Mise en perspective

Amaury Hayat
CERMICS - Ecole des Ponts - Institut Polytechnique de Paris
Korean Institute for Advanced Studies

AI4Maths Workshop of the SMF - SMAI - SFdS

–

Institut Henri Poincaré

–

November 18th 2025

Introduction

Social networks

Emails and apps

Virtual assistants

Platforms

Introduction

From computer scientists awareness...

- Convolutional Neural Networks (LeNet-5, 1998,..., ResNet, 2015)
- Generative adversarial networks (2014)
- Transformer (2017)
- AlphaZero (2018)

... to global awareness

- ChatGPT (followed by Llama, Mistral, etc.)
- Diffusion model (Dall·E, Midjourney, Stable Diffusion)
- 2024 Nobel Prizes in Physics **and** in Chemistry

Introduction

An idea that started taking roots:

How can AI help mathematicians ?

Basic usage:

- asking a math question to an LLM / asking to prove a small lemma

Introduction

An idea that started taking roots:

How can AI help mathematicians ?

Basic usage:

- asking a math question to an LLM / asking to prove a small lemma
  - terrible answers (2022-2023)

Introduction

An idea that started taking roots:

How can AI help mathematicians ?

Basic usage:

- asking a math question to an LLM / asking to prove a small lemma
  - terrible answers (2022-2023)
  - but models are evolving quickly

    Deepseek R3; OpenAI o3,o4, GPT-5; Mistral Magistral, Gemini-2.5 Pro

Introduction

Lemma 3. The following holds: $\partial_\eta f_M(\gamma^*, 0) \neq 0$.

Proof. From the classical formula of the differential of the determinant,

$$\partial_\eta f_M(\gamma^*, 0) = \mathrm{Tr}(\mathrm{com}(A - Id)^T \partial_\eta \hat{m}(\gamma^*, 0)) = \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right) \mathrm{Tr}(\mathrm{com}(A - Id)^T A). \tag{2.54}$$

Since

$$\begin{aligned} \mathrm{Tr}(\mathrm{com}(A - Id)^T A) &= \mathrm{Tr}(\mathrm{com}(A - Id)^T (A - Id)) + \mathrm{Tr}(\mathrm{com}(A - Id)^T) \\ &= \mathrm{Tr}((A - Id)\mathrm{com}(A - Id)^T) + \mathrm{Tr}(\mathrm{com}(A - Id)^T) \\ &= 4\det(A - Id) + \mathrm{Tr}(\mathrm{com}(A - Id)^T), \end{aligned} \tag{2.55}$$

we have, since $\det(A - Id) = 0$

$$\partial_\eta f_M(\gamma^*, 0) = \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right) \mathrm{Tr}(\mathrm{com}(A - Id)^T). \tag{2.56}$$

From this, only two cases can occur. Either $\mathrm{rank}(A - I) = 3$ in which case, since

$$(A - Id)\mathrm{com}(A - Id)^T = \det(A - Id) Id = 0, \tag{2.57}$$

then each column of $\mathrm{com}(A - Id)^T$ belongs to $\mathrm{Ker}(A - I)$ which has dimension 1 so $\mathrm{rank}(\mathrm{com}(A - Id)^T)$ is at most equal to 1. And, since $\mathrm{rank}(A - I) = 3$, there is at least one principal minor that is

Introduction

Timothy Gowers @wtgowers @wtgowers · Oct 31

I crossed an interesting threshold yesterday, which I think many other mathematicians have been crossing recently as well. In the middle of trying to prove a result, I identified a statement that looked true and that would, if true, be useful to me. 1/3

51 336 2K 611K

Timothy Gowers @wtgowers @wtgowers · Oct 31

Replying to @wtgowers

the time it would have taken me would probably have been of order of magnitude an hour (an estimate that comes with quite wide error bars). So it looks as though we have entered the brief but enjoyable era where our research is greatly sped up by AI but AI still needs us. 3/3

17 37 964 71K

Timothy Gowers @wtgowers @wtgowers · Oct 31

Introduction

Sebastien Bubeck @SebastienBubeck

Claim: gpt-5-pro can prove new interesting mathematics.

Proof: I took a convex optimization paper with a clean open problem in it and asked gpt-5-pro to work on it. It proved a better bound than what is in the paper, and I checked the proof it's correct.

Details below.

ChatGPT 5 Pro

2503.10138v1.pdf
PDF

Can you improve the condition on the step-size in Theorem 1? I don't want to add any more hypothesis, I just want you to work to improve the step-size condition under the same assumptions as Theorem 1.

Request for GPT-5 Pro
Reasoned for 17m 35s

Yes. Under the *same* assumptions as Theorem 1 (convex L-smooth f and fixed step size for gradient descent), you can enlarge the guaranteed-convexity window from $\eta \leq 1/L$ to

$$\boxed{\eta \leq \tfrac{3}{2L}}$$

i.e., the optimization curve $n \mapsto f(x_n)$ is convex for any step size $\eta \in (0, 3/(2L)]$. This strictly improves Theorem 1's bound and still lies below the non-convexity regime exhibited in Theorem 2. See Theorem 1

Sebastien Bubeck @SebastienBubeck · Aug 20

By the way this is the proof it came up with:

Write $g_k := \nabla f(x_k)$ and $\Delta_k := g_{k+1} - g_k$. We compare two consecutive decreases D_k and D_{k+1}.

1. **Lower bound for D_k with a Bregman term.**
   For convex L-smooth f, the Bregman divergence obeys

$$\frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \leq f(x) - f(y) - \langle \nabla f(y), x - y\rangle$$

(see the inequality used repeatedly in §3, Eq. (3) / Theorem 2.1.5 of Nesterov as cited there). Applying it with $x = x_k$, $y = x_{k+1}$ and noting $x_k - x_{k+1} = \eta g_k$ gives

$$D_k \geq \eta \langle g_{k+1}, g_k\rangle + \frac{1}{2L}\|\Delta_k\|^2. \quad \text{(A)}$$

2. **Upper bound for D_{k+1} by convexity.**
   By convexity, $f(x) - f(y) \leq \langle \nabla f(x), x - y\rangle$, so with $x = x_{k+1}$, $y = x_{k+2}$ and $x_{k+1} - x_{k+2} = \eta g_{k+1}$,

$$D_{k+1} \leq \eta \|g_{k+1}\|^2. \quad \text{(B)}$$

3. **Subtract and use cocoercivity once.**
   From (A)–(B),

$$D_k - D_{k+1} \geq \eta \langle g_{k+1}, g_k - g_{k+1}\rangle + \frac{1}{2L}\|\Delta_k\|^2 = -\eta \langle g_{k+1}, \Delta_k\rangle + \frac{1}{2L}\|\Delta_k\|^2.$$

Since $\langle g_{k+1}, \Delta_k\rangle = \langle g_k, \Delta_k\rangle + \|\Delta_k\|^2$,

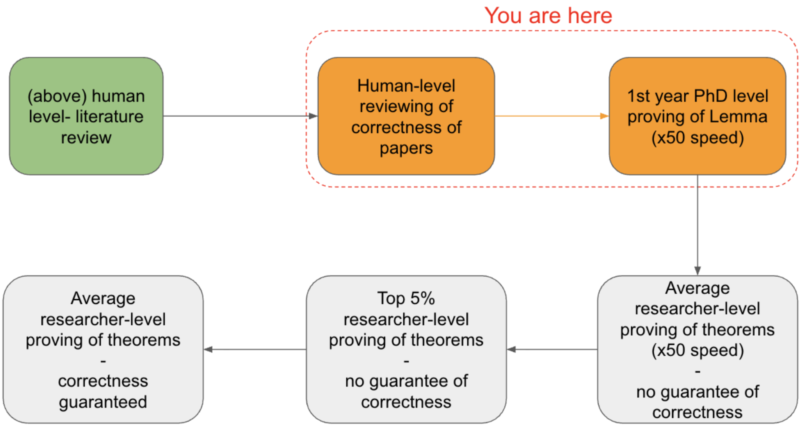
$$D_k - D_{k+1} \geq -\eta \langle g_k, \Delta_k\rangle + \left(\frac{1}{2L} - \eta\right)\|\Delta_k\|^2.$$

Now apply the standard cocoercivity inequality (Eq. (3) in §3),

$$\langle \Delta_k,\ x_{k+1} - x_k\rangle \geq \frac{1}{L}\|\Delta_k\|^2,$$

and use $x_{k+1} - x_k = -\eta g_k$ to get $-\eta \langle g_k, \Delta_k\rangle \geq \frac{1}{L}\|\Delta_k\|^2$. Therefore,

Introduction



Introduction

ArxivMathBench (Peyronnet, Glöckle, A.H., 2025): a live benchmark of Lemma extracted from latest arxiv papers

- Directly a research problem
- Mitigate data-contamination by updating regularly from the latest arxiv paper

Model	**Proof Accept. (%)**	**Human confidence score (%)**
GPT-5	12.3	86
Gem-2.5	7	88
Deepseek-R	11.9	81

Introduction

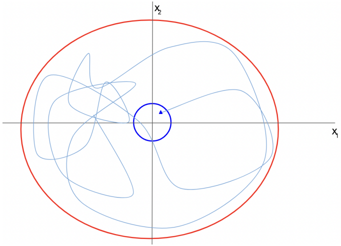
ArxivMathBench (Peyronnet, Glöckle, A.H., 2025): a live benchmark of Lemma extracted from latest arxiv papers

- Directly a research problem
- Mitigate data-contamination by updating regularly from the latest arxiv paper

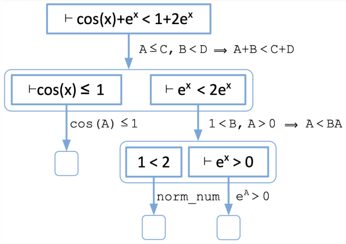
Model	**Proof Accept. (%)**	**Human confidence score (%)**
GPT-5	12.3	86
Gem-2.5	7	88
Deepseek-R	11.9	81

AI for mathematics $\neq$ generic LLM

Outline of the talk

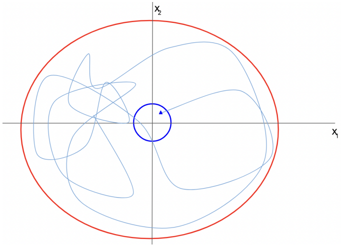


1. AI as a tool for math discovery

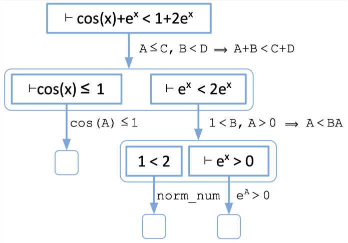


2. Will AI prove theorems on its own?

Outline of the talk

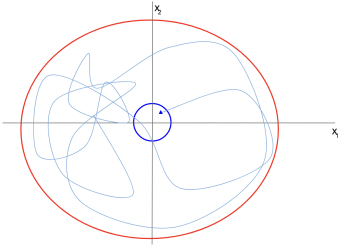


1. AI as a tool for math discovery

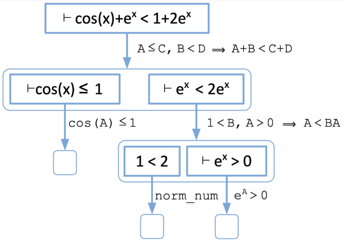


2. When will AI prove theorems on its own?

Outline of the talk



1. AI as a tool for math discovery

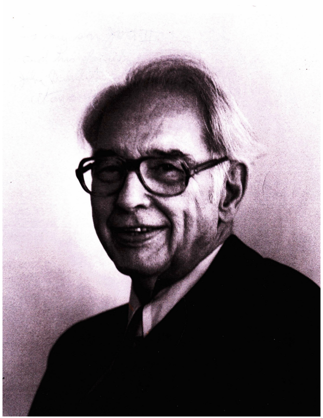


2. When will AI prove theorems on its own?

Solving maths problems with a computer

Solving maths problems with a computer

- Atanasoff-Berry Computer (designed 1937, created in 1942).
- First electronic computer
- Solving systems of linear equations



Solving maths problems with a computer

Conjecture (Euler, 1769)

If there exist integers a_1, a_2,...,a_k, b, *and* n *such that*

$$a_1^n + a_2^n + \ldots + a_k^n = b^n,$$

then $k \geq n$.

Solving maths problems with a computer

Conjecture (Euler, 1769)

If there exist integers a_1, a_2,...,a_k, b, and n such that

$$a_1^n + a_2^n + \ldots + a_k^n = b^n,$$

then $k \geq n$.

A problem open for almost 200 years

Solving maths problems with a computer

Lander and Parkin (1966)

Solving maths problems with a computer

Lander and Parkin (1966)

$$27^5 + 84^5 + 110^5 + 133^5 = 144^5$$

Solving maths problems with a computer

COUNTEREXAMPLE TO EULER'S CONJECTURE ON SUMS OF LIKE POWERS

BY L. J. LANDER AND T. R. PARKIN

Communicated by J. D. Swift, June 27, 1966

A direct search on the CDC 6600 yielded

$$27^5 + 84^5 + 110^5 + 133^5 = 144^5$$

as the smallest instance in which four fifth powers sum to a fifth power. This is a counterexample to a conjecture by Euler [1] that at least n nth powers are required to sum to an nth power, $n > 2$.

REFERENCE

1. L. E. Dickson, *History of the theory of numbers*, Vol. 2, Chelsea, New York, 1952, p. 648.

Conclusion: computers have been used to prove theorems for a long time.

Conclusion: computers have been used to prove theorems for a long time.

Can AI be useful to solve more complicated problems?

Problems where the difficulty is not just a high number of case-checking?

AI for mathematical discovery

- Paradigm: Intuition, an important concept in mathematics

AI for mathematical discovery

- Paradigm: Intuition, an important concept in mathematics
- In a number of open problems, intuition resembles a kind of pattern recognition. You've seen plenty of examples, and this gives you an idea of how to proceed in a case you've never seen.

AI for mathematical discovery

- Paradigm: Intuition, an important concept in mathematics
- In a number of open problems, intuition resembles a kind of pattern recognition. You've seen plenty of examples, and this gives you an idea of how to proceed in a case you've never seen.
- Can we train an AI to have better mathematical intuition than us?

AI for mathematical discovery

- Paradigm: Intuition, an important concept in mathematics
- In a number of open problems, intuition resembles a kind of pattern recognition. You've seen plenty of examples, and this gives you an idea of how to proceed in a case you've never seen.
- Can we train an AI to have better mathematical intuition than us?

Yes

AI for mathematical discovery

- Paradigm: Intuition, an important concept in mathematics
- In a number of open problems, intuition resembles a kind of pattern recognition. You've seen plenty of examples, and this gives you an idea of how to proceed in a case you've never seen.
- Can we train an AI to have better mathematical intuition than us?

Yes

One example from one field of mathematics: stability of dynamical systems

A system of differential equations

$$\dot{x}(t) = f(x(t)),$$

where $x(t) \in \mathbb{R}^n$, $f \in C^1(\mathbb{R}^n)$ and $f(0) = 0$.

A system of differential equations

$$\dot{x}(t) = f(x(t)),$$

where $x(t) \in \mathbb{R}^n$, $f \in C^1(\mathbb{R}^n)$ and $f(0) = 0$.

Question (System Stability)

Is it true that for every $\varepsilon > 0$, there exists $\delta > 0$ such that if the initial condition satisfies $\|x(0)\| \leq \delta$ then the solution $x(t)$ exists for all $t \in [0, +\infty)$ and

$$\|x(t)\| \leq \varepsilon, \ \forall\, t \in [0, +\infty).$$

A system of differential equations

$$\dot{x}(t) = f(x(t)),$$

where $x(t) \in \mathbb{R}^n$, $f \in C^1(\mathbb{R}^n)$ and $f(0) = 0$.

Question (System Stability)

Is it true that for every $\varepsilon > 0$, there exists $\delta > 0$ such that if the initial condition satisfies $\|x(0)\| \leq \delta$ then the solution $x(t)$ exists for all $t \in [0, +\infty)$ and

$$\|x(t)\| \leq \varepsilon, \ \forall\, t \in [0, +\infty).$$

- Are all solutions arbitrarily bounded if the initial condition is sufficiently small?

A system of differential equations

$$\dot{x}(t) = f(x(t)),$$

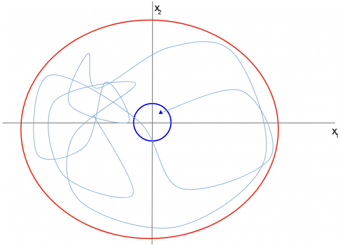
where $x(t) \in \mathbb{R}^n$, $f \in C^1(\mathbb{R}^n)$ and $f(0) = 0$.

Question (System Stability)

Is it true that for every $\varepsilon > 0$, there exists $\delta > 0$ such that if the initial condition satisfies $\|x(0)\| \leq \delta$ then the solution $x(t)$ exists for all $t \in [0, +\infty)$ and

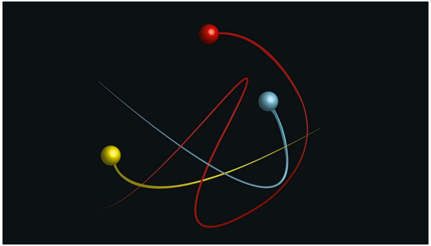
$$\|x(t)\| \leq \varepsilon, \ \forall\, t \in [0, +\infty).$$

- Are all solutions arbitrarily bounded if the initial condition is sufficiently small?

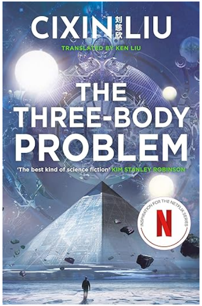


Stability of Dynamical Systems

A problem that has interested mathematicians for over a hundred years.



Stability of Dynamical Systems



Stability of Dynamical Systems

A significant advancement: Lyapunov functions

Theorem

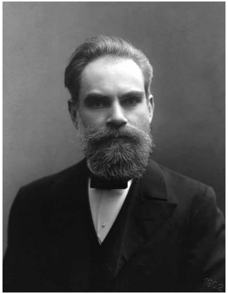
If there exists a function $V \in C^1(\mathbb{R}^n;\mathbb{R})$ *such that for all* $x \in \mathbb{R}^n$

$$V(x) > V(0), \quad \textit{and} \quad \nabla V(x) \cdot f(x) \leq 0,$$

and

$$\lim_{\|x\| \to +\infty} V(x) = +\infty,$$

then the system is stable.



A. Lyapunov (1857-1918)

Stability of Dynamical Systems

A significant advancement: Lyapunov functions

Theorem

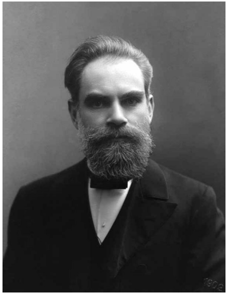
If there exists a function $V \in C^1(\mathbb{R}^n;\mathbb{R})$ *such that for all* $x \in \mathbb{R}^n$

$$V(x) > V(0), \quad \textit{and} \quad \nabla V(x) \cdot f(x) \leq 0,$$

and

$$\lim_{\|x\| \to +\infty} V(x) = +\infty,$$

then the system is stable.



A. Lyapunov (1857-1918)

Nothing tells us how to find such a function V...

Stability of Dynamical Systems

Nothing tells us how to find such a function V...

And it's not a simple problem.

Stability of Dynamical Systems

Nothing tells us how to find such a function V...

And it's not a simple problem.

$$\dot{x}(t) = \begin{pmatrix} -6x_1^4(t)x_2^5(t) - 3x_1^7(t)x_3^2(t) \\ 3x_1^9(t) - 6x_1^2(t)x_2^5(t)x_3^2(t) \\ -4x_1^2(t)x_3^5(t) \end{pmatrix}$$

Stability of Dynamical Systems

Nothing tells us how to find such a function V...

And it's not a simple problem.

$$\dot{x}(t) = \begin{pmatrix} -6x_1^4(t)x_2^5(t) - 3x_1^7(t)x_3^2(t) \\ 3x_1^9(t) - 6x_1^2(t)x_2^5(t)x_3^2(t) \\ -4x_1^2(t)x_3^5(t) \end{pmatrix}$$

The system is stable, a Lyapunov function is

$$V(x) = x_1^6 + 2(x_2^6 + x_3^4)$$

Stability of Dynamical Systems

Motivating example: boundary stabilization of the Saint-Venant equations

$$\begin{cases} \partial_t H + \partial_x(HV) = 0, \\ \partial_t V + \partial_x(\frac{V^2}{2} + gH) - S_b(x) + S(V,H,x) = 0 \\ V(t,0) = G_1(A(t,0)), \quad V(t,L) = G_2(A(t,L)) \end{cases}$$

Theorem (A.H., Shang, 2019)

There exists explicit simple conditions on G_1 and G_2 such that system is exponentially stable for the H^2 norm for any $L > 0$, S_b and S.

Remarkable:

- Explicit control which does not need the knowledge of S and S_b
- Holds for any length of the domain $\rightarrow$ something experts in the field thought impossible.

Stability of Dynamical Systems

Motivating example: boundary stabilization of the Saint-Venant equations

$$\begin{cases} \partial_t H + \partial_x(HV) = 0, \\ \partial_t V + \partial_x\left(\frac{V^2}{2} + gH\right) - S_b(x) + S(V,H,x) = 0 \\ V(t,0) = G_1(A(t,0)), \quad V(t,L) = G_2(A(t,L)) \end{cases}$$

Theorem (A.H., Shang, 2019)

There exists explicit simple conditions on G_1 and G_2 such that system is exponentially stable for the H^2 norm for any $L > 0$, S_b and S.

Remarkable:

- Explicit control which does not need the knowledge of S and S_b
- Holds for any length of the domain $\rightarrow$ something experts in the field thought impossible.

Behind: a Lyapunov function, hard to find

Stability of Dynamical Systems

Today, more than a hundred years later, it is still an open question:

there is still no systematic way to construct a Lyapunov function.

Stability of Dynamical Systems

Today, more than a hundred years later, it is still an open question:

there is still no systematic way to construct a Lyapunov function.

$\longrightarrow$ We resort to intuition

Stability of Dynamical Systems

Today, more than a hundred years later, it is still an open question:

there is still no systematic way to construct a Lyapunov function.

$\rightarrow$ We resort to intuition

Can we train an AI to have better mathematical intuition than us?

$$\dot{x}(t) = \begin{pmatrix} -6x_1^4(t)x_2^5(t) - 3x_1^7(t)x_3^2(t) \\ 3x_1^9(t) - 6x_1^2(t)x_2^5(t)x_3^2(t) \\ -4x_1^2(t)x_3^5(t) \end{pmatrix} \quad \rightarrow \quad \text{Yes, } V(x) = x_1^6 + 2(x_2^6 + x_3^4)$$

Stability of Dynamical Systems

Train an AI to have an intuition of Lyapunov functions

Global Lyapunov functions: a long-standing open problem in mathematics, with symbolic transformers (NeurIPS, Alfarano, Charton, A.H., 2024)

Neural network architecture: Transformer ($\sim$1000 smaller than GPT-3)

Procedure:

1. Generate a set of systems and associated Lyapunov functions.
2. Encode the examples
3. Train the language model (supervised learning)

Stability of Dynamical Systems

Train an AI to have an intuition of Lyapunov functions

Global Lyapunov functions: a long-standing open problem in mathematics, with symbolic transformers (NeurIPS, Alfarano, Charton, A.H., 2024)

Neural network architecture: Transformer ($\sim$1000 smaller than GPT-3)

Procedure:

1. Generate a set of systems and associated Lyapunov functions.
2. Encode the examples
3. Train the language model (supervised learning)

Stability of Dynamical Systems

How to generate examples of systems and solutions, for an open problem?

Stability of Dynamical Systems

How to generate examples of systems and solutions, for an open problem?

$\longrightarrow$ Use a backward approach: instead of finding a solution from the problem, we find the problem from the solution.

Stability of Dynamical Systems

How to generate examples of systems and solutions, for an open problem?

$\rightarrow$ Use a backward approach: instead of finding a solution from the problem, we find the problem from the solution.

- Find a mathematical way to get:

$$\underbrace{V(x)}_{\text{positive, random}} \quad \rightarrow \text{all systems with Lyapunov function } V$$

  Then sample at random.

- In spirit: finding a Lyapunov function is a hard problem, checking that a function is a Lyapunov function is easier. NP-ish flavor in some sense.

Stability of Dynamical Systems

How to generate examples of systems and solutions, for an open problem?

$\rightarrow$ Use a backward approach: instead of finding a solution from the problem, we find the problem from the solution.

- Find a mathematical way to get:

$$\underbrace{V(x)}_{\text{positive, random}} \quad \rightarrow \text{all systems with Lyapunov function } V$$

  Then sample at random.

- In spirit: finding a Lyapunov function is a hard problem, checking that a function is a Lyapunov function is easier. NP-ish flavor in some sense.

Limitations: even with a perfect generator, it biases the distribution (and it matters).

Stability of Dynamical Systems

Train an AI to have an intuition of Lyapunov functions

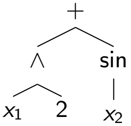
Global lyapunov functions: a long-standing open problem in mathematics, with symbolic transformers (NeurIPS, Alfarano, Charton, A.H., 2024)

Neural network architecture: Transformer ($\sim$1000 smaller than ChatGPT)

Procedure:

1. Generate a set of systems and associated Lyapunov functions.
2. Encode the examples

$(x_1^2 + \sin(x_2)) \quad \rightarrow$



$\rightarrow \quad " + "," \wedge "," x_1 "," 2 "," \sin "," x_2 "$

3. Train the language model (supervised learning)

Stability of Dynamical Systems

Train an AI to have an intuition of Lyapunov functions

Global lyapunov functions: a long-standing open problem in mathematics, with symbolic transformers (NeurIPS, Alfarano, Charton, A.H., 2024)

Neural network architecture: Transformer ($\sim$1000 smaller than ChatGPT)

Procedure:

1. Generate a set of systems and associated Lyapunov functions.
2. Encode the examples
3. Train the language model (supervised learning)
   - Symbolic training (use a cross-entropy loss)
   - Standard techniques: priming approach, repeated examples, etc.

Stability of Dynamical Systems

Results

It works! The AI learns a mathematical intuition of Lyapunov functions.

[1] Existing method for some polynomial systems.

Stability of Dynamical Systems

Results

It works! The AI learns a mathematical intuition of Lyapunov functions.

Type	n equations	SOS algorithms[1]	AI
polynomial (train distrib.)	2-5	15%	**99%**
polynomial (fwd distrib.)	2-3	47%	**84-93%**
Non-polynomial	2-3	$\sim$ 0%	**87%**

[1] Existing method for some polynomial systems.

Stability of Dynamical Systems

Results

It works! The AI learns a mathematical intuition of Lyapunov functions.

Type	n equations	SOS algorithms[1]	AI
polynomial (train distrib.)	2-5	15%	**99%**
polynomial (fwd distrib.)	2-3	47%	**84-93%**
Non-polynomial	2-3	$\sim$ 0%	**87%**

[1] Existing method for some polynomial systems.

Stability of Dynamical Systems

Results

It works! The AI learns a mathematical intuition of Lyapunov functions.

Type	n equations	SOS algorithms[1]	AI
polynomial (train distrib.)	2-5	15%	**99%**
polynomial (fwd distrib.)	2-3	47%	**84-93%**
Non-polynomial	2-3	$\sim$ 0%	**87%**

Master students accuracy: $\sim$ 10%

[1] Existing method for some polynomial systems.

Stability of Dynamical Systems

Results

It works! The AI learns a mathematical intuition of Lyapunov functions.

Type	n equations	SOS algorithms[2]	AI
polynomial (train distrib.)	2-5	15%	**99%**
polynomial (fwd distrib.)	2-3	47%	**84-93%**
Non-polynomial	2-3	$\sim$ 0%	**87%**

My accuracy (!): $\sim$ 25%

[2] Existing classical method for some polynomial systems

Stability of Dynamical Systems

Results

It works! The AI learns a mathematical intuition of Lyapunov functions.

Type	n equations	SOS algorithms[2]	AI
polynomial (train distrib.)	2-5	15%	**99%**
polynomial (fwd distrib.)	2-3	47%	**84-93%**
Non-polynomial	2-3	$\sim$ 0%	**87%**

My accuracy (!): $\sim$ 25%

Can we understand what is going on / how the model learns? $\rightarrow$ Talk of François Charton at 4:15pm

[2] Existing classical method for some polynomial systems

NewScientist

Sign in

Enter search keywords

Meta says its AI could help mathematicians

Tada Images/Shutterstock

An AI system developed by Meta can find solutions to maths problems that have eluded mathematicians for over a century, researchers at the firm claim.

The problems involve mathematical tools called Lyapunov functions, named after mathematician Aleksandr Lyapunov, which analyse whether a system will remain stable over time, meaning its behaviour can be predicted. One famous example of such a system is the motion of three celestial bodies as a result of their mutual gravitational interactions – describing the behaviour of this "three-body problem" is extremely challenging.

Read more Incredible maths proof...


AIが挑む
リアプノフ関数！
複雑系を
制御する力
2024年
10月
Global Lyapunov functions: a
problem in mathematics, with
Alberto Alfarano
FAIR, Meta
albealfa@meta.com
Amaury Hayat
CERMICS, Ecole des Ponts - Institut Polytechnique de Paris
amaury.hayat@enpc.fr
Abstract
1 Introduct
0:00 / 6:08

Summary of the approach

Paradigm: Training a Transformer to have a mathematical intuition on a problem

Key points:

- Generating data in a backward fashion (solution $\to$ problem)
- Test out-of-distribution on "real" instances of the problem.

Used in many frameworks: explicit solutions to ODE; local controllability; eigenvalues of random matrices; GCD; equilibrium of bio-networks, to predict quantities in elliptic curves, for cryptography, etc.

Try it yourself:

`https://github.com/ahayat16/Lyapunov/`

and customize on your favorite math problem

Many other examples

- in topology feedforward and MPNNs to guess links between different mathematical quantities, in particular hyperbolic and algebraic invariant of knots (a conjecture that was later proved) [Davies et al., 2021]

- in group theory path-finding for large graphs with ResMLP and RL [Chervov et al. 2025]

- in partial differential equations PINNs to find an exact self-similar solution to 3D Euler [Wang, Lai, Gómez-Serrano, Buckmaster, 2023] **(see also Victorita Dolean-Maini's talk at 3:30pm)**

ARTIFICIAL INTELLIGENCE >

Spanish mathematician Javier Gómez Serrano and Google DeepMind team up to solve the Navier-Stokes million-dollar problem

A team of researchers and engineers has been secretly working for three years on one of humanity's most devilish enigmas, the solution of which is considered imminent thanks to artificial intelligence

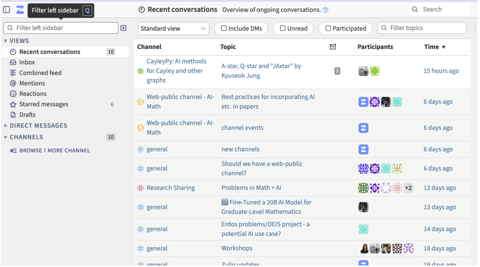
- to find good mathematical constructions PatternBoost [Charton, Ellenberg, Williamson, Wagner], AlphaEvolve [Georgiev, Gómez-Serrano, Tao, Wagner, 2025] (released two weeks ago –**see Adam Wagner's talk at 2pm**)

- ... and in many others fields

Advertising

A Zulip forum on AI for Mathematics

- by mathematicians
- for mathematicians (...and AI scientists)



https://ai-math.zulipchat.com/join/gjeretjgqhgchcjwsh2fn7g7/

with the support of ZULIP

You are most welcome to join if interested: amaury.hayat@enpc.fr

Advertising

Automath: a Paris-based seminar for the automation of mathematics (bi-monthly)

Automath !

19 octobre 2025 Recherche Séminaires

Automath ! est un projet collectif pour faire communauté en région parisienne autour de l'informatisation des mathématiques :

- usage de l'apprentissage automatique (notamment par réseaux de neurones) pour faire des mathématiques : ChatGPT, Gemini, ...
- usage de la formalisation des mathématiques et des assistants de preuve : Lean, Coq/Rocq, ...
- usage de la combinaison des deux !

Site : https://automath.dma.ens.fr/

Opening seminar in January!

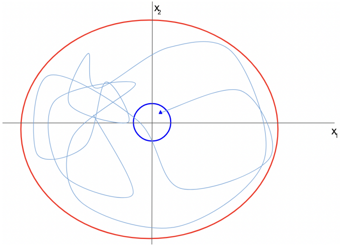
AI as a tool for math discovery

- AI are already useful in the practice of mathematics and can help solve difficult problems.
- AI are trained to have better intuition than humans on a specific problem.
- This augmented intuition allows us to bypass the difficulty of the problem.

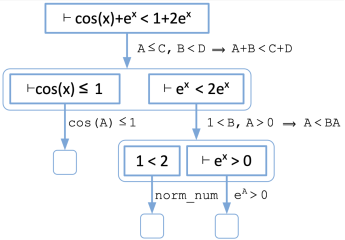
Future of Mathematical AI

Can an AI prove a mathematical result on its own?

Outline of the talk



1. AI as a tool for math discovery



2. When will AI prove theorems?

AI for Mathematical Proof

Can a trained AI find a proof for a mathematical statement?

- A much harder problem
- Shocking question: calls into question our very vision of mathematics
- A science-fiction future that is probably (much) closer than we imagine

AI for Mathematical Proof

Will Mathematics Exist in 2099? (*GAFA*, W.T. Gowers, 2000, Rough structure and classification)

Mathematician. Is the following true? Let $\delta > 0$. Then for N sufficiently large, every set $A \subset \{1, 2, \ldots, N\}$ of size at least δN contains a subset of the form $\{a, a+d, a+2d\}$?
Computer. Yes. If A is non-empty, choose $a \in A$ and set $d = 0$.
M. All right all right, but what if d is not allowed to be zero?
C. Have you tried induction on N, with some $\delta = \delta(N)$ tending to zero?
M. That idea is no help at all. Give me some examples please.
C. The obvious greedy algorithm gives the set

$$\{1, 2, 4, 5, 10, 11, 13, 14, 28, 29, 31, 32, 37, 38, 40, 41, \ldots\}.$$

C. [*Pauses for 0.001 seconds*] Actually it isn't. Behrend found a much better bound in 1946. [*Downloads paper*]
M. Oh dear, I'm out of ideas then. Could you give me a suggestion by any chance?
C. We have a set A. We want to prove that a subset of a certain form exists. The best way of proving existence is often to count.
M. [*Intrigued*] Yes, but what would that mean for a problem like this?
C. Here we wish to count the number of solutions (x, y, z) of the single

AI for Mathematical Proof

Will Mathematics Exist in 2099? (*GAFA*, W.T. Gowers, 2000, Rough structure and classification)

84 W.T. GOWERS GAFA2000

techniques may be helpful. Rather than giving several examples of the use of standard methods to solve problems, let me return to the question of automating mathematics and present an imagined dialogue between a mathematician and a computer in two or three decades' time. The idea

AI for Mathematical Proof

First approach: training a Transformer (GPT-f, Polu, Sutskever, 2020)

Question

Let $a > 0$ *and* $b > 0$*, such that* $ab = b - a$*, show that*

$$\frac{a}{b} + \frac{b}{a} - ab = 2$$

GPT

Proof

AI for Mathematical Proof

First approach: training a Transformer (GPT-f, Polu, Sutskever, 2020)

Question

Let $a > 0$ *and* $b > 0$*, such that* $ab = b - a$*, show that*

$$\frac{a}{b} + \frac{b}{a} - ab = 2$$

GPT

Proof

Ilya Sutskever: The OpenAI Genius Who Told Sam Altman He Was Fired

Company's chief scientist led a board coup against one of the most prominent figures in Silicon Valley

Ilya Sutskever

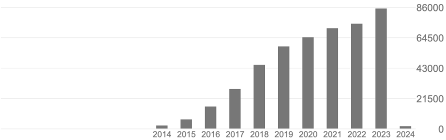
SUIVRE

Co-Founder and Chief Scientist of OpenAI
Adresse e-mail validée de openai.com - Page d'accueil

Machine Learning Neural Networks Artificial Intelligence Deep Learning

ARTICLES CITÉE PAR ACCÈS PUBLIC COAUTEURS

	Toutes	Depuis 2019
Citations	462122	356473
indice h	84	79
indice i10	114	110



AI for Mathematical Proof

First approach: training a Transformer (GPT-f, Polu, Sutskever, 2020)

Question

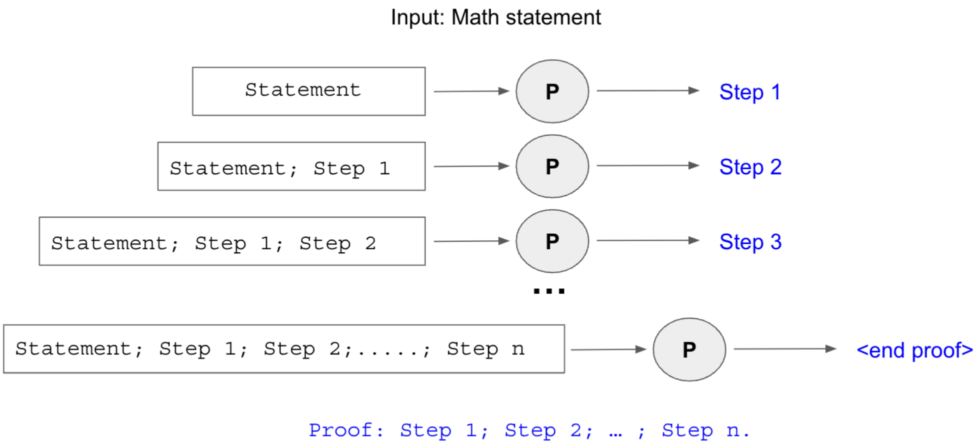
Let $a > 0$ and $b > 0$, such that $ab = b - a$, show that

$$\frac{a}{b} + \frac{b}{a} - ab = 2$$

GPT

Proof

AI for Mathematical Proof



AI for Mathematical Proof

```
theorem Exercice_1
(a b : ℝ)
(h₀: a > 0)
(h₁: b > 0)
(h₂: a*b = b-a) :
a/b+b/a-2*(a*b) = 2 :=
begin
sorry,
end
```

Procedure: train it with examples: (exercises, proofs)

- The hope is that by showing it enough examples, the AI will be capable of learning to reason, just by learning to predict the next step each time.

AI for Mathematical Proof

```
theorem Exercice_1
(a b : ℝ)
(h₀: a > 0)
(h₁: b > 0)
(h₂: a*b = b-a) :
a/b+b/a-2*(a*b) = 2 :=
begin
sorry,
end
```

Proof

Procedure: train it with examples: (exercises, proofs)

- The hope is that by showing it enough examples, the AI will be capable of learning to reason, just by learning to predict the next step each time.

AI for Mathematical Proof

```
theorem Exercice_1
(a b : ℝ)
(h₀: a > 0)
(h₁: b > 0)
(h₂: a*b = b-a) :
a/b+b/a-2*(a*b) = 2 :=
begin
sorry,
end
```

Procedure: train it with examples: (exercises, proofs)

- The hope is that by showing it enough examples, the AI will be capable of learning to reason, just by learning to predict the next step each time.

Enough = sufficiently diverse and sufficiently numerous

→ Limitation: lack of data

AI for Mathematical Proof

We have very few data available (especially formal).

AI for Mathematical Proof

We have very few data available (especially formal).

Question

Let $a > 0$ *and* $b > 0$*, such that* $ab = b - a$*, show that*

$$\frac{a}{b} + \frac{b}{a} - ab = 2$$

informal language

```
theorem Exercice_1
(a b : ℝ)
(h₀: a > 0)
(h₁: b > 0)
(h₂: a*b = b-a) :
a/b+b/a-2*(a*b) = 2 :=
begin
sorry,
end
```

formal language

AI for Mathematical Proof

We have very few data available (especially formal).

Question

Let $a > 0$ and $b > 0$, such that $ab = b - a$, show that

$$\frac{a}{b} + \frac{b}{a} - ab = 2$$

informal language

```
theorem Exercice_1
(a b : ℝ)
(h₀: a > 0)
(h₁: b > 0)
(h₂: a*b = b-a) :
a/b+b/a-2*(a*b) = 2 :=
begin
sorry,
end
```

formal language

→ See Patrick Massot's talk at 11am

AI for Mathematical Proof

We have very few data available (especially formal).

Question

Let $a > 0$ and $b > 0$, such that $ab = b - a$, show that

$$\frac{a}{b} + \frac{b}{a} - ab = 2$$

informal language

```
theorem Exercice_1
(a b : ℝ)
(h₀: a > 0)
(h₁: b > 0)
(h₂: a*b = b-a) :
a/b+b/a-2*(a*b) = 2 :=
begin
sorry,
end
```

formal language

→ See Patrick Massot's talk at 11am

Lean: ∼300,000 theorems. A large dataset for humans, a small dataset for AI.
→ Limit of the approach

AI for Mathematical Proof

How to tackle this limit ?

Many subsequent improvements and variations

- Training a retriever on the library to suggest relevant theorems: **LeanDojo** (Yang et al., 2023)
- Fine-tuning better base LLM: **LeanLlama** (Glöckle et al., 2023)

Tackle the lack of data

- Using additional data gathered on the internet: **Llemma** (Azerbayev et al., 2023), **InternLM-2** (Wu et al. 2024)
- Using synthetic data or automatic formalization: **HyperTree Proof Search** (Lample et al. 2022), **DeepSeek-Prover** (Xin et al. 2024), **Kimina-Prover** (Wang et al. 2025) **REAL-Prover** (Shen et al. 2025), **Goedel-Prover-v2** (Lin et al. 2025)

AI for Mathematical Proof

Autoformalization: a natural idea to get more formal data → train a model to translate from "natural language proof" to formal and verifiable data.

Exercice 1

Let $a > 0$ and $b > 0$, such that $ab = b - a$, show that

$$\frac{a}{b} + \frac{b}{a} - ab = 2$$

informal language

→

```
theorem Exercice_1
(a b : ℝ)
(h₀: a > 0)
(h₁: b > 0)
(h₂: a*b = b-a) :
a/b+b/a-2*(a*b) = 2 :=
begin
sorry,
end
```

formal language

Autoformalization: both a means and an end:

- a means: more data to train AI models
- an end: checking the correctness of new mathematical theories

AI for Mathematical Proof

Autoformalization: both a means and an end

Example of contradictory papers:

- Schumacher, G., Tsuji, H. (2004). Quasi-projectivity of moduli spaces of polarized varieties. *Annals of mathematics*, 597-639.
- Kollár, J. (2006). Non-quasi-projective moduli spaces. *Annals of mathematics*, 1077-1096.

A quickly improving field:

- (2024) Very good statement autoformalization for olympiad style exercise (e.g. Herald, Numina, AlphaProof)
- (2025) 4% of statements in arxiv papers autoformalized automatically in most field of mathematics
- (2025) two first arxiv paper completely autoformalized by Morph Labs / Math inc (proofs included) with little human intervention
- (2025) Start of MALINCA - an ERC Synergy Grant project

Still challenges for statement and proof autoformalization in most fields of mathematics

AI for Mathematical Proof

How to tackle this limit ?

Many subsequent improvements and variations

- Training a retriever on the library to suggest relevant theorems: **LeanDojo** (Yang et al., 2023)
- Fine-tuning better base LLM: **LeanLlama** (Glöckle et al., 2023)

Tackle the lack of data

- Using additional data gathered on the internet: **Llemma** (Azerbayev et al., 2023), **InternLM-2** (Wu et al. 2024)
- Using synthetic data or automatic formalization: **HyperTree Proof Search** (Lample et al. 2022), **DeepSeek-Prover** (Xin et al. 2024), **Kimina-Prover** (Wang et al. 2025) **REAL-Prover** (Shen et al. 2025), **Goedel-Prover-v2** (Lin et al. 2025)

AI for Mathematical Proof

How to tackle this limit ?

Many subsequent improvements and variations

- Training a retriever on the library to suggest relevant theorems: **LeanDojo** (Yang et al., 2023)
- Fine-tuning better base LLM: **LeanLlama** (Glöckle et al., 2023)

Tackle the lack of data

- Using additional data gathered on the internet: **Llemma** (Azerbayev et al., 2023), **InternLM-2** (Wu et al. 2024)
- Using synthetic data or automatic formalization: **HyperTree Proof Search** (Lample et al. 2022), **DeepSeek-Prover** (Xin et al. 2024), **Kimina Prover** (Wang et al. 2025) **REAL-Prover** (Shen et al. 2025), **Goedel-Prover-v2** (Lin et al. 2025)

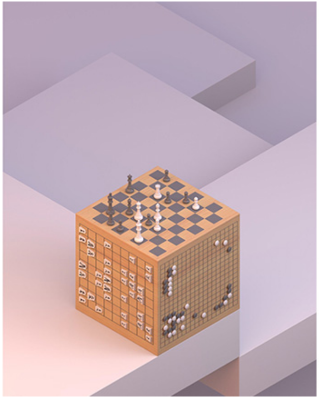
How to go further?

AI for Mathematical Proof

Second approach: **treat mathematics as a game** (Lample, Lachaux, Lavril, Martinet, Hayat, Ebner, Rodriguez, Lacroix, 2022);

AI for Mathematical Proof

Second approach: **treat mathematics as a game** (Lample, Lachaux, Lavril, Martinet, Hayat, Ebner, Rodriguez, Lacroix, 2022);

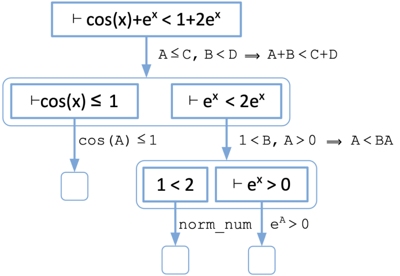


Deepmind (2017)

AI for Mathematical Proof

Second approach: **treat mathematics as a game** (Lample, Lachaux, Lavril, Martinet, Hayat, Ebner, Rodriguez, Lacroix, 2022);

Deepmind (2017)

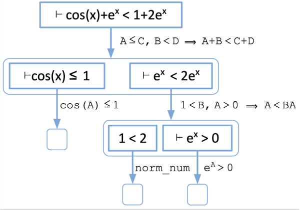


You won !

AI for Mathematical Proof

Main difficulties:

- two-player game vs. solo against a goal.
- In chess, when you play a move you always have a single game. In mathematics: one statement $\rightarrow$ multiple statements
- Difficult in mathematics to know automatically in the middle of a proof what the probability of succeeding is.
- The number of possibilities is much, much larger in mathematics

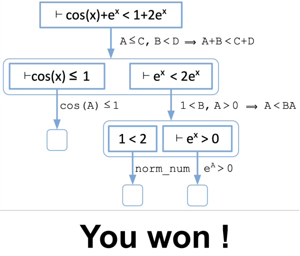


You won !

AI for Mathematical Proof

Main difficulties:

- two-player game vs. solo against a goal.
- In chess, when you play a move you always have a single game. In mathematics: one statement $\rightarrow$ multiple statements
- Difficult in mathematics to know automatically in the middle of a proof what the probability of succeeding is.
- The number of possibilities is much, much larger in mathematics



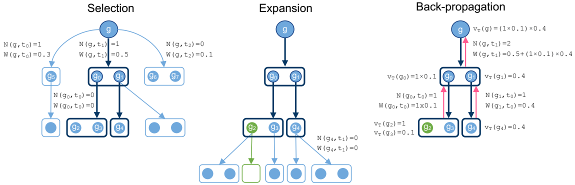
Much more difficult than chess

AI for Mathematical Proof

In practice

- Two transformers: P_θ which predicts a tactic, c_θ which predicts the difficulty of proving a statement (goal, hypothesis, etc.).

- An intelligent proof search that sees the proof as a tree and combines P_θ, c_θ and a tree



expansion.

- Continuously training of P_θ and c_θ on successful proofs

AI for Mathematical Proof

Results

Exercises at the undergraduate level...

...60% of middle school / high school exercises up to Olympiad level...

...and a few exercises from the International Mathematical Olympiads

AI for Mathematical Proof

Results

Exercises at the undergraduate level...

...60% of middle school / high school exercises up to Olympiad level...

...and a few exercises from the International Mathematical Olympiads

Exercise

Show that for all $n \in \mathbb{N}$, *7 does not divide* $2^n + 1$.

AI for Mathematical Proof

Results

Exercises at the undergraduate level...

...60% of middle school / high school exercises up to Olympiad level... More recent approach reached $90-99\%$ of success (Deepseek, **Kimina**, Seed, 2025)
$\rightarrow$ Talk of Yann Fleureau at 11:30am

...and a few exercises from the International Mathematical Olympiads

Exercise

Show that for all $n \in \mathbb{N}$, 7 does not divide $2^n + 1$.

More recent approaches reached up to a gold medal level at the International Mathematical Olympiads (AlphaProof, 2024, Kimina, Aristotle, Seed, 2025)

AI for Mathematical Proof

ABEL: Hypertree proof search 2.0 (Glöckle, Limperg, Synnaeve, A.H. 2024)

ABEL:Sample Efficient Online Reinforcement Learning for Neural Theorem Proving

A better AlphaZero style proof-search in the era of pre-trained models:

- on-par with state of the art on high-school exercise (in Oct. 2024)
- state-of-the-art on PutnamBench for 3,5 months
- Very very few additional data used in the training ($\sim$240 examples)

AI for Mathematical Proof

ABEL: Hypertree proof search 2.0 (Glöckle, Limperg, Synnaeve, A.H. 2024)

ABEL:Sample Efficient Online Reinforcement Learning for Neural Theorem Proving

A better AlphaZero style proof-search in the era of pre-trained models:

- on-par with state of the art on high-school exercise (in Oct. 2024)
- state-of-the-art on PutnamBench for 3,5 months
- Very very few additional data used in the training ($\sim$240 examples)

$\rightarrow$ Progress are orthogonal to recent progress made with autoformalized or synthetic data

AI for Mathematical Proof

ABEL: Hypertree proof search 2.0 (Glöckle, Limperg, Synnaeve, A.H. 2024)

ABEL:Sample Efficient Online Reinforcement Learning for Neural Theorem Proving

A better AlphaZero style proof-search in the era of pre-trained models:

- on-par with state of the art on high-school exercise (in Oct. 2024)
- state-of-the-art on PutnamBench for 3,5 months
- Very very few additional data used in the training ($\sim$240 examples)

$\rightarrow$ Progress are orthogonal to recent progress made with autoformalized or synthetic data

Should be open-sourced soon!

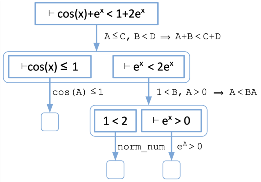
AI for Mathematical Proof

Two types of reinforcement learning approaches:

- Proof search, step by step (HTPS, ABEL, etc.)

- Whole-proof generation: a new paradigm since 2024 (Deepseek-prover, Kimina-prover, etc.)

  Principle: train a model with a specific type of reinforcement learning with **verifiable answers** $\rightarrow$ the underlying RL (GRPO) makes it a "cheaper" method



You won !

AI for Mathematical Proof

Two types of reinforcement learning approaches:

- Proof search, step by step (HTPS, ABEL, etc.)
- Whole-proof generation: a new paradigm since 2024 (Deepseek-prover, Kimina-prover, etc.)

  Principle: train a model with a specific type of reinforcement learning with **verifiable answers** → the underlying RL (GRPO) makes it a “cheaper” method

  → This is at the origin of the so-called reasoning models (GPT-5, o3, deepseek-R1, Magistral, etc.)

AI for Mathematical Proof

Two types of reinforcement learning approaches:

- Proof search, step by step (HTPS, ABEL, etc.)
- Whole-proof generation: a new paradigm since 2024 (Deepseek-prover, Kimina-prover, etc.)

  Principle: train a model with a specific type of reinforcement learning with **verifiable answers** → the underlying RL (GRPO) makes it a "cheaper" method

  → This is at the origin of the so-called reasoning models (GPT-5, o3, deepseek-R1, Magistral, etc.)

AI for Mathematical Proof

Two types of reinforcement learning approaches:

- Proof search, step by step (HTPS, ABEL, etc.)
- Whole-proof generation: a new paradigm since 2024 (Deepseek-prover, Kimina-prover, etc.)

  Principle: train a model with a specific type of reinforcement learning with **verifiable answers** $\rightarrow$ the underlying RL (GRPO) makes it a "cheaper" method

  $\rightarrow$ This is at the origin of the so-called reasoning models (GPT-5, o3, deepseek-R1, Magistral, etc.)

Main limit of reinforcement learning: collapse of diversity

AI for Mathematical Proof

Two types of reinforcement learning approaches:

- Proof search, step by step (HTPS, ABEL, etc.)
- Whole-proof generation: a new paradigm since 2024 (Deepseek-prover, Kimina-prover, etc.)

  Principle: train a model with a specific type of reinforcement learning with **verifiable answers** → the underlying RL (GRPO) makes it a "cheaper" method

  → This is at the origin of the so-called reasoning models (GPT-5, o3, deepseek-R1, Magistral, etc.)

Main limit of reinforcement learning: collapse of diversity
→ How to tackle this ?

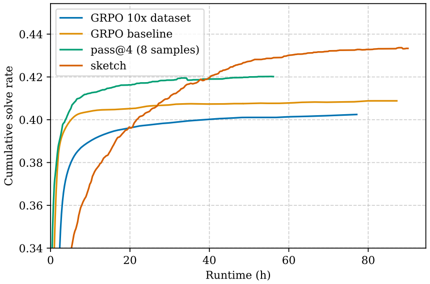
AI for Mathematical Proof

End-to-end sketching and proving (Glöckle, Gu, Synnaeve, A.H., 2025)

Instead of generating whole-proofs, train together:

- A sketcher is trained to provide skeleton of proofs with many lemmas
- A prover is trained to prove the lemmas

Rationale: decomposition and hierarchical recursion is a natural way to decrease the complexity when several attempts are possible.



AI for Mathematical Proof

A very (very) fast-moving field

2019

$(r_1 - r_2)r_3 = r_1 r_3 - r_2 r_3$

HOList, LPLG

2022

$\forall\ n \in \mathbb{N},\ \neg 7 \mid 2^n + 1$

GPT-f, Thor/DSP, HTPS

2024

Silver medal Inter. Math. Olymp.

Llemma, LeanDojo, InternML, DeepSeek, ABEL, AlphaProof, etc. 2025 wouldn't fit on the slide

Today:

- Several models obtain a gold medal at the Inter. Math. Olymp.
- AI models included in Lean tactics (e.g. Lean hammer)
- Models starts to have close to human performances on some problems

Conclusion

- AI methods are already useful in the practice of mathematics
- AI and LLM for proving theorems is only beginning, and there are many ideas... and much to do.
- LLMs will not be the final AI tool which will be used to mathematics.
- The practice of mathematics will probably change... and that's okay.
- AI will not replace mathematicians but will instead enhance them.

A great thank to my co-authors

Jean-Michel Coron, Peipei Shang, Georges Bastin, Shengquan Xiang, Christophe Zhang, Benedetto Piccoli, Thibault Liard, Alexandre Bayen, Ludovick Gagnon, Swann Marx, Maria Laura Delle Monache, Benjamin Seibold, Dan Work, Jonathan Sprinkle, Jonathan Lee, Arwa Alanqary, Rahul Bhadani, Christopher Denaro, Ryan J. Weightman, Matthew Bunting, Anish Gollakota, Matthew W. Nice, Derek Gloudemans, Gergely Zachar, Jon F. Davis, Virginie Ehrlacher, Alberto Alfarano, Guillaume Lample, Gabriel Synnaeve, Timothée Lacroix, Jean Cauvin-Vila, Yating Hu, Epiphane Loko, Nathan Lichtlé, Fabian Glöckle François Charton, Vincent Boulard
...and many others

A great thank to my co-authors


Alberto Alfarano


François Charton


Fabian Glöckle


Guillaume Lample


Antoine Peyronnet


Gabriel Synnaeve

Conclusion

Thank you for your attention