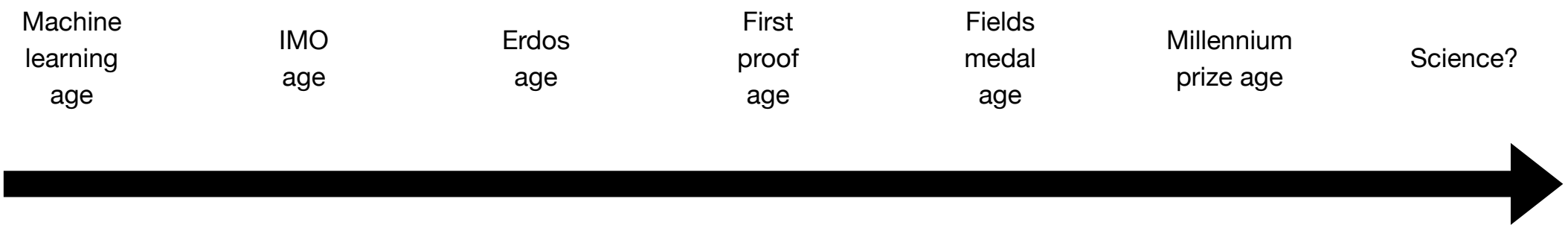
Autoformalization and the future of math and science

How far will we go? How fast?

Patrick Shafto
Professor / Program Manager
Rutgers / DARPA

Ages of AI math



Ages of AI math

Machine learning age

IMO age

Erdos age

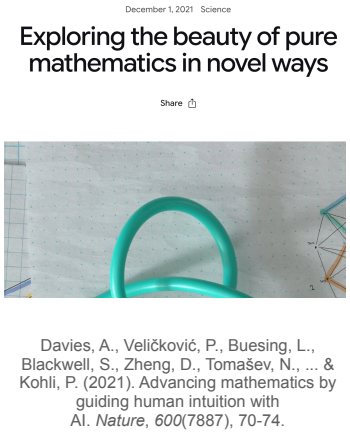
First proof age

Fields medal age

Millennium prize age

Science?

Dec 2021



December 1, 2021 Science

Exploring the beauty of pure mathematics in novel ways

Share

Davies, A., Veličković, P., Buesing, L., Blackwell, S., Zheng, D., Tomašev, N., ... & Kohli, P. (2021). Advancing mathematics by guiding human intuition with AI. *Nature*, *600*(7887), 70-74.

Ages of AI math

Machine learning age

Dec 2021

IMO age

Erdos age

First proof age

Fields medal age

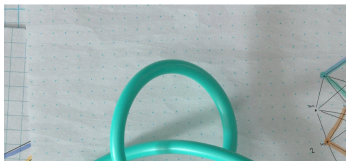
Millennium prize age

Science?

December 1, 2021 Science

Exploring the beauty of pure mathematics in novel ways

Share



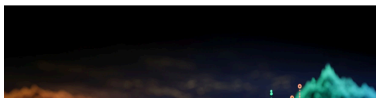
Davies, A., Veličković, P., Buesing, L., Blackwell, S., Zheng, D., Tomašev, N., ... & Kohli, P. (2021). Advancing mathematics by guiding human intuition with AI. *Nature*, *600*(7887), 70-74.

May 14, 2025 Science

AlphaEvolve: A Gemini-powered coding agent for designing advanced algorithms

AlphaEvolve team

Share



Romera-Paredes, B., Barekatain, M., Novikov, A., Balog, M., Kumar, M. P., Dupont, E., ... & Fawzi, A. (2024). Mathematical discoveries from program search with large language models. *Nature*, *625*(7995), 468-475.

Francois Charton

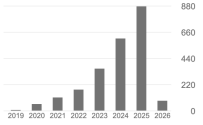
Member of technical staff, Axiom Math
Verified email at axiommath.ai
Artificial Intelligence

FOLLOW

TITLE	CITED BY	YEAR
Extrapolating jet radiation with autoregressive transformers A Butter, F Charton, J Mariño Villadamigo, A Ore, T Plehn, J Spinner SciPost Physics 20 (1), 004	5	2026
Pangenome-Informed Language Models for Synthetic Genome Sequence Generation P Huang, F Charton, JNM Schmelzle, SS Darnell, P Prins, E Garrison, ... bioRxiv, 2024.09. 18.612131	3	2025
Transformers know more than they can tell--Learning the Collatz sequence F Charton, A Narayanan arXiv preprint arXiv:2511.10811		2025
TAPAS: Datasets for Learning the Learning with Errors Problem E Saxena, A Alfarano, F Charton, E Wenger, K Lauter arXiv preprint arXiv:2510.08797		2025
Transforming calabi-yau constructions: Generating new calabi-yau manifolds with transformers JHT Yip, C Arnal, F Charton, G Shiu arXiv preprint arXiv:2507.03732	3	2025

Cited by

	All	Since 2021
Citations	2295	2225
h-index	20	20
i10-index	27	27



Public access VIEW ALL

0 articles 9 articles

not available available

Based on funding mandates

Ages of AI math

Machine learning age

IMO age

Erdos age

First proof age

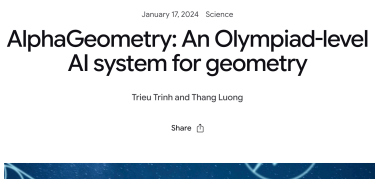
Fields medal age

Millennium prize age

Science?

Dec 2021

Jan 2024



Trinh, T. H., Wu, Y., Le, Q. V., He, H., & Luong, T. (2024). Solving olympiad geometry without human demonstrations. *Nature*, *625*(7995), 476-482.

July 21, 2025 Research

Advanced version of Gemini with Deep Think officially achieves gold-medal standard at the International Mathematical Olympiad

Thang Luong and Edward Lockhart

Jul 28, 2025 6:42 PM Eastern Daylight Time

Harmonic Announces IMO Gold Medal-Level Performance & Launch of First Mathematical Superintelligence (MSI) AI App

NEWS | 04 December 2025

DeepSeek's self-correcting AI model aces tough maths proofs

The mathematical reasoning model performed as well as humans at prestigious international mathematics competitions.

Ages of AI math

Machine learning age

IMO age

Jan 2024

Erdos age

Dec 2025

First proof age

Fields medal age

Millennium prize age

Science?

Dec 2021

1. AI-generated solutions, partial solutions, or negative results for previously open problems

Problem	AI tools used	Date	Outcome
[36]	AlphaEvolve	3 Nov, 2025	🟡 Slight improvement to past construction
[52]	AlphaEvolve	3 Nov, 2025	🔴 Did not match past constructions
[64]	AlphaEvolve	3 Nov, 2025	No counterexample found
[67]	AlphaEvolve	3 Nov, 2025	🔴 Did not match past constructions
[106]	AlphaEvolve	3 Nov, 2025	Matched past construction
[124]	Aristotle	29 Nov, 2025	🟡 Partial result
[391]	AlphaEvolve	3 Nov, 2025	🔴 Did not match past constructions
[493]	AlphaEvolve	3 Nov, 2025	No counterexample found
[507]	AlphaEvolve	3 Nov, 2025	🟡 Surpassed some past constructions
[728]	Aristotle, ChatGPT 5.2 Pro	6 Jan, 2026	🟢 Full solution (Lean), using arguments similar to 🟡 Pomerance (2015)
[1097]	AlphaEvolve	3 Nov, 2025	🟡 Slight improvement to past construction

Logical Intelligence ✓
@logic_int

🚀 Aleph prover just went BEAST MODE
4 math problems unsolved for 20+ years. Formal proofs in Lean 4. Less than 48 hours. Under $5k total.

✅ Binomial tail bounds conjecture (Telgarsky, 2009)
✅ Quantum gate lattice approximation (Greene & Damelin, 2015)*
✅ Erdős 124
✅ Erdős 481
✅ #1 on PutnamBench leaderboard

The era of AI mathematics is here.

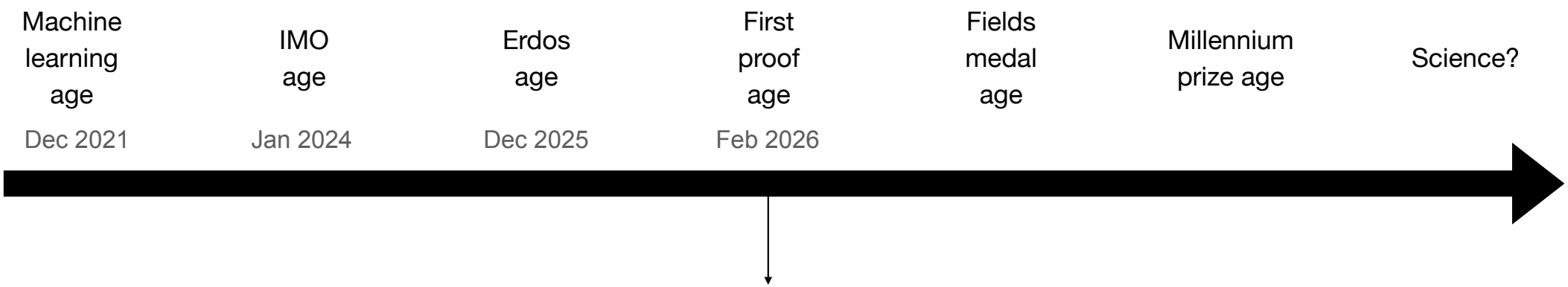
Special thanks to @BorisHanin and @ylecun for helping bring this to life 🙏
And massive kudos to the @LeanFRO team — none of this is possible without the incredible foundation you've built.

Aleph will be soon available to the public, stay tuned!

*conditional on results from Sardari (2015), formalization pending

3:12 PM · Dec 2, 2025 · **115.6K** Views

Ages of AI math



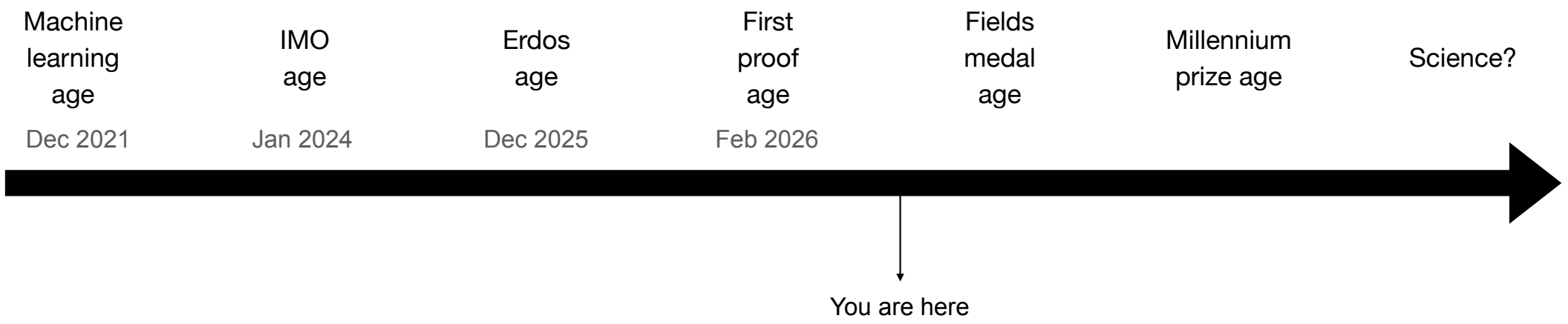
[Submitted on 5 Feb 2026]

First Proof

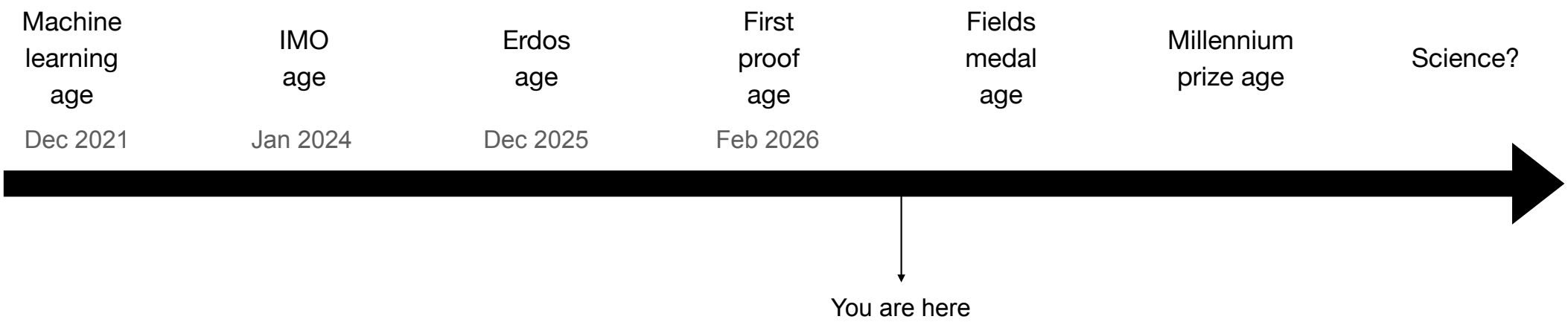
Mohammed Abouzaid, Andrew J. Blumberg, Martin Hairer, Joe Kileel, Tamara G. Kolda, Paul D. Nelson, Daniel Spielman, Nikhil Srivastava, Rachel Ward, Shmuel Weinberger, Lauren Williams

To assess the ability of current AI systems to correctly answer research-level mathematics questions, we share a set of ten math questions which have arisen naturally in the research process of the authors. The questions had not been shared publicly until now; the answers are known to the authors of the questions but will remain encrypted for a short time.

Ages of AI math



Ages of AI math



Largely “hammer and chisel” efforts. Humans are a backstop/bottleneck.

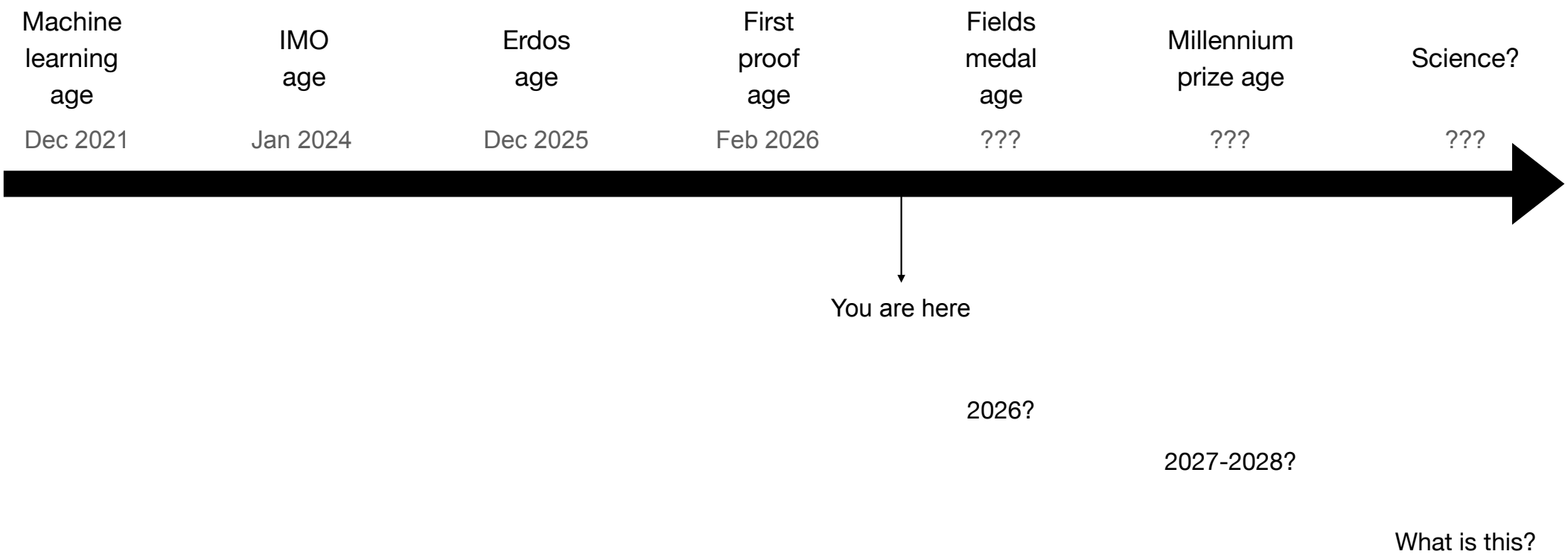
Grothendieck describes two styles in mathematics. If you think of a theorem to be proved as a nut to be opened, so as to reach "the nourishing flesh protected by the shell", then the *hammer and chisel* principle is: "put the cutting edge of the chisel against the shell and strike hard. If needed, begin again at many different points until the shell cracks—and you are satisfied". He says:

> I can illustrate the second approach with the same image of a nut to be opened. The first analogy that came to my mind is of immersing the nut in some softening liquid, and why not simply water? From time to time you rub so the liquid penetrates better, and otherwise you let time pass. The shell becomes more flexible through weeks and months—when the time is ripe, hand pressure is enough, the shell opens like a perfectly ripened avocado!
>
> A different image came to me a few weeks ago. The unknown thing to be known appeared to me as some stretch of earth or hard marl, resisting penetration... the sea advances insensibly in silence, nothing seems to happen, nothing moves, the water is so far off you hardly hear it... yet it finally surrounds the resistant substance. [Grothendieck 1985–1987, pp. 552-3][1]

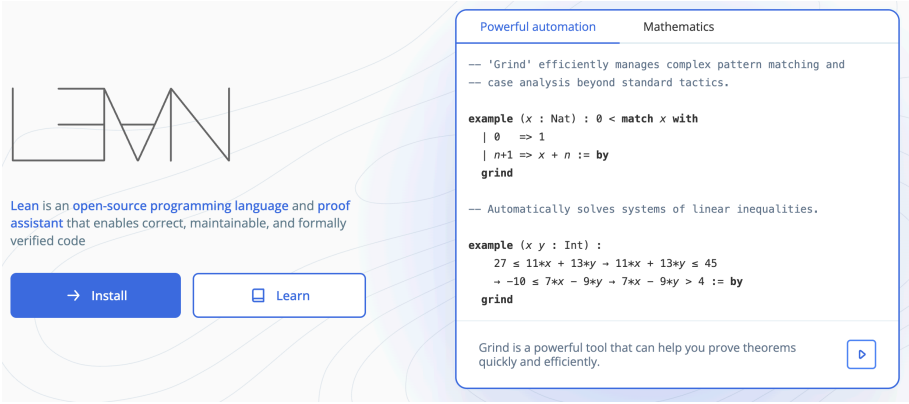
In this "rising sea" the theorem is "submerged and dissolved by some more or less vast theory, going well beyond the results originally to be established" [Grothendieck 1985–1987, p. 555].[2] Grothendieck calls this his approach and Bourbaki's. Here as so often he sees math research, exposition, and teaching as all the same. He

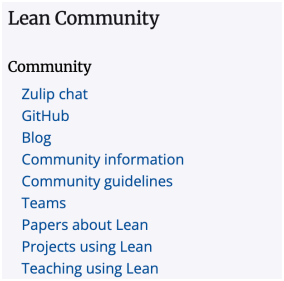
McLarty, C., Gray, J., & Parshall, K. (2007). The rising sea: Grothendieck on simplicity and generality. *Episodes in the history of modern algebra (1800–1950)*, *32*, 301-325.

Ages of AI math



Formal verification languages offer a potential solution…





A mathlib overview

The goal of this web page is to give a rough list of topics currently covered in mathlib, and provide pointers for exploration. This is not meant to be an exhaustive list, and could be outdated (see the full index for exhaustive and up to date information).

Here topics are listed in the greatest generality we currently have in mathlib, hence some things may be difficult to recognize. We also have a page dedicated to undergraduate mathematics which may be easier to read, as well as a page listing undergraduate maths topics that are not yet in mathlib.

To make updates to this list, please make a pull request to mathlib after editing the yaml source file. This can be done entirely on GitHub, see "Editing files in another user's repository".

But still requires extensive human work.

If the future of mathematics is automatic formalization…

- Automatically review papers for correctness
- Enable every mathematician to work more efficiently and effectively
- Facilitate dynamic synthesis of literatures and new forms of metamathematics
- Broaden the possibilities of who can contribute to mathematics
- Provide access to and training in pure mathematics to those without

There are significant challenges remain to be overcome.

Four challenges for large-scale autoformalization:

- Challenge 1: Is the formalization the same?
- Challenge 2: Mind the (informal-formal) gap
- Challenge 3: Ensuring the proof proves
- Challenge 4: Informalization

All are about minimizing human effort for formalization.

ON A CONJECTURE OF MARTON

W. T. GOWERS, BEN GREEN, FREDDIE MANNERS, AND TERENCE TAO

ABSTRACT. We prove a conjecture of K. Marton, widely known as the polynomial Freiman–Ruzsa conjecture, in characteristic 2. The argument extends to odd characteristic, with details to follow in a subsequent paper.

1. INTRODUCTION

In this paper we prove a conjecture of Katalin Marton (see [28]), widely known in the literature as the *polynomial Freiman–Ruzsa conjecture* in characteristic 2. The conjecture has many equivalent forms, one of which is the following statement.

Conjecture 1.1. *Suppose that $A \subset \mathbf{F}_2^n$ is a set with $|A+A| \leqslant K|A|$. Then A is covered by at most $2K^C$ cosets[1] of some subgroup $H \leqslant \mathbf{F}_2^n$ of size at most $|A|$.*

Note that if A is covered by at most r cosets of a subgroup H of size at most $|A|$, then $|A+A| \leqslant (\binom{r}{2}+1)|A|$, so Conjecture 1.1 allows one to pass from the property of having a small doubling constant to the property of being contained in a small union of cosets of a subgroup of size at most $|A|$ and back again with at most a polynomial loss in the parameters.

The first result in this direction was due to Imre Ruzsa [28], who obtained an upper bound of $2K^2 2^{K^4}$ in place of the desired $2K^C$. It was Ruzsa [28], [29, Conjecture 2.2.2] who attributed Conjecture 1.1

Gowers, W. T., Green, B., Manners, F., & Tao, T. (2023). On a conjecture of Marton. arXiv preprint arXiv:2311.05762.

to Marton. Sanders [32] made a dramatic breakthrough by proving the first bounds of shape $\ll \exp(\log^{C_1} K)$, showing that $C_1 = 4+\varepsilon$ is permissible here. A variant of Sanders' argument due to Konyagin (not published by Konyagin, but described by Sanders in [33], and see also the comments on page 8 of [11]) showed that $C_1 = 3+\varepsilon$ is permissible. We also remark that Conjecture 1.1 (with worse values of C) was previously established in the special case when A was a downset in [9]. We refer to [7, 8, 20] for surveys on this conjecture.

Our main result is the following.

Theorem 1.2. *Conjecture 1.1 is true with $C = 12$.*

An elaboration of our arguments gives corresponding results in vector spaces over $\mathbf{F}_p$, p odd. This is notationally somewhat heavier and makes the underlying ideas slightly harder to appreciate so we give this argument in a separate paper [6].

Since the initial release of this preprint, Jyun-Jie Liao has given a refinement of the argument that improves the constant $C = 12$ in Theorem 1.2 to $C = 11$ (with corresponding numerical improvements to our other main results and applications); see [17, 18].

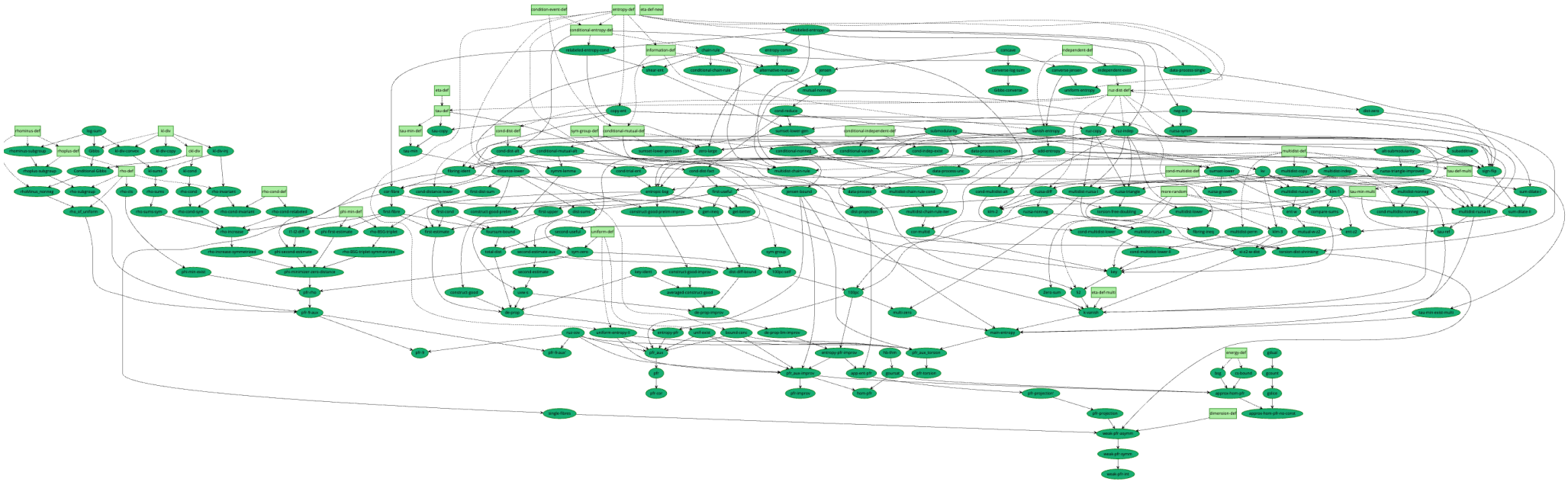
Applications. It was shown in [11, Theorem 1.11] that Theorem 1.8 (or Theorem 1.2) in characteristic 2 implies the so-called 'weak' polynomial Freiman–Ruzsa conjecture over $\mathbf{Z}$, so that is now a theorem as well.

Theorem 1.3. *Let A be a finite subset of $\mathbf{Z}^D$ for some D, and suppose that $|A+A| \leqslant K|A|$. Then there is some set $A' \subseteq A$, $|A'| \geqslant K^{-C_1/2}|A|$, with $\dim A' \leqslant C_2 \log K$, for some absolute[2] constants $C_1, C_2 > 0$.*

Theorem 1.3 solves a conjecture stated in [3, 11, 23, 26]. We remark that there is also a 'strong' PFR conjecture over $\mathbf{Z}$ (though its formulation requires care; see [21] and [24] for more discussion). This remains a challenging open problem.

As mentioned above, the polynomial Freiman–Ruzsa conjecture is equivalent to many other statements in additive combinatorics. We mention a few of these here.

Corollary 1.4. *There is a constant C_3 such that if $f : \mathbf{F}_2^m \to \mathbf{F}_2^n$ is a function such that the set $\{f(x)+f(y)-f(x+y) : x, y \in \mathbf{F}_2^m\}$ has*



Lemma 6.6 (Distance lower bound)

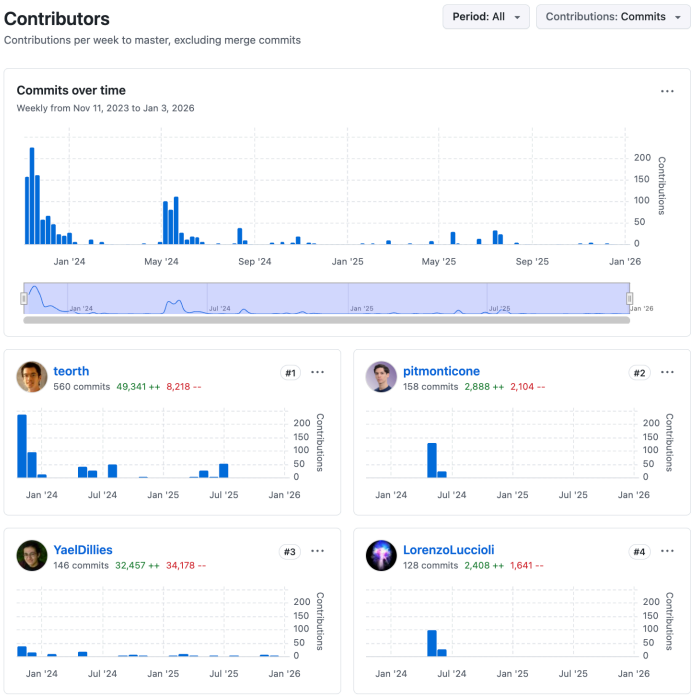
For any G-valued random variables X_1', X_2', one has

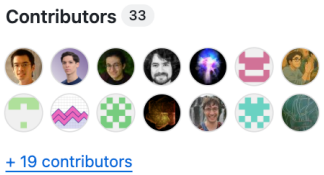
$$d[X_1'; X_2'] \geq k - \eta(d[X_1^0; X_1'] - d[X_1^0; X_1]) - \eta(d[X_2^0; X_2'] - d[X_2^0; X_2]).$$

LaTeX *Lean*

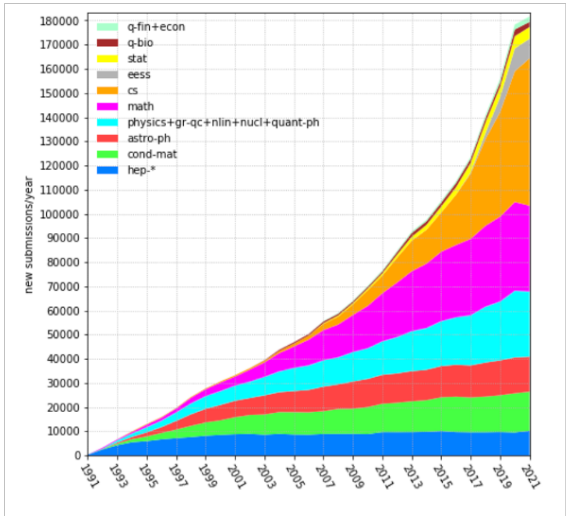
```
theorem distance_ge_of_min                                                source
    {Ω₀₁ : Type u_1} {Ω₀₂ : Type u_2} [MeasureTheory.MeasureSpace Ω₀₁]
    [MeasureTheory.MeasureSpace Ω₀₂] {G : Type uG} [AddCommGroup G]
    [MeasurableSpace G] (p : refPackage Ω₀₁ Ω₀₂ G) {Ω : Type u_3}
    {Ω'₁ : Type u_7} {Ω'₂ : Type u_8} [MeasureTheory.MeasureSpace Ω]
    [hΩ₁ : MeasureTheory.MeasureSpace Ω'₁]
    [hΩ₂ : MeasureTheory.MeasureSpace Ω'₂]
    [MeasureTheory.IsProbabilityMeasure MeasureTheory.volume]
    [MeasureTheory.IsProbabilityMeasure MeasureTheory.volume] {X₁ X₂ : Ω → G}
    {X₁' : Ω'₁ → G} {X₂' : Ω'₂ → G} (h : tau_minimizes p X₁ X₂)
    (h1 : Measurable X₁') (h2 : Measurable X₂') :
  d[X₁ # X₂] - p.η * (d[p.X₀₁ # X₁'] - d[p.X₀₁ # X₁]) - p.η * (d[p.X₀₂ # X₂'] -
      d[p.X₀₂ # X₂]) ≤ d[X₁' # X₂']
```

Imagine the cost of rolling out this effort across the math literature….





Math papers on arXiv: ~22% of 2,000,000+ papers (1991- 1/2022)

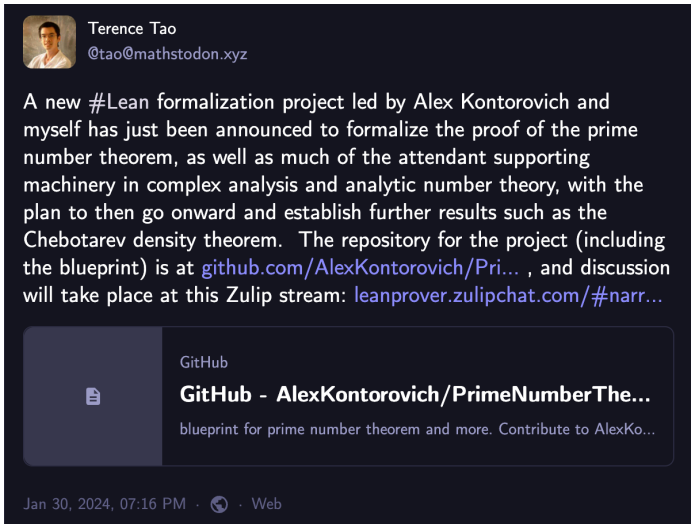


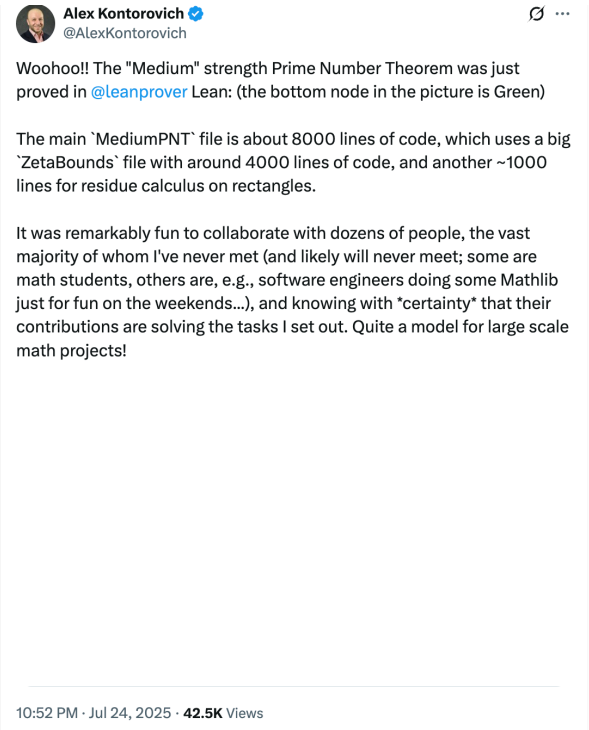
Even if an amazing advance in collaboration, the cost is prohibitive.

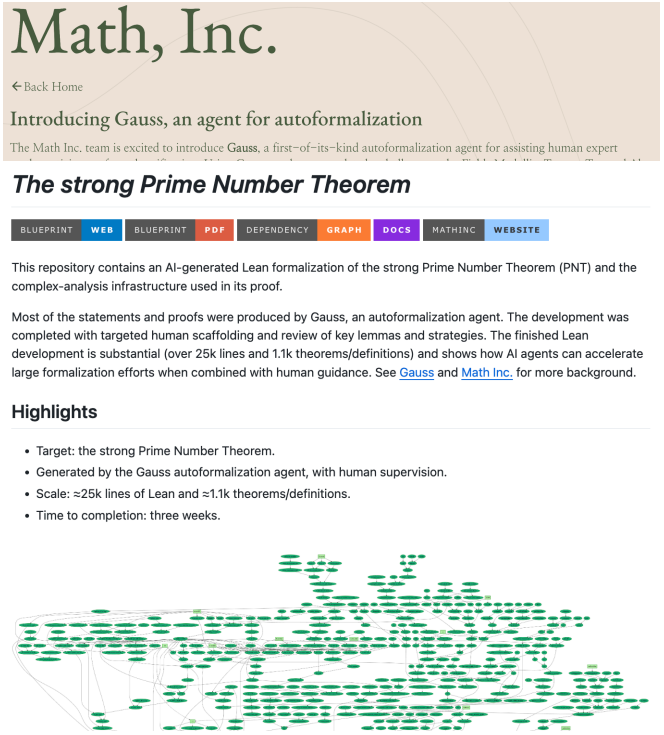
Amazing progress toward autoformalization in 2025…

Prime number theorem and beyond project…

Completed by autoformalization

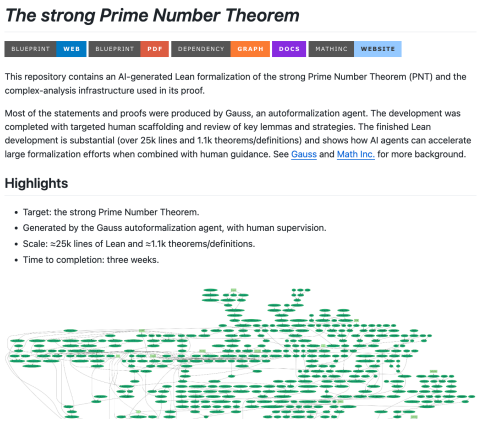






But human checked/corrected translations, manually elaborated proof…

Amazing progress toward autoformalization in 2025…

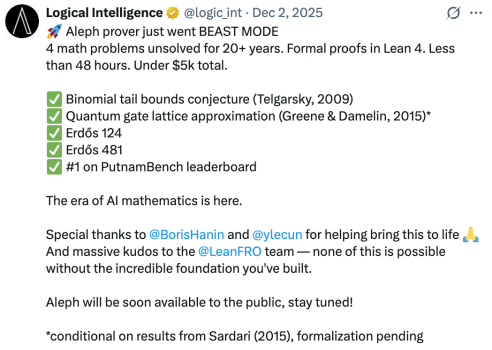


Math, Inc

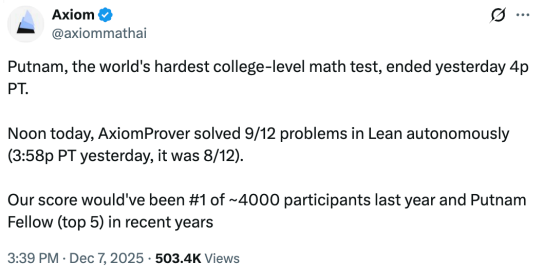
Jul 28, 2025 6:42 PM Eastern Daylight Time

Harmonic Announces IMO Gold Medal-Level Performance & Launch of First Mathematical Superintelligence (MSI) AI App

Harmonic



Logical Intelligence



Axiom

Amazing progress in 2025…

MATH-AI

The 5th Workshop on Mathematical Reasoning and AI

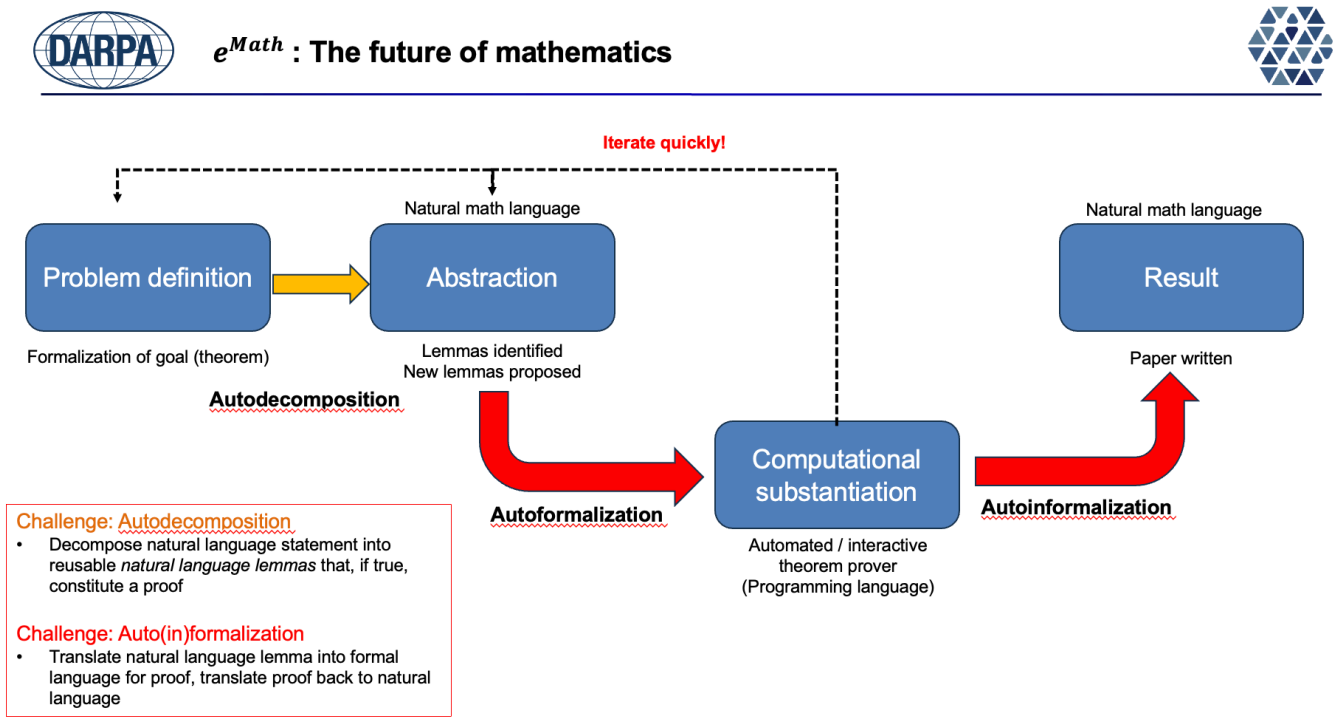
NeurIPS 2025, San Diego Convention Center (Upper Level Ballroom 6A), December 6th, 2025

Schedule

All times are in Pacific Time (PT)

8:25am - 8:30am	Opening Remarks
8:30am - 9:00am	Invited Talk: Weiyang Liu (Chinese University of Hong Kong), "Scalable Formal Verification Enables Novel Design"
9:00am - 9:30am	Invited Talk: Hannaneh Hajishirzi (UW & AI2), "Olmo 3 Model Flow for Mathematical Reasoning"
9:30am - 10:00am	Invited Talk: Max Tegmark (MIT), "Vericoding: Formally Verified Program Synthesis"
10:00am - 10:30am	Invited Talk: Leonardo de Moura (Lean FRO & AWS), "Teaching AI to Configure Proof Automation in Lean"
10:30am - 11:15am	Panel Discussion: Tengyu Ma (Stanford), Tom Kalil (Renaissance Philanthropy), Patrick Shafto (DARPA & Rutgers), Jonathan Thomm (Harmonic)

Plug for my DARPA Exponentiating Mathematics (expMath) program…



Publicly announcing teams imminently.

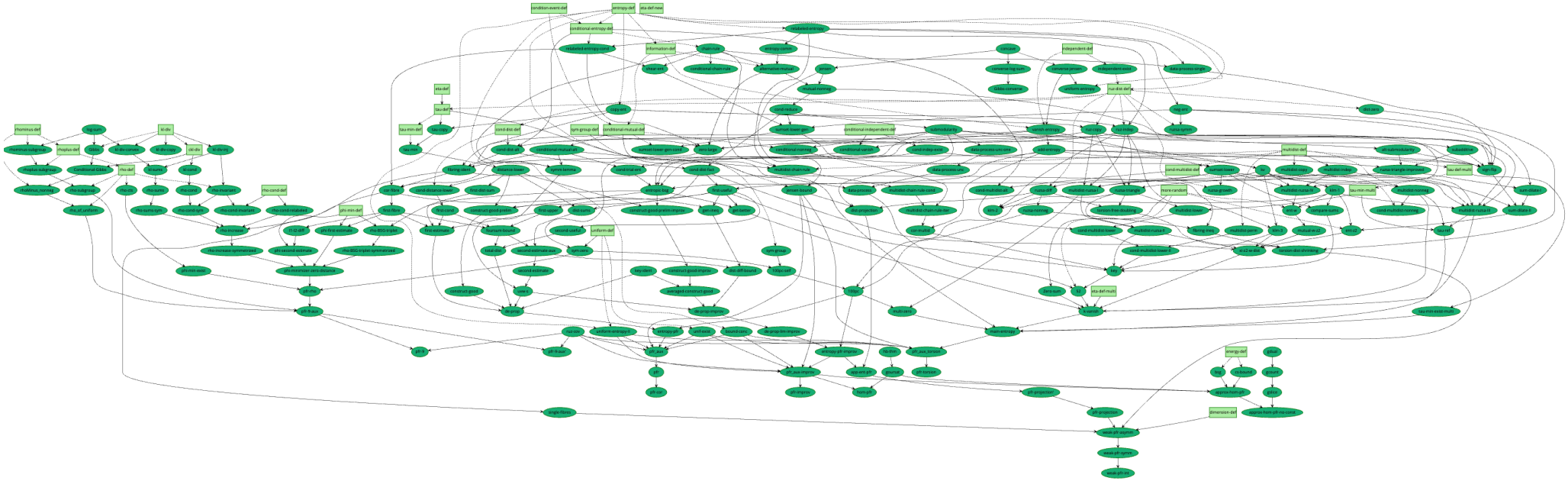
Contact me if you want to receive updates, join events, etc.

Four challenges for large-scale autoformalization:

- Challenge 1: Is the formalization the same?
- Challenge 2: Mind the (informal-formal) gap
- Challenge 3: Ensuring the proof proves
- Challenge 4: Informalization

All are about minimizing human effort for formalization.

Challenge 1: Is the formalization the same?



Lemma 6.6 (Distance lower bound)

For any G-valued random variables X_1', X_2', one has

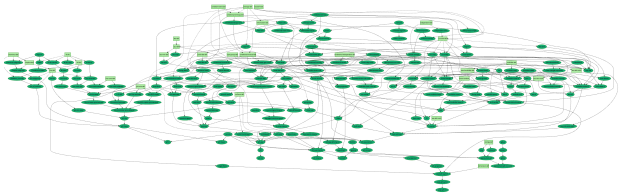
$$d[X_1'; X_2'] \geq k - \eta(d[X_1^0; X_1'] - d[X_1^0; X_1]) - \eta(d[X_2^0; X_2'] - d[X_2^0; X_2]).$$

LaTeX *Lean*

```
theorem distance_ge_of_min                                                source
    {Ω₀₁ : Type u_1} {Ω₀₂ : Type u_2} [MeasureTheory.MeasureSpace Ω₀₁]
    [MeasureTheory.MeasureSpace Ω₀₂] {G : Type uG} [AddCommGroup G]
    [MeasurableSpace G] (p : refPackage Ω₀₁ Ω₀₂ G) {Ω : Type u_3}
    {Ω'₁ : Type u_7} {Ω'₂ : Type u_8} [MeasureTheory.MeasureSpace Ω]
    [hΩ₁ : MeasureTheory.MeasureSpace Ω'₁]
    [hΩ₂ : MeasureTheory.MeasureSpace Ω'₂]
    [MeasureTheory.IsProbabilityMeasure MeasureTheory.volume]
    [MeasureTheory.IsProbabilityMeasure MeasureTheory.volume] {X₁ X₂ : Ω → G}
    {X₁' : Ω'₁ → G} {X₂' : Ω'₂ → G} (h : tau_minimizes p X₁ X₂)
    (h1 : Measurable X₁') (h2 : Measurable X₂') :
  d[X₁ # X₂] - p.η * (d[p.X₀₁ # X₁'] - d[p.X₀₁ # X₁]) - p.η * (d[p.X₀₂ # X₂'] -
      d[p.X₀₂ # X₂]) ≤ d[X₁' # X₂']
```

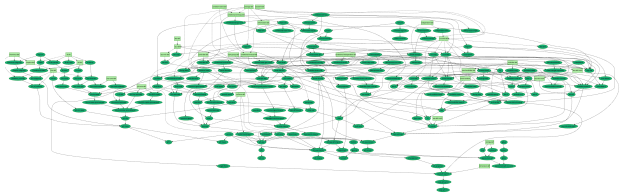
Formal-informal alignment

Existing literature: alignment of implications



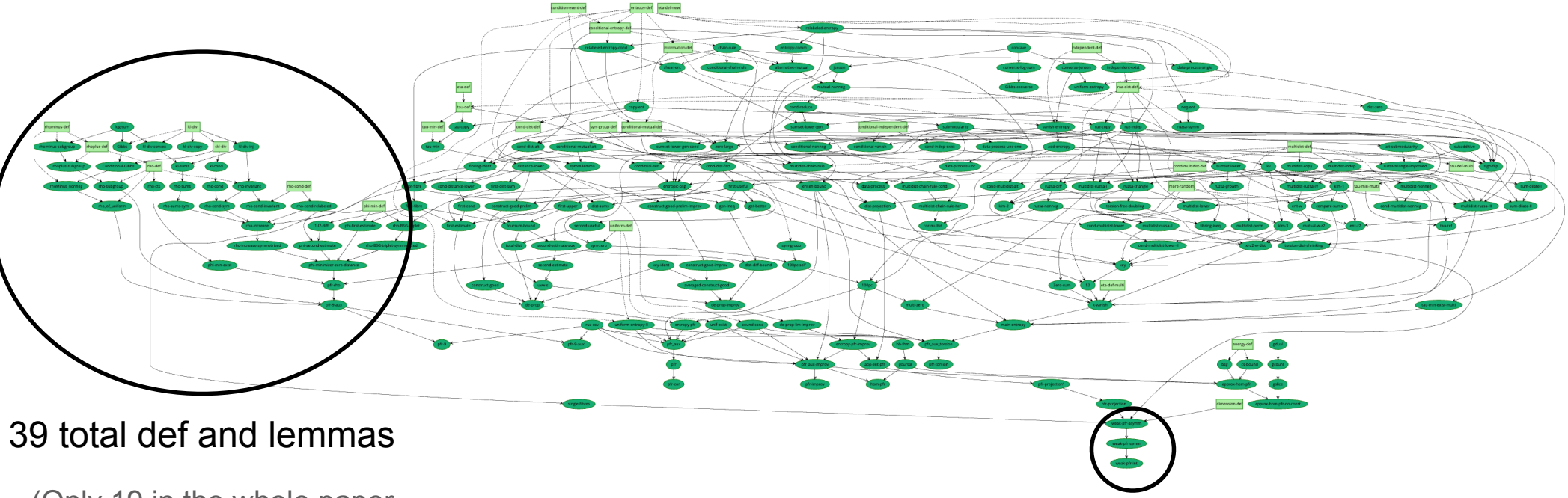
For a node high in the graph, derive formal implications. Compare those formal implications to informally stated implications.

Future literature: Need to ensure formalized math has informal counterpart



For a node low in the graph, may not have informal comparison. Would need to informalize and backward chain argument. (See challenge 4)

Challenge 2: Filling in the gaps



39 total def and lemmas

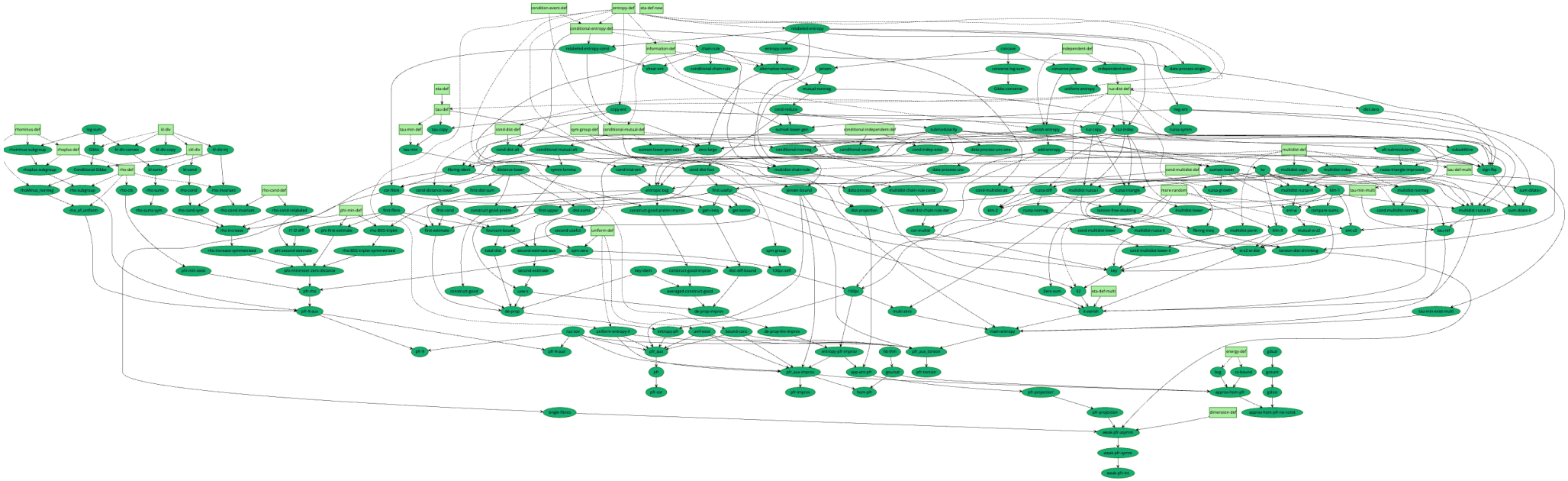
(Only 19 in the whole paper, not including equations)

Thm 1.3

Informal math:

- Omits information considered unnecessary
- Organizes information to prioritize important

Curriculum for progress toward gap filling: Progressive missing value prediction



Lemma 6.6 (Distance lower bound)

For any G-valued random variables X_1', X_2', one has

$$d[X_1'; X_2'] \geq k - \eta(d[X_1^0; X_1'] - d[X_1^0; X_1]) - \eta(d[X_2^0; X_2'] - d[X_2^0; X_2]).$$

LaTeX *Lean*

theorem distance_ge_of_min source

```
{Ω₀₁ : Type u_1} {Ω₀₂ : Type u_2} [MeasureTheory.MeasureSpace Ω₀₁]
[MeasureTheory.MeasureSpace Ω₀₂] {G : Type uG} [AddCommGroup G]
[MeasurableSpace G] (p : refPackage Ω₀₁ Ω₀₂ G) {Ω : Type u_3}
{Ω'₁ : Type u_7} {Ω'₂ : Type u_8} [MeasureTheory.MeasureSpace Ω]
[hΩ₁ : MeasureTheory.MeasureSpace Ω'₁]
[hΩ₂ : MeasureTheory.MeasureSpace Ω'₂]
[MeasureTheory.IsProbabilityMeasure MeasureTheory.volume]
[MeasureTheory.IsProbabilityMeasure MeasureTheory.volume] {X₁ X₂ : Ω → G}
{X₁' : Ω'₁ → G} {X₂' : Ω'₂ → G} (h : tau_minimizes p X₁ X₂)
(h1 : Measurable X₁') (h2 : Measurable X₂') :
d[X₁ # X₂] - p.η * (d[p.X₀₁ # X₁'] - d[p.X₀₁ # X₁]) - p.η * (d[p.X₀₂ # X₂'] -
    d[p.X₀₂ # X₂]) ≤ d[X₁' # X₂']
```

Challenge 3: Ensuring the proof proves

channel: new members | topic: Budden's claims about Navier Stokes | near: 566150614 | Search

new members > **Budden's claims about Navier Stokes** JAN 3

Jason Rute 7:43 AM

I just want to point out that the Lean code Budden has posted (at least one screen shot, 2 files on separate occasions, and a file he shared privately with Elliot Glazer---but tweeted about sharing) tell a fairly simple story. At first, his files were very circular in their logic with lots of axioms and assumptions but which also didn't compile. After coaching from Elliot, it seems Budden has learned how to use Lean correctly. His only Lean file which typechecks is also quite incomplete (and I think it uses the sup norm when he meant to use the L2 norm FWIW). It is not a complete proof of NS by any means. And finally he posted a screen shot showing the many errors he is still trying to fix, which is the last we heard of him.

I know there is a lot of worry that his "proof" is using Lean exploits, but it really seems that he has never had anything remotely resembling a working Lean proof, and any public claims to that effect on Budden's part have been ambitious promises to deliver, not based on him having a "working" Lean proof up his sleeve. (I'm mostly getting this by him saying things about writing all the code at the last minute, implying that he never had full Lean proof. And also that none of his code seems to be very complicated.) 7:43 AM

👍 Mirek Olšák, Elliot Glazer, Eric Vergo

I'm personally not worried about exploits for a Lean millennium problem proof. This community would quickly scrutinize a Lean millennium problem solution, and be able to reduce it to either interesting non-trivial mathematics or to trivial broken Lean. We have many tools for inspecting proofs and guarding against exploits and we would come up with many more if a Millennium problems was involved. (And no journal or prize committee would be quick to announce a resolution until the Lean community was happy with the Lean proof.) I worry a lot more about the little theorems that aren't going to get the attention of a Millennium problem (which is why I keep harping about coming up principled and complete standards of correctness for Lean proofs). 7:43 AM

👍 7 💯 2

Elliot Glazer
@ElliotGlazer

Congratulations to my good ol' pal and colleague James Hanson for proving FLT in Lean with standard axioms **in a way that fools SafeVerify and lean4checker!**

Surely this time it's legit right?

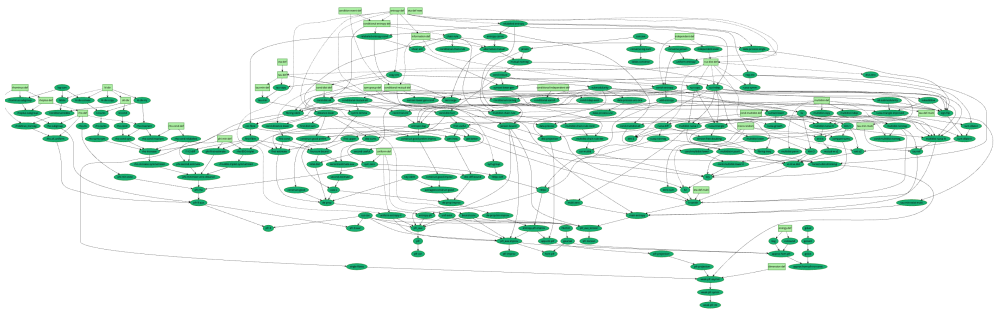
lean4 > SafeVerify-passing kernel soundness exploit

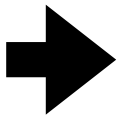
James E Hanson EDITED

```
import Mathlib.Data.Nat.Basic
import Mathlib.Tactic
import Mathlib.NumberTheory.FLT.Basic
import Lean

open Lean

unsafe def badNatUnsafe : ℕ := unsafeCast (-4294967296 : ℤ)

@[implemented_by badNatUnsafe] opaque badNatVal : ℕ

run_elab
  addDecl <| .defnDecl {
    name        := .str .anonymous "badNat"
    levelParams := []
    type        := .const ``Nat []
    value       := .lit <| .natVal badNatVal
    hints       := .opaque
    safety      := .safe
  }

theorem FLT : FermatLastTheorem := by
  have truly_marvelous_0 : ¬badNat ≤ 9223372036854775807 := by decide
  have truly_marvelous_1 : ¬9223372036854775807 < badNat := by decide
  simp_all only [not_le]

#print axioms FLT
-- 'FLT' depends on axioms: [propext]
```

5:14 PM · Jan 4, 2026 · **5,234** Views

Challenge 4: Informalization





ON A CONJECTURE OF MARTON

W. T. GOWERS, BEN GREEN, FREDDIE MANNERS, AND TERENCE TAO

ABSTRACT. We prove a conjecture of K. Marton, widely known as the polynomial Freiman–Ruzsa conjecture, in characteristic 2. The argument extends to odd characteristic, with details to follow in a subsequent paper.

1. INTRODUCTION

In this paper we prove a conjecture of Katalin Marton (see [28]), widely known in the literature as the *polynomial Freiman–Ruzsa conjecture* in characteristic 2. The conjecture has many equivalent forms, one of which is the following statement.

Conjecture 1.1. *Suppose that $A \subset \mathbf{F}_2^n$ is a set with $|A+A| \leqslant K|A|$. Then A is covered by at most $2K^C$ cosets[1] of some subgroup $H \leqslant \mathbf{F}_2^n$ of size at most $|A|$.*

Note that if A is covered by at most r cosets of a subgroup H of size at most $|A|$, then $|A+A| \leqslant (\binom{r}{2}+1)|A|$, so Conjecture 1.1 allows one to pass from the property of having a small doubling constant to the property of being contained in a small union of cosets of a subgroup of size at most $|A|$ and back again with at most a polynomial loss in the parameters.

The first result in this direction was due to Imre Ruzsa [28], who obtained an upper bound of $2K^2 2^{K^4}$ in place of the desired $2K^C$. It was Ruzsa [28], [29, Conjecture 2.2.2] who attributed Conjecture 1.1

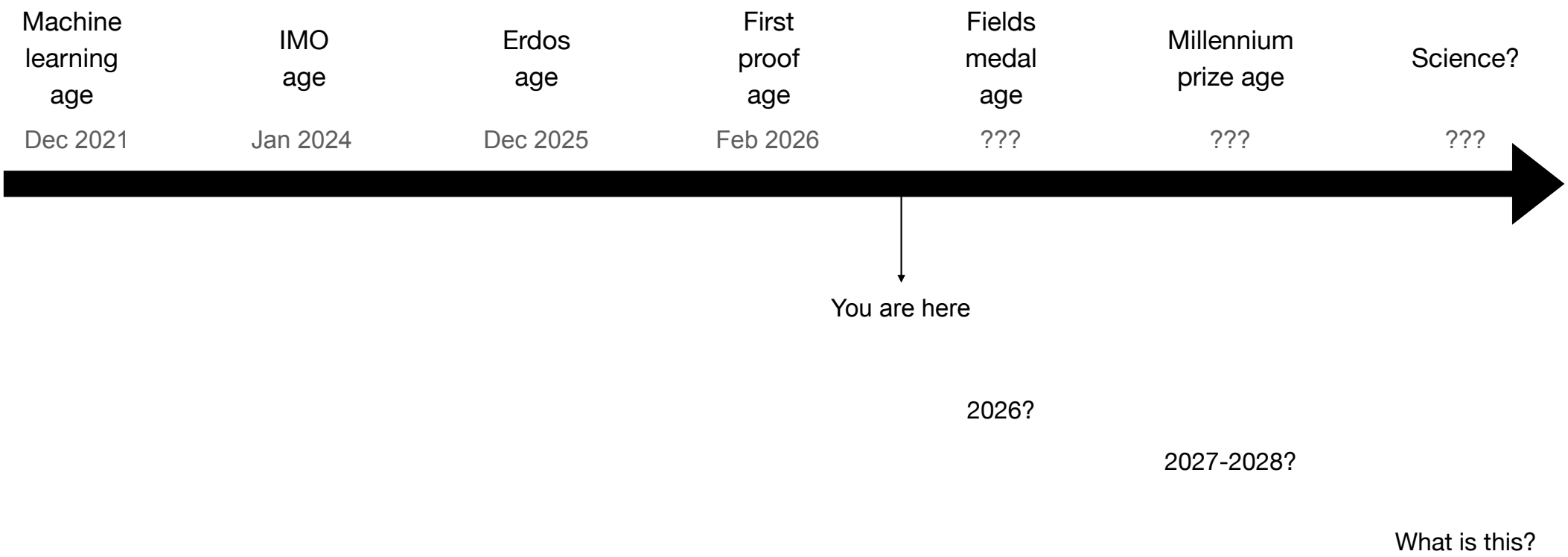
Formalized proofs contain more information than humans deemed necessary.
How to effectively translate results from code to human useable form?

Four challenges for large-scale autoformalization:

- Challenge 1: Is the formalization the same?
- Challenge 2: Mind the (informal-formal) gap
- Challenge 3: Ensuring the proof proves
- Challenge 4: Informalization

All are about minimizing human effort for formalization.

Ages of AI math



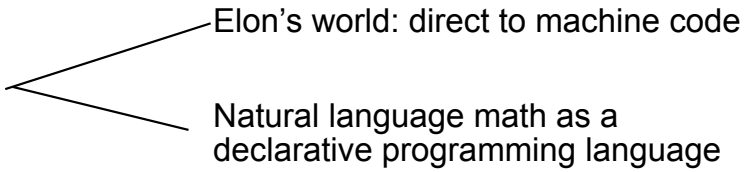
If the future of ~~mathematics~~ Science is automatic formalization…

- Automatically review papers for correctness
- Enable every mathematician Scientist to work more efficiently and effectively
- Facilitate dynamic synthesis of literatures and new forms of metamathematics Science
- Broaden the possibilities of who can contribute to mathematics Science
- Provide access to pure and training in pure ~~mathematics~~ Science to those without access

There are significant challenges remain to be overcome.

Three example cases of formalization in science:

- Example 1: Computer science (programming)



- Example 2: Quantum computing / information

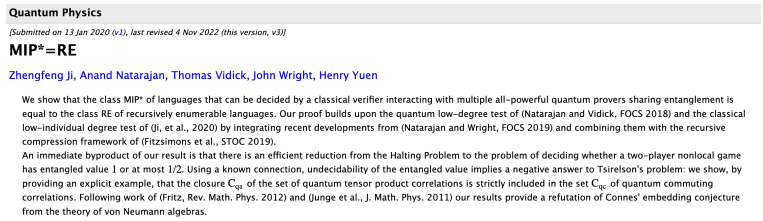
- Example 3: Behavioral science

(Many others: material science, chemistry, biology, etc)

Three example cases of formalization in science:

- Example 1: Computer science (programming)
- Example 2: Quantum computing / information
- Example 3: Behavioral science

Domain is amenable. Discussion of an effort at formalization.



Quantum Physics

[Submitted on 13 Jan 2020 (v1), last revised 4 Nov 2022 (this version, v3)]

MIP*=RE

Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, Henry Yuen

We show that the class MIP* of languages that can be decided by a classical verifier interacting with multiple all-powerful quantum provers sharing entanglement is equal to the class RE of recursively enumerable languages. Our proof builds upon the quantum low-degree test of (Natarajan and Vidick, FOCS 2018) and the classical low-individual degree test of (Ji, et al., 2020) by integrating recent developments from (Natarajan and Wright, FOCS 2019) and combining them with the recursive compression framework of (Fitzsimons et al., STOC 2019).
An immediate byproduct of our result is that there is an efficient reduction from the Halting Problem to the problem of deciding whether a two-player nonlocal game has entangled value 1 or at most 1/2. Using a known connection, undecidability of the entangled value implies a negative answer to Tsirelson's problem: we show, by providing an explicit example, that the closure C_{qs} of the set of quantum tensor product correlations is strictly included in the set C_{qc} of quantum commuting correlations. Following work of (Fritz, Rev. Math. Phys. 2012) and (Junge et al., J. Math. Phys. 2011) our results provide a refutation of Connes' embedding conjecture from the theory of von Neumann algebras.

(Many others: material science, chemistry, biology, etc)

Quantum Physics

[Submitted on 13 Jan 2020 (v1), last revised 4 Nov 2022 (this version, v3)]

MIP*=RE

Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, Henry Yuen

We show that the class MIP* of languages that can be decided by a classical verifier interacting with multiple all-powerful quantum provers sharing entanglement is equal to the class RE of recursively enumerable languages. Our proof builds upon the quantum low-degree test of (Natarajan and Vidick, FOCS 2018) and the classical low-individual degree test of (Ji, et al., 2020) by integrating recent developments from (Natarajan and Wright, FOCS 2019) and combining them with the recursive compression framework of (Fitzsimons et al., STOC 2019).

An immediate byproduct of our result is that there is an efficient reduction from the Halting Problem to the problem of deciding whether a two-player nonlocal game has entangled value 1 or at most $1/2$. Using a known connection, undecidability of the entangled value implies a negative answer to Tsirelson's problem: we show, by providing an explicit example, that the closure C_{qa} of the set of quantum tensor product correlations is strictly included in the set C_{qc} of quantum commuting correlations. Following work of (Fritz, Rev. Math. Phys. 2012) and (Junge et al., J. Math. Phys. 2011) our results provide a refutation of Connes' embedding conjecture from the theory of von Neumann algebras.

Three example cases of formalization in science:

- Example 1: Computer science (programming)
- Example 2: Quantum computing / information
- Example 3: Behavioral science

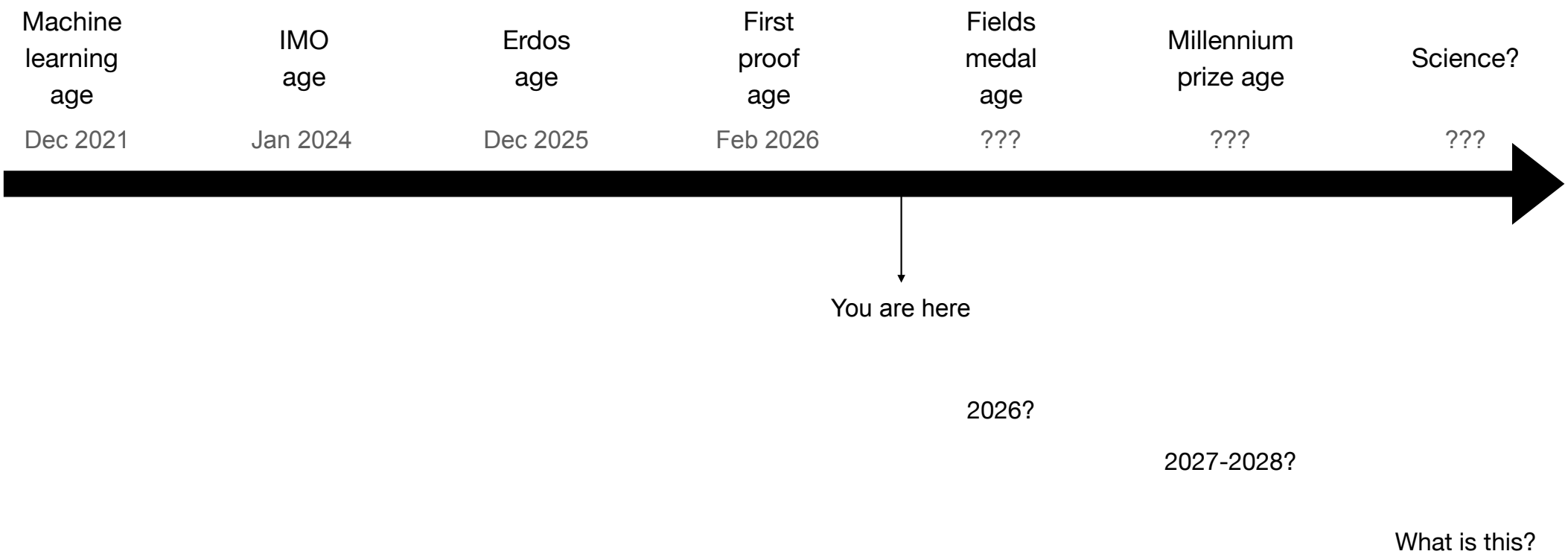
Domain is not amenable. Not totally clear what to formalize, given the limited state of theory.

One candidate is formalization of preregistration?

Despite obvious challenges, biggest potential win here.

(Many others: material science, chemistry, biology, etc)

Ages of AI math



Autoformalization and the future of mathematics

Patrick Shafto
Professor / Program Manager
Rutgers / DARPA