed). Another run only proved a consequence of vSC that it needs, and showed that one can take $C_\alpha = 10$ for all $\alpha > 1/2$.

See also [8] on spin parity of differentials and [11] on syzygies.

What's next?



Present and Future

So far, we have seen many examples where

- AI helps mathematicians to discover new mathematical objects and conjectures,
- AI proves modest open problems in mathematics,
- AI helps formalization of existing mathematical theorems.

But we don't have an example where AI creates **genuinely new ideas**[11] **to solve important open problems that many people care about**.

[11] This is a very subjective term, but most will agree on this point.

FrontierMath Open Problems

Epoch AI benchmarked *FrontierMath* which are high school olympiad to research level problems (Tier 1 - 4). Recently, they announced *FrontierMath Open Problems*[12], where AI models are challenged to solve open problems in mathematics.

Inverse Galois Problem

Find a degree 23 polynomial in $\mathbb{Z}[x]$ whose splitting field over $\mathbb{Q}$ has Galois group M_{23}.

It genuinely requires new mathematical ideas to solve this problem.

[12] https://epoch.ai/frontiermath/open-problems

As an early career mathematician

When I attended FrontierMath Symposium (for Tier 4 dataset), I was able to use the paid version of ChatGPT for the first time (`o3` and `o4-mini`), without paying (thanks to OpenAI). And I was quite shocked by the fact that many of the proposals submitted by the participants were solved by the models in a few minutes.

As an early career mathematician

When I attended FrontierMath Symposium (for Tier 4 dataset), I was able to use the paid version of ChatGPT for the first time (`o3` and `o4-mini`), without paying (thanks to OpenAI). And I was quite shocked by the fact that many of the proposals submitted by the participants were solved by the models in a few minutes.

Q. Will I be replaced by AI mathematicians?

As an early career mathematician

When I attended FrontierMath Symposium (for Tier 4 dataset), I was able to use the paid version of ChatGPT for the first time (`o3` and `o4-mini`), without paying (thanks to OpenAI). And I was quite shocked by the fact that many of the proposals submitted by the participants were solved by the models in a few minutes.

Q. Will I be replaced by AI mathematicians?

A. No! (hopefully) We need to *collaborate* with AI, not compete against AI.

As an early career mathematician

Paata Ivanisvili, Professor at UC Irvine[13]

"I also notice that PhD students who actively use AI tend to move noticeably faster than those who are pessimistic or dismissive of the technology. This is not an advertisement for paying hundreds of dollars per month for frontier models–but it is a reminder to stay open-minded, curious, and willing to try new tools rather than reject them a priori."

[13] https://x.com/PI010101/status/2016632840780140675?s=20

If you want to know more about the area

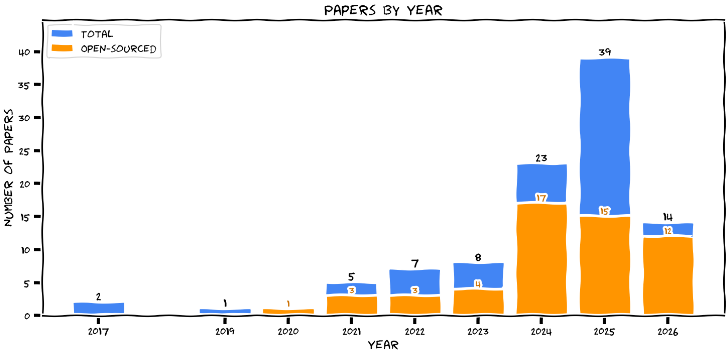


Figure 10: seewoo5.github.io/awesome-ai-for-math

References i

[1] T. Achim, A. Best, A. Bietti, K. Der, M. Fédérico, S. Gukov, D. Halpern-Leistner, K. Henningsgard, Y. Kudryashov, A. Meiburg, et al.
Aristotle: Imo-level automated theorem proving.
arXiv preprint arXiv:2510.01346, 2025.

[2] A. Asadi, K. Chatterjee, E. Goharshady, M. Karrabi, A. Montaseri, and C. Pagano.
Strongly Polynomial Time Complexity of Policy Iteration for L_∞ Robust MDPs.
arXiv preprint arXiv:2601.23229, 2026.

[3] K. Barreto, J. Kang, S.-h. Kim, V. Kovač, and S. Zhang.
Irrationality of rapidly converging series: a problem of Erdős and Graham.
arXiv preprint arXiv:2601.21442, 2026.

[4] G. Bérczi, B. Hashemi, and J. Klüver.
Flow-based Extremal Mathematical Structure Discovery.
arXiv preprint arXiv:2601.18005, 2026.

[5] A. Blokhuis, A. E. Brouwer, D. Jungnickel, V. Krčadinac, S. Rottey, L. Storme, T. Szőnyi, and P. Vandendriessche.
Blocking sets of the classical unital.
Finite fields and their applications, 35:1–15, 2015.

References ii

[6] B. Bukh and T.-W. Chao.
Sharp density bounds on the finite field Kakeya problem.
arXiv preprint arXiv:2108.00074, 2021.

[7] F. Charton, J. S. Ellenberg, A. Z. Wagner, and G. Williamson.
Patternboost: Constructions in mathematics with a little help from ai.
arXiv preprint arXiv:2411.00566, 2024.

[8] D. Chen, E. Chen, K. Lau, K. Ono, and J. Zhang.
Parity of k-differentials in genus zero and one.
arXiv preprint arXiv:2602.03722, 2026.

[9] E. Chen, C. Cummins, B. Eltschig, D. Grubisic, L. Haller, L. Hong, A. Kurghinyan, K. Lau, H. Leather, S. Lee, A. Markosyan, K. Ono, M. Patel, G. Pendharkar, V. Rathi, A. Schneidman, V. Seeker, S. Sengupta, I. Sinha, J. Xin, and J. Zhang.
Almost all primes are partially regular.
arXiv preprint arXiv:2602.05090, 2026.

[10] E. Chen, C. Cummins, B. Eltschig, D. Grubisic, L. Haller, L. Hong, A. Kurghinyan, K. Lau, H. Leather, S. Lee, A. Markosyan, K. Ono, M. Patel, G. Pendharkar, V. Rathi, A. Schneidman, V. Seeker, S. Sengupta, I. Sinha, J. Xin, and J. Zhang.
Dead ends in square-free digit walks.
arXiv preprint arXiv:2602.05095, 2026.

References iii

[11] E. Chen, C. Cummins, D. Grubisic, L. Haller, L. Hong, A. Kurghinyan, K. Lau, H. Leather, S. Lee, A. Markosyan, K. Ono, M. Patel, G. Pendharkar, V. Rathi, A. Schneidman, V. Seeker, S. Sengupta, I. Sinha, J. Xin, and J. Zhang.
Fel's Conjecture on Syzygies of Numerical Semigroups.
arXiv preprint arXiv:2602.03716, 2026.

[12] L. Chen, J. Gu, L. Huang, W. Huang, Z. Jiang, A. Jie, X. Jin, X. Jin, C. Li, K. Ma, et al.
Seed-prover: Deep and broad reasoning for automated theorem proving.
arXiv preprint arXiv:2507.23726, 2025.

[13] A. Fawzi, M. Balog, A. Huang, T. Hubert, B. Romera-Paredes, M. Barekatain, A. Novikov, F. J. R. Ruiz, J. Schrittwieser, G. Swirszcz, et al.
Discovering faster matrix multiplication algorithms with reinforcement learning.
Nature, 610(7930):47–53, 2022.

[14] T. Feng.
Eigenweights for arithmetic Hirzebruch Proportionality.
arXiv preprint arXiv:2601.23245, 2026.

References iv

[15] T. Feng, T. Trinh, G. Bingham, J. Kang, S. Zhang, S. hyun Kim, K. Barreto, C. Schildkraut, J. Jung, J. Seo, C. Pagano, Y. Chervonyi, D. Hwang, K. Hou, S. Gukov, C.-C. Tsai, H. Choi, Y. Jin, W.-Y. Li, H.-A. Wu, R.-A. Shiu, Y.-S. Shih, Q. V. Le, and T. Luong.
Semi-Autonomous Mathematics Discovery with Gemini: A Case Study on the Erdös Problems.
arXiv preprint arXiv:2601.22401, 2026.

[16] T. Feng, Z. Yun, and W. Zhang.
Arithmetic volumes of moduli stacks of shtukas.
arXiv preprint arXiv:2601.18557, 2026.

[17] T. Hales, M. Adams, G. Bauer, T. D. Dang, J. Harrison, L. T. Hoang, C. Kaliszyk, V. Magron, S. McLaughlin, T. T. Nguyen, et al.
A formal proof of the Kepler conjecture.
In *Forum of mathematics, Pi*, volume 5, page e2. Cambridge University Press, 2017.

[18] T. C. Hales.
A proof of the Kepler conjecture.
Annals of mathematics, pages 1065–1185, 2005.

[19] Y.-H. He, K.-H. Lee, and T. Oliver.
Machine-Learning Number Fields.
Mathematics, Computation and Geometry of Data, 2(1):49–66, 2022.

References v

[20] Y.-H. He, K.-H. Lee, T. Oliver, and A. Pozdnyakov.
Murmurations of elliptic curves.
Experimental Mathematics, 34(3):528–540, 2025.

[21] X. Huang, B. Jackson, and K.-H. Lee.
From Black Box to Bijection: Interpreting Machine Learning to Build a Zeta Map Algorithm.
arXiv preprint arXiv:2511.12421, 2025.

[22] J. Lee and J. Seo.
Lower bounds for multivariate independence polynomials and their generalisations.
arXiv preprint arXiv:2602.02450, 2026.

[23] K.-H. Lee and S. Lee.
Machines Learn Number Fields, But How? The Case of Galois Groups.
arXiv preprint arXiv:2508.06670, 2025.

[24] J. Liu, Z. Zhou, Z. Zhu, M. D. Santos, W. He, J. Liu, R. Wang, Y. Xie, J. Zhao, Q. Wang, et al.
Numina-Lean-Agent: An Open and General Agentic Reasoning System for Formal Mathematics.
arXiv preprint arXiv:2601.14027, 2026.

[25] S. J. Miller, F. Peng, T. Popescu, J. M. Siktar, N. Wattanawanichkul, and T. P. R. Program.
Walking to Infinity Along Some Number Theory sequences.
arXiv preprint arXiv:2010.14932, 2020.

[26] A. Novikov, N. Vũ, M. Eisenberger, E. Dupont, P.-S. Huang, A. Z. Wagner, S. Shirobokov, B. Kozlovskii, F. J. Ruiz, A. Mehrabian, et al.
AlphaEvolve: A coding agent for scientific and algorithmic discovery.
arXiv preprint arXiv:2506.13131, 2025.

[27] B. Romera-Paredes, M. Barekatain, A. Novikov, M. Balog, M. P. Kumar, E. Dupont, F. J. Ruiz, J. S. Ellenberg, P. Wang, O. Fawzi, et al.
Mathematical discoveries from program search with large language models.
Nature, 625(7995):468–475, 2024.

[28] J. Schmitt.
Extremal descendant integrals on moduli spaces of curves: An inequality discovered and proved in collaboration with AI.
arXiv preprint arXiv:2512.14575, 2025.

[29] N. Sothanaphan.
Resolution of Erdős Problem #728: a writeup of Aristotle's Lean proof.
arXiv preprint arXiv:2601.07421, 2026.

[30] T. Tao.
New Nikodym set constructions over finite fields.
arXiv preprint arXiv:2511.07721, 2025.