

AI-Assisted Scientific Research

Yan-Qing Ma (Peking University)

yqma@pku.edu.cn

Workshop on Quarkonium Physics in High Energy Nuclear Collisions
CCNU, Wuhan, 2026/10/10



北京大学



Outline

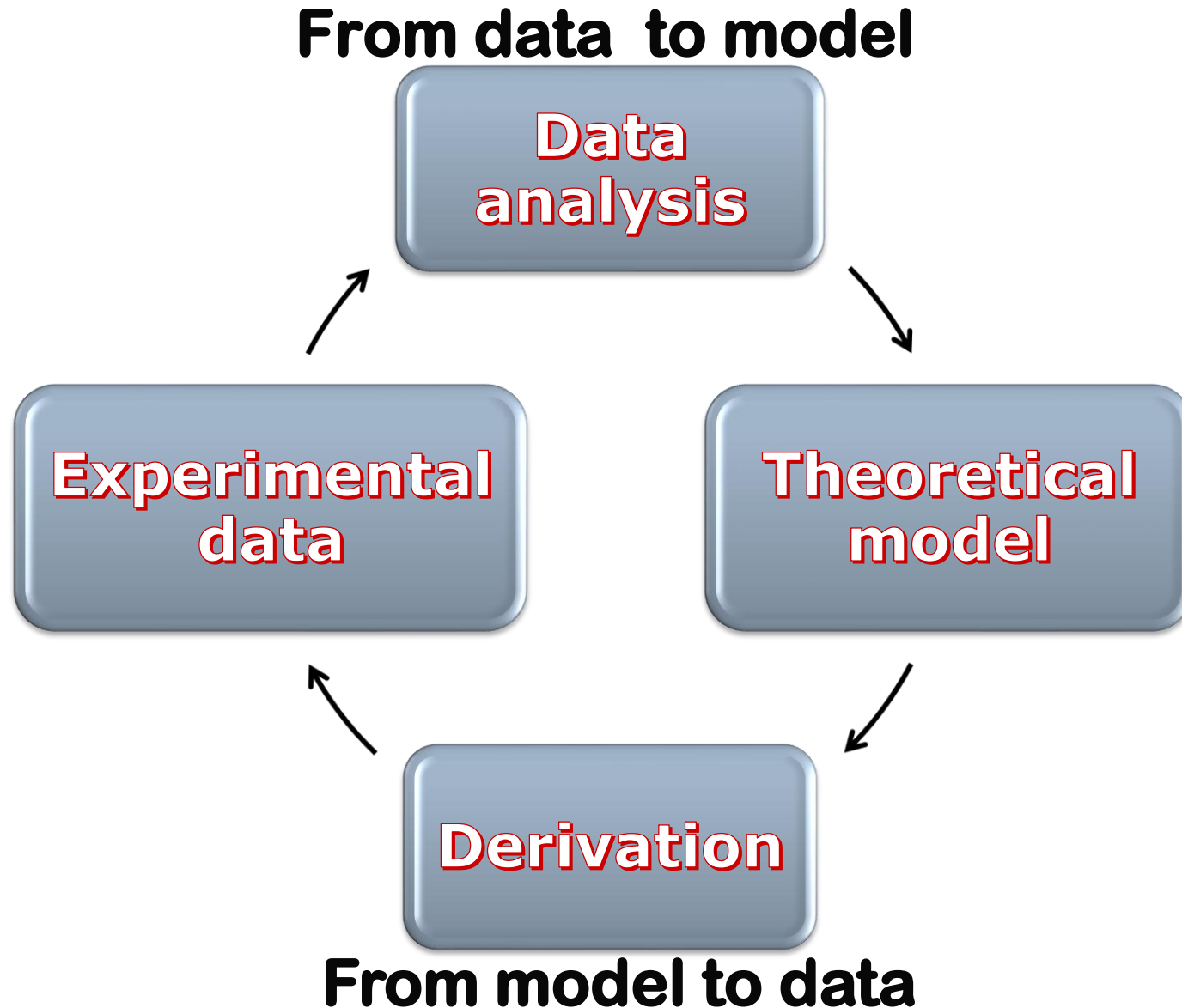
I. Background

II. Progress in Mathematics

III. Progress in Physics

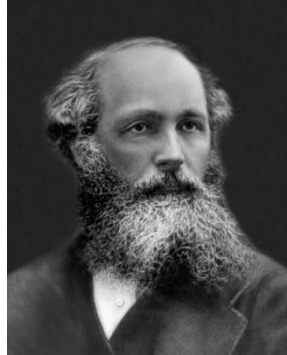
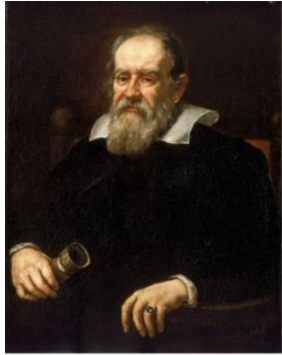
IV. AI-era research

The circle of scientific discovery



Human scientific discovery

Fundamental natural laws: human contributions



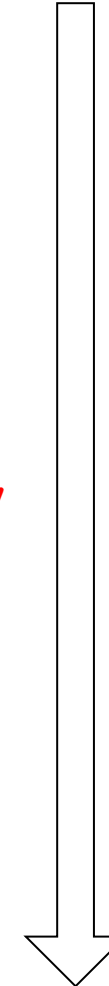
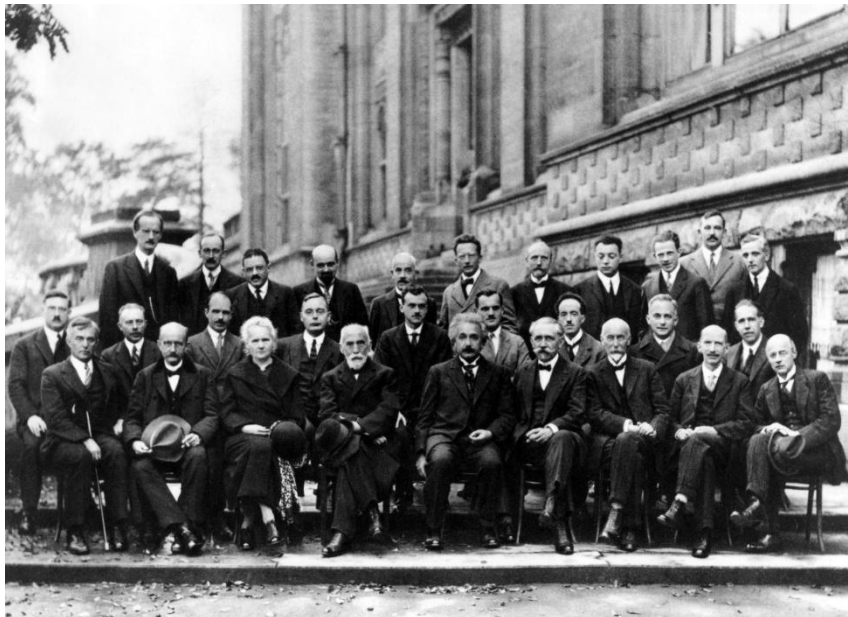
Galileo's
laws of motion

Newton's
laws of motion

Maxwell's
electromagnetic theory

Theory of relativity

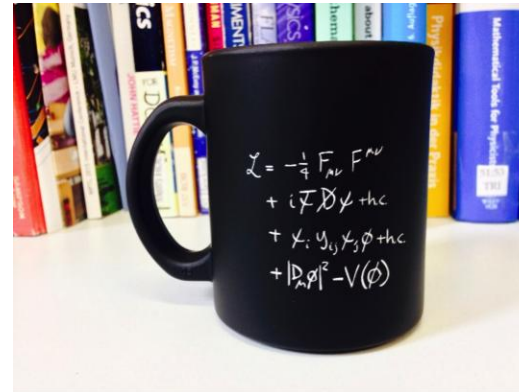
Quantum theory



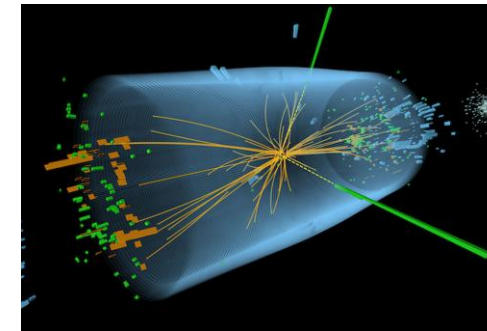
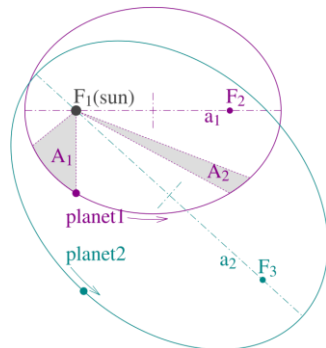
Reflection

Human exploration of natural laws:

- **Advantages:** interpretability, conciseness, **universality**

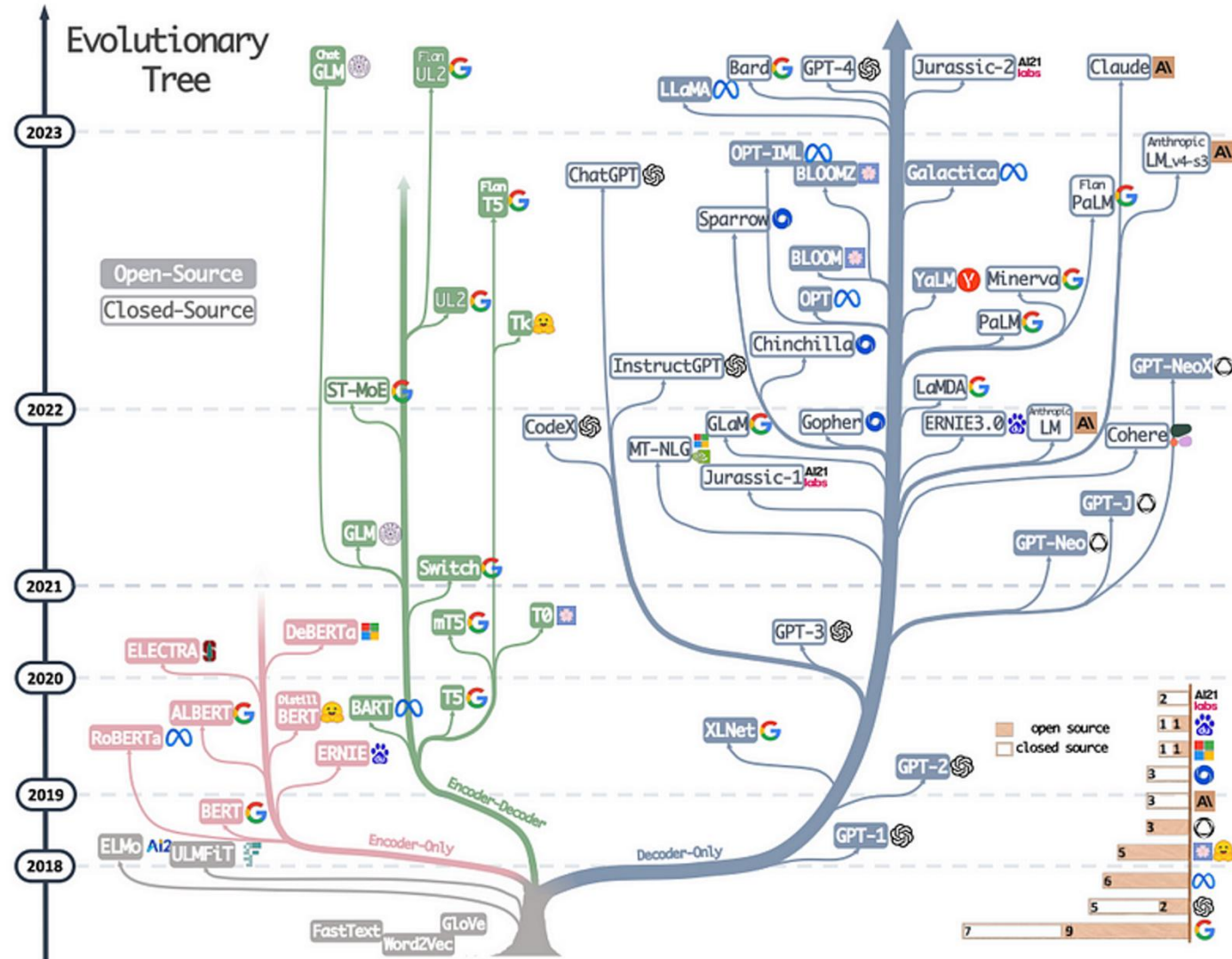


- **Disadvantages:** long period, preconceived notion, insufficient ability to handle complex problems

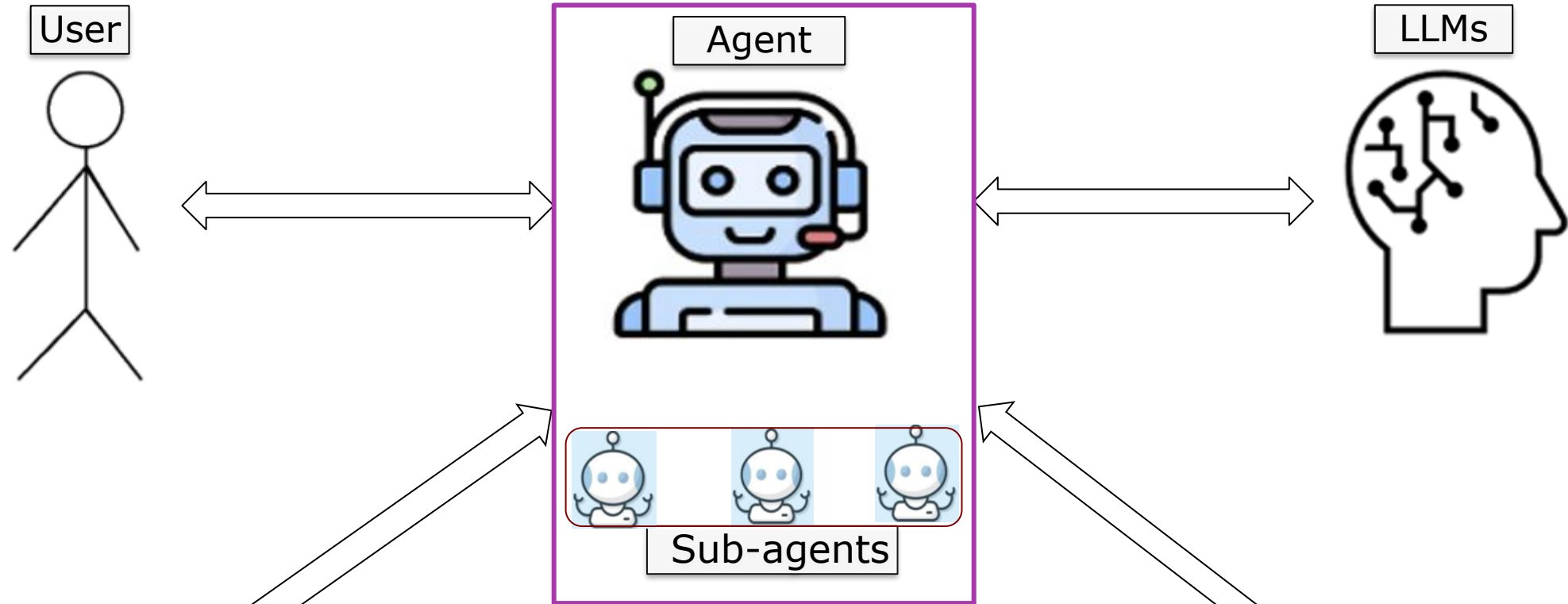


Advancement in LLMs

LLMs: rapid advancement in recent years



AI Toolchain Diagram



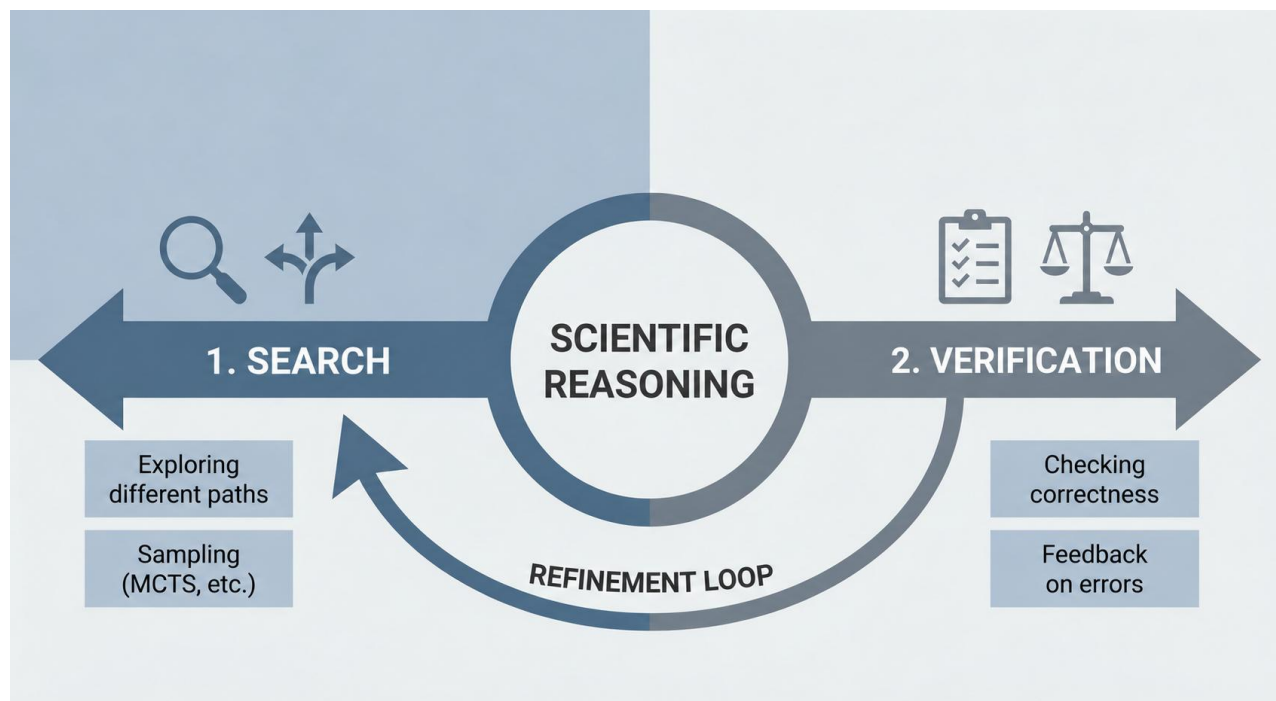
Skills: specific knowledge, specialized functions



Tools: mount existing tools

Test-driving research paradigm

- **Scientific tasks = search + verification**



⑩ Search: Explore useful candidates.

□ Verification: Check, learn and refine.

Search = model capabilities + agent harness

Verification = independent checks + a reliable verifier

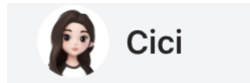
Creative methods can emerge inside the loop

AI: Conquering One Field after Another

The beginning
of the end:

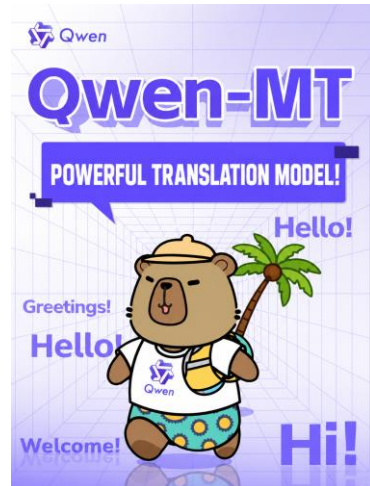
2022

- Chat client



2023

- Translation



2024

- Coding



2025

- Mathematics



Outline

I. Background

II. Progress in Mathematics

III. Progress in Physics

IV. AI-era research

AI reaches IMO gold-medal standard

Google DeepMind · July 2025

Can AI write proofs at IMO gold-medal level?

- Gemini Deep Think solved five of six problems, earning 35 out of 42 points.
- IMO coordinators certified the score. The model wrote proofs in natural language.

35 / 42



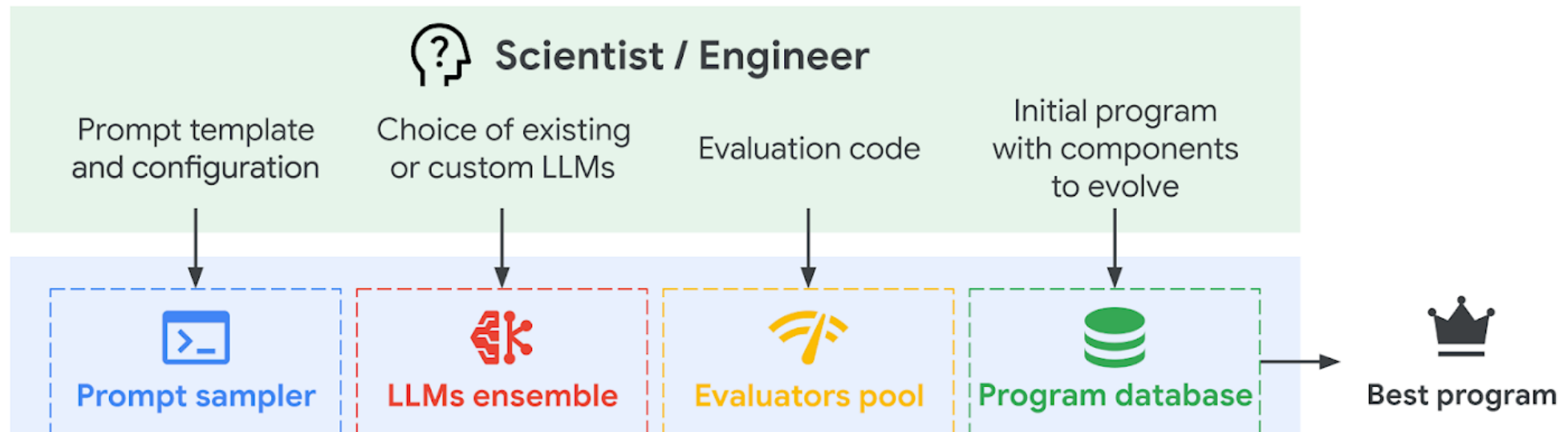
2024 and 2025 systems (Google DeepMind)

AlphaEvolve and mathematical discovery

Google DeepMind and mathematical collaborators · May–November 2025

Evolutionary search for mathematical constructions

- May: a new 11-dimensional kissing configuration with 593 outer spheres.
- November: 67 research-level problems studied, with several improved constructions.



AlphaEvolve components (original diagram, excerpt)

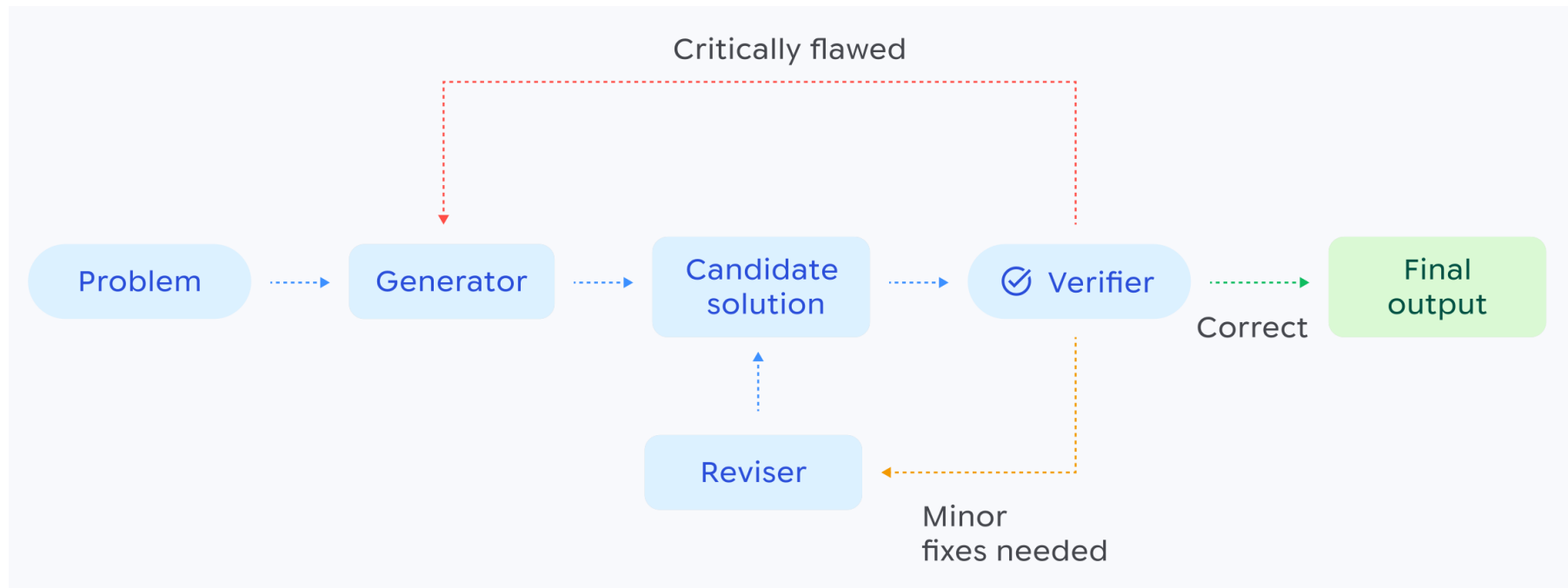
[May announcement](#) [November mathematics paper](#)

Aletheia and the Erdős problems

Google DeepMind · February 2026

Aletheia's research and verification loop

- 700 Erdős problems assessed, with resolutions of 652, 654, 1040 and 1051.



Aletheia workflow (Google DeepMind)

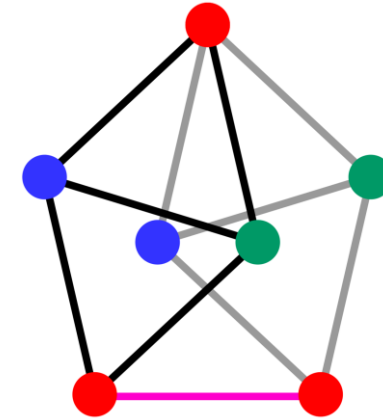
[DeepMind announcement](#) [Erdős case study](#) [Collaboration case studies](#)

A counterexample to unit-distance growth

OpenAI · May 2026

How many pairs at unit distance can a planar point set contain?

- Planar point sets have n^{1+c} unit-distance pairs for a fixed $c > 0$ and arbitrarily large n .
- Golod–Shafarevich field towers refute the expected near-linear upper bound.



7 points; 11 pairs at distance 1 (classical example)

AI finds unexpected constructions that mathematicians can verify

The Jacobian conjecture counterexample

Levent Alpöge with Claude · July 2026

Does local polynomial invertibility imply global invertibility?

- An explicit polynomial map on $\mathbb{C}^3$ has constant nonzero Jacobian, yet distinct inputs collide.
- The counterexample extends to dimensions ≥ 3 . The two-dimensional question remains open.

$$\mathbf{F}(\mathbf{a}) = \mathbf{F}(\mathbf{b}), \mathbf{a} \neq \mathbf{b}$$

$$\det J_F = -2.$$

$$F(x, y, z) = \begin{pmatrix} u^3 z + y^2 u(4 + 3xy) \\ y + 3xu^2 z + 3xy^2(4 + 3xy) \\ 2x - 3x^2 y - x^3 z \end{pmatrix}$$

$$\left(0, 0, -\frac{1}{4}\right), \quad \left(1, -\frac{3}{2}, \frac{13}{2}\right), \quad \left(-1, \frac{3}{2}, \frac{13}{2}\right)$$



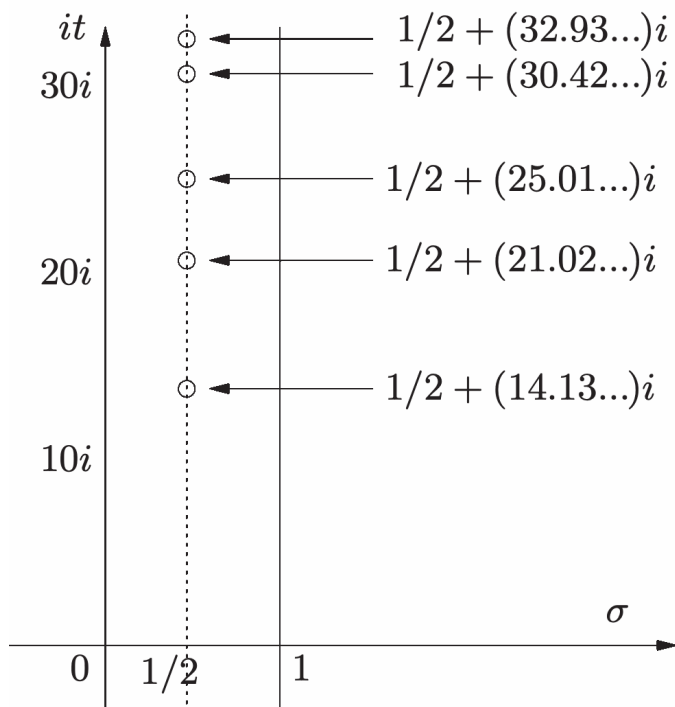
$$F = \left(-\frac{1}{4}, 0, 0\right)$$

3D complex counterexample; 2D open

Riemann hypothesis: a new lower bound

Anthropic · 10 August 2026

Do all nontrivial zeros of zeta function lie on $\text{Re}(s) = 1/2$?



Known zeros:

$\text{Re}(s) = 1/2$

The Riemann hypothesis remains open

Critical-line proportion: lower bound

Claude · Aug 2026



67.2%

Previous bound

41.6%

0.0% 20.0% 40.0% 60.0% 80.0% 100.0%

~60 coordinated agents

Numerical checks and mutual review

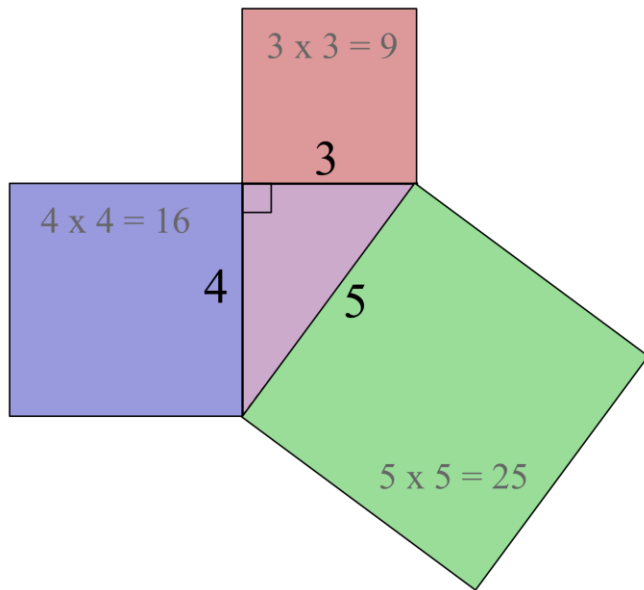
Expert review and Lean followed

Fermat's Last Theorem in Lean

Anthropic · 4 September 2026

Can $a^n + b^n = c^n$ hold for positive integers when $n > 2$?

$$3^2 + 4^2 = 5^2$$



$$a^n + b^n \neq c^n$$

Positive integers a, b, c ; integer $n > 2$

11 days

Largely autonomous formalization

**Lean checks each step
of the complete proof**

An existing human proof, now checked by computer at scale.

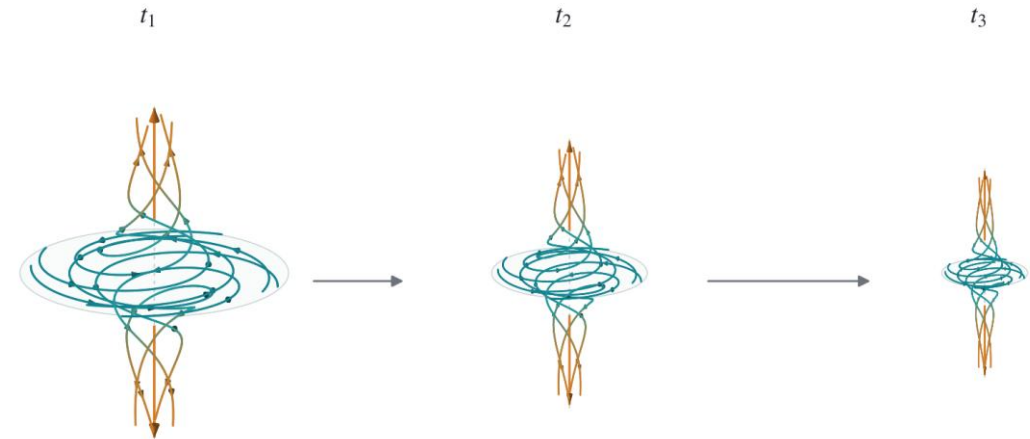
Navier–Stokes finite-time blow-up

OpenAI · 8 September 2026

Can a smooth, finite-energy fluid develop a singularity?

$$\begin{aligned}\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u} &= -\nabla p + \nu \Delta \mathbf{u} + \mathbf{f} \\ \nabla \cdot \mathbf{u} &= 0\end{aligned}$$

- The release reports finite-time blow-up under smooth forcing, with finite energy.
- About 10,000 agents worked for 88 hours. Lean formalization followed in 17 hours.



One of the Seven Millennium Prize Problems, Clay Mathematics Institute

Smooth forcing, finite energy; reported proof and Lean formalization.

[Research announcement](#) [Lean proof repository](#)

OpenAI's mathematics release

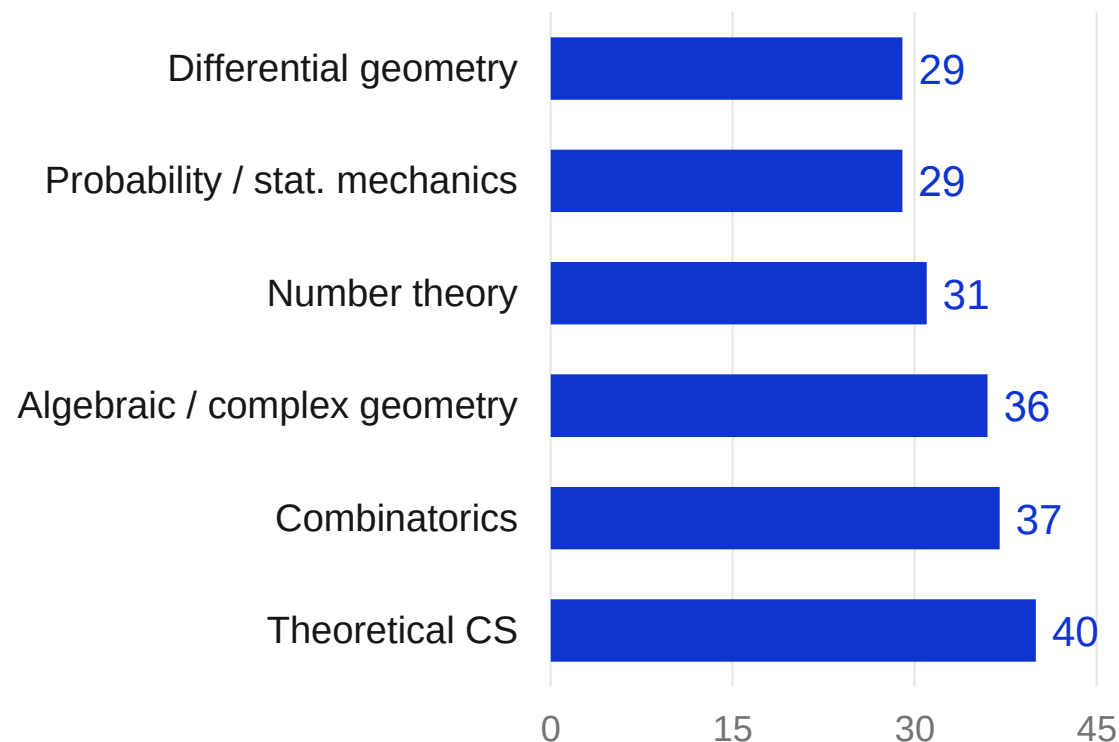
Released 6 October 2026 · Catalogue snapshot: 9 October 2026

700+ manuscripts across 17 mathematical fields

- 719 manuscripts belong to 372 families of related results.
- Each family may include a main result, companion papers or alternative proofs.

719 manuscripts

Families in the six largest fields



Counts from the catalogue; the other 11 fields contain 170 families.

Proof verification and open review

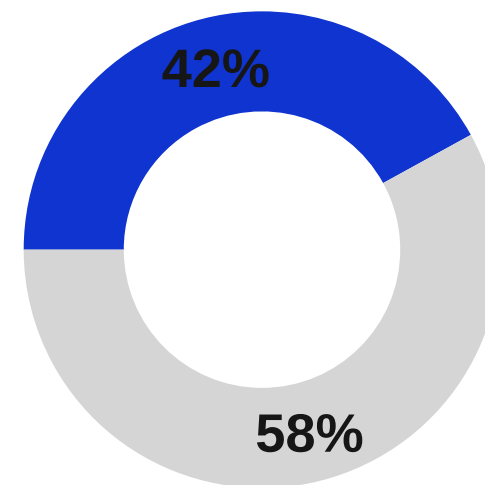
OpenAI repository · Status snapshot: 9 October 2026

Lean certificates and a versioned public record

- The repository reports about 42% formalization, with Lean files available for checking.
- The release includes 10 reasoning summaries. New versions retain earlier manuscripts.

≈42% formalized

Repository-reported formalization status



■ Lean formalized ■ Other stages

Approximate formalization share reported by the repository.

A lower bound for the plane-coloring problem

OpenAI mathematics release · 6 October 2026

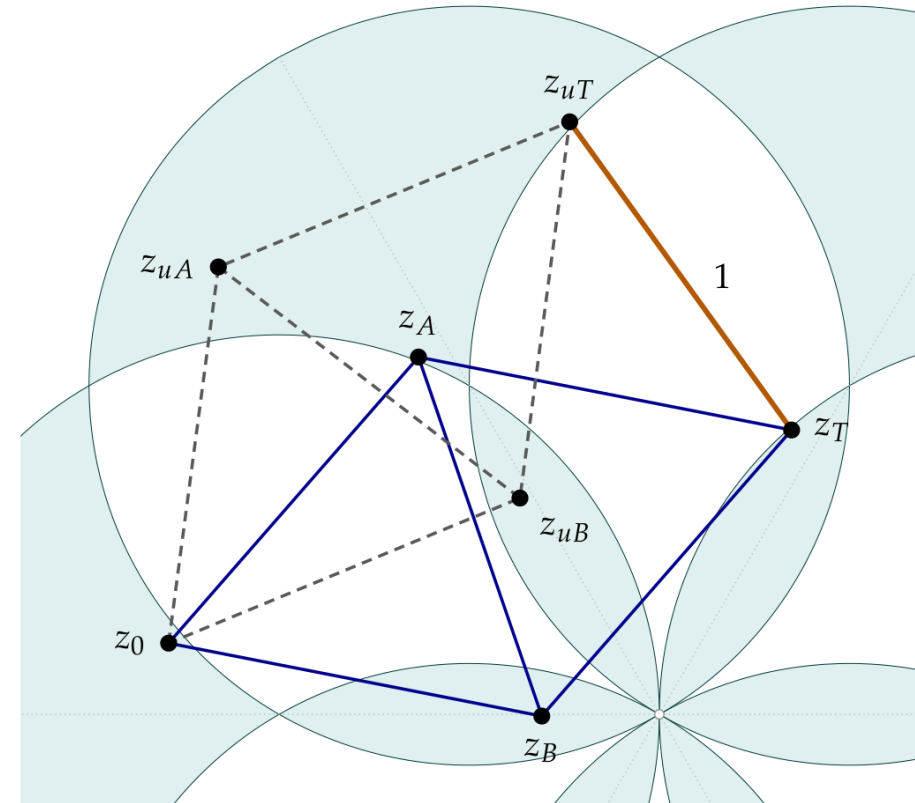
How many colors avoid monochromatic unit-distance pairs?

- The October catalogue reports that the Euclidean plane is not five-colorable.
- The classical seven-color upper bound remains. The exact value is six or seven.

$$6 \leq \chi(\mathbb{R}^2) \leq 7$$

Reported release claim, verification status varies.

[Result catalogue](#) [Associated manuscript](#)



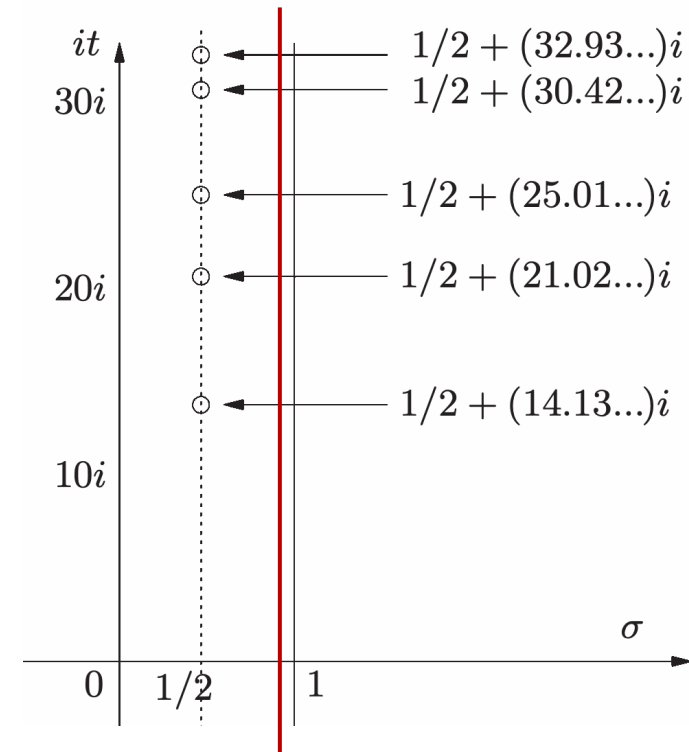
Moser-graph component (paper Fig. 3)

A fixed zero-free strip for $\zeta(s)$

OpenAI mathematics release · 6 October 2026

How far left can a zero-free region extend uniformly?

- The October collection reports a fixed zero-free region right of $\operatorname{Re}(s) = 11/12$.
- Verification stages vary. This claim does not establish the Riemann hypothesis.



$$\zeta(s) \neq 0$$

$\operatorname{Re}(s) > 11/12$, excluding the pole at $s = 1$.

[Repository scope and status](#) [Release announcement](#)



Agent

LLM

AI

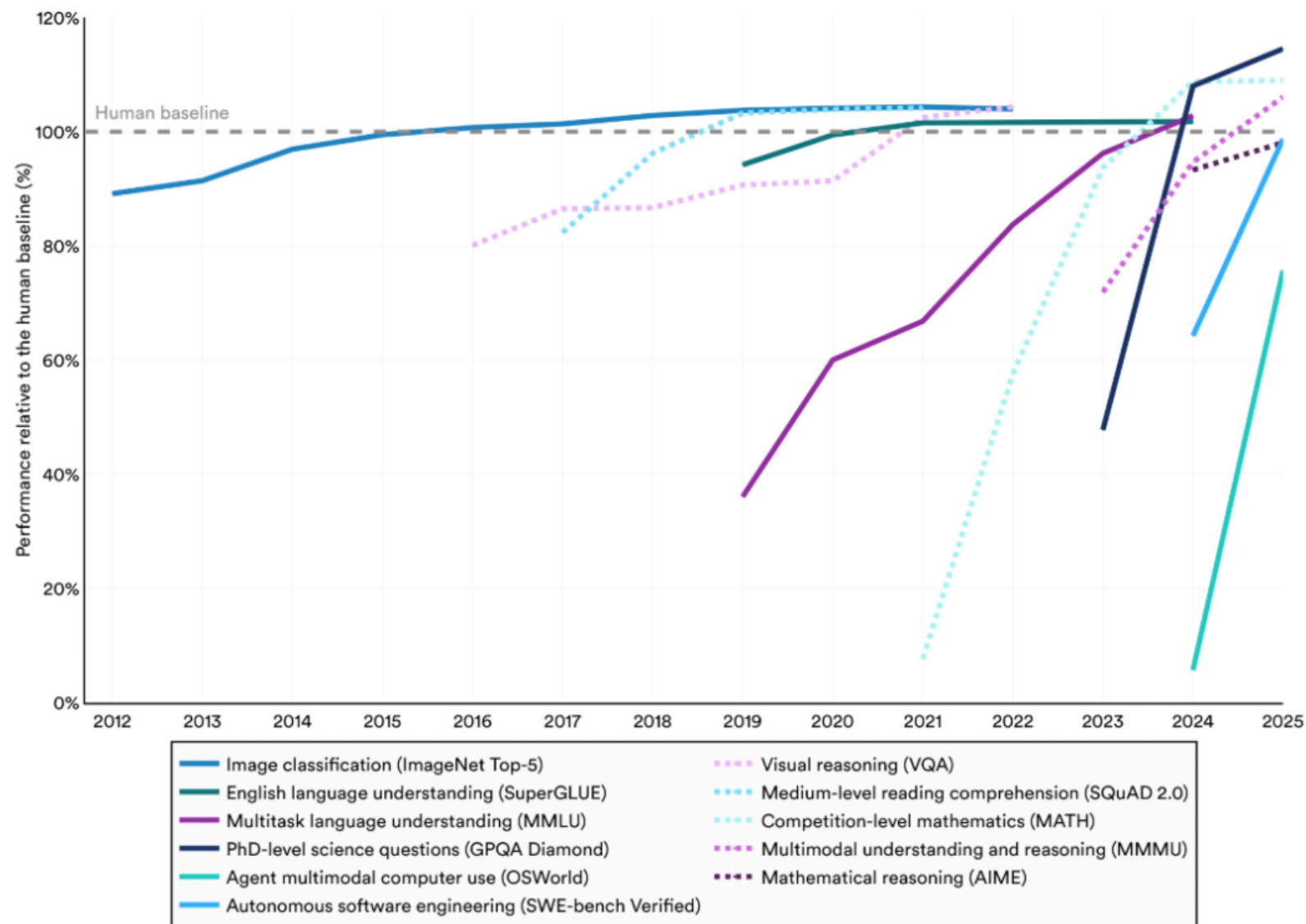
Mathematics

Physics

AI and human benchmarks

Select AI Index technical performance benchmarks vs. human performance

Source: AI Index, 2026 | Chart: 2026 AI Index report



Outline

I. Background

II. Progress in Mathematics

III. Progress in Physics

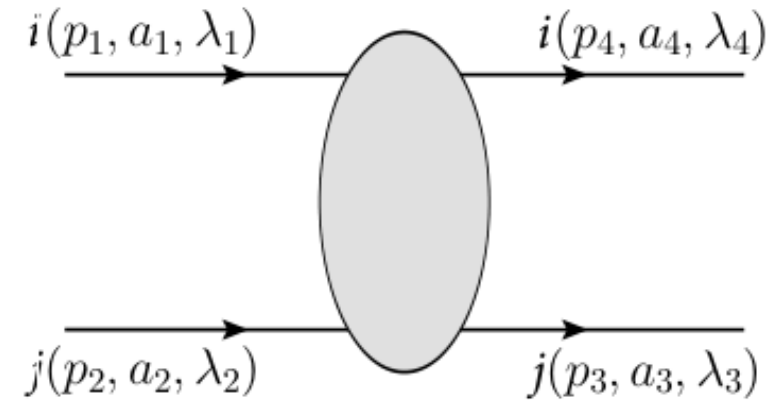
IV. AI-era research

Single-minus amplitudes

OpenAI and collaborators · February 2026

When can a single-minus tree amplitude be nonzero?

- A nonzero amplitude appears in the half-collinear regime, outside the generic-momentum assumption.
- GPT-5.2 proposed a general formula; analytical checks and a graviton extension followed.



$$n \leq 6 \rightarrow \text{all } n$$

Half-collinear; Klein (2,2) kinematics

$$A(1^-, 2^+, \dots, n^+) \neq 0$$

Half-collinear kinematics only.

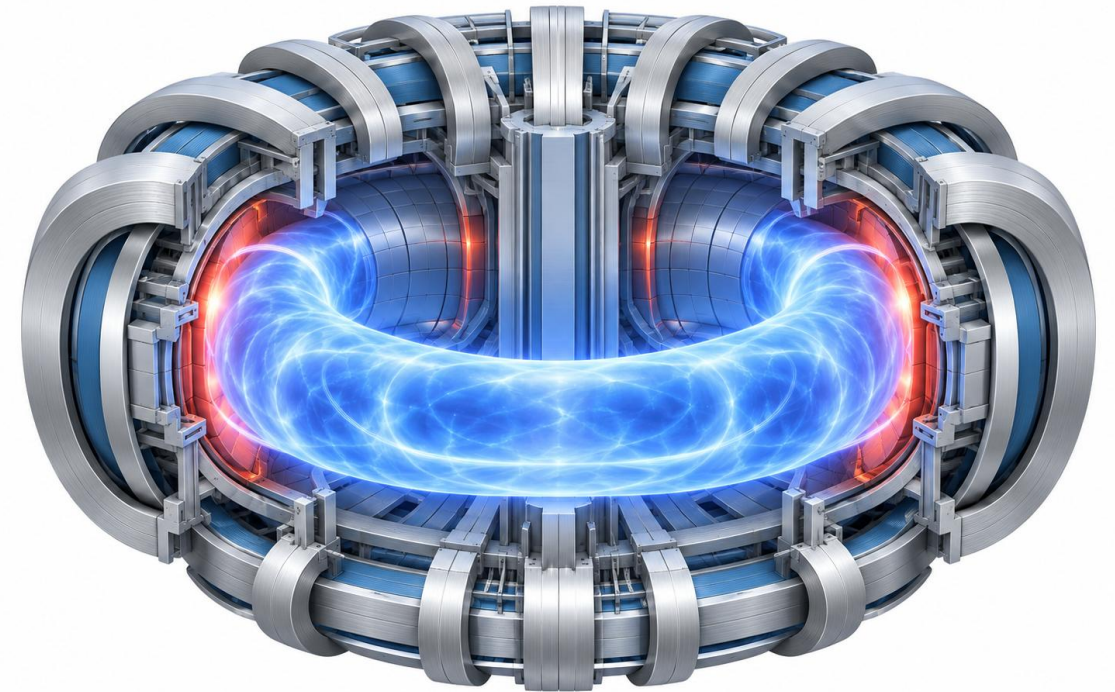
[OpenAI research](#) [Gluon preprint](#) [Graviton preprint](#)

Reinforcement learning on tokamaks

EXL-50U and DIII-D teams · May–August 2026

Can learned controllers regulate real plasmas?

- EXL-50U: stable control in more than ten discharges, with typical tracking errors of 1–5 mm.
- DIII-D: learned policies reached the device for rotation and dynamic shape control.



Concept illustration for nuclear fusion in tokamaks

A joint weather and cyclone model

Google DeepMind · August 2026

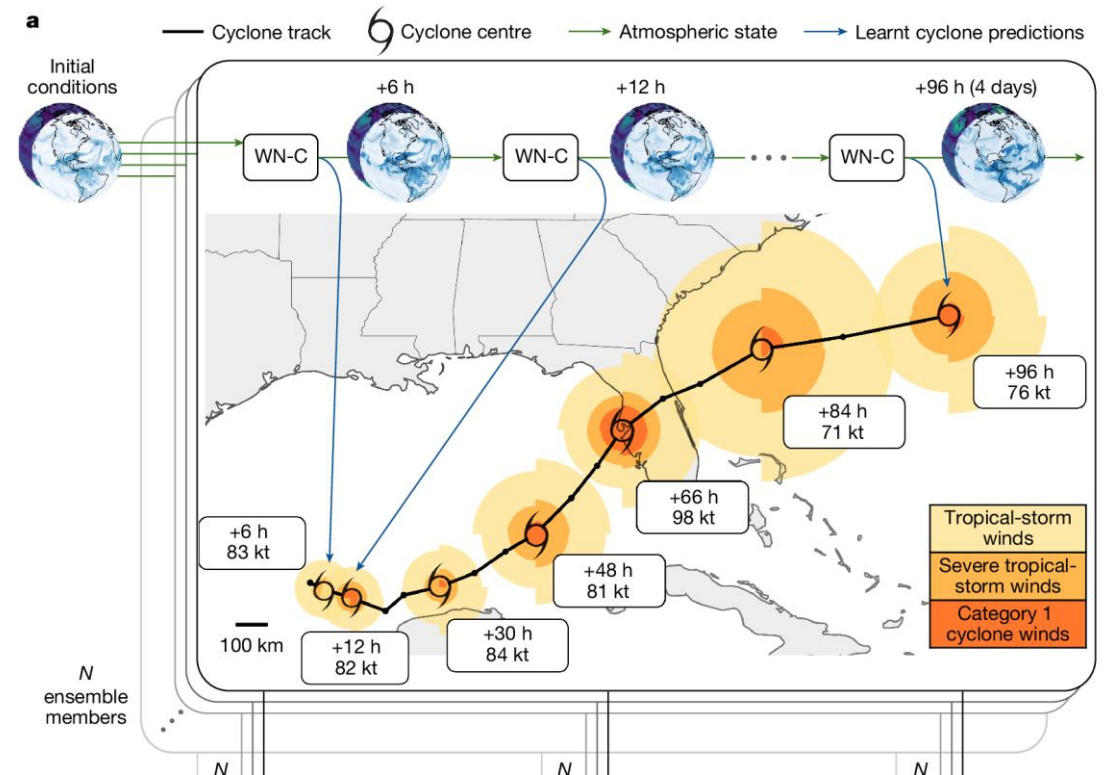
Can one ensemble forecast track, intensity and wind structure?

- One generative ensemble predicts global weather and cyclone track, intensity and wind structure.
- Historical tests show at least one extra day of lead time for key cyclone forecasts.

+24 h

Lead-time gains vary with variable and baseline.

[Nature paper](#) [Code and model resources](#)



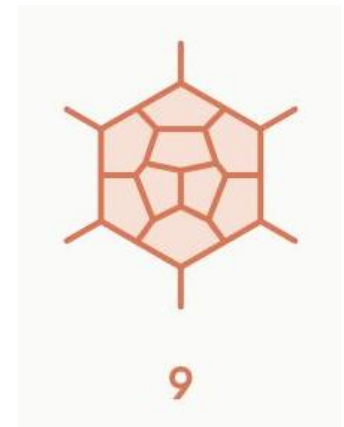
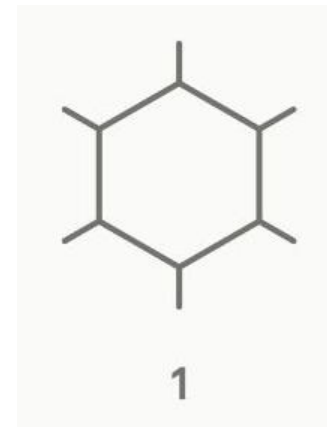
Joint cyclone forecast (Nature Fig. 1a)

Six-gluon amplitudes at nine loops

Claude-assisted computation · September 2026

How far can the hexagon-function bootstrap go?

- A Claude-assisted computation extends the planar N=4 SYM six-gluon amplitude to nine loops.
- Three years after eight-loop result obtained by human. Verified by experts.



8 → 9 loops

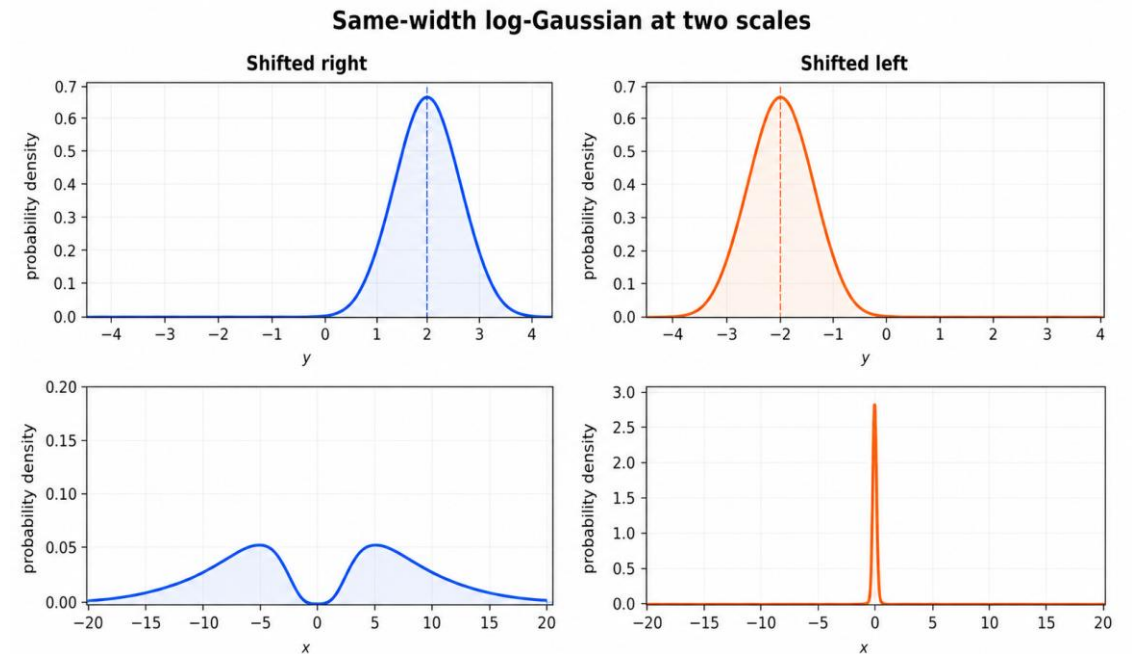
Optimal quantum ordered search

Joseph Carolan and Andrew M. Childs · 28 September 2026

What is the best quantum query complexity for a sorted list?

- Claude found a zero-error algorithm through a continuum relaxation.
- GPT-5.6-Sol then found an exact algorithm. Both attain the optimal leading query constant.

$$(1/\pi) \ln n + o(\log n)$$



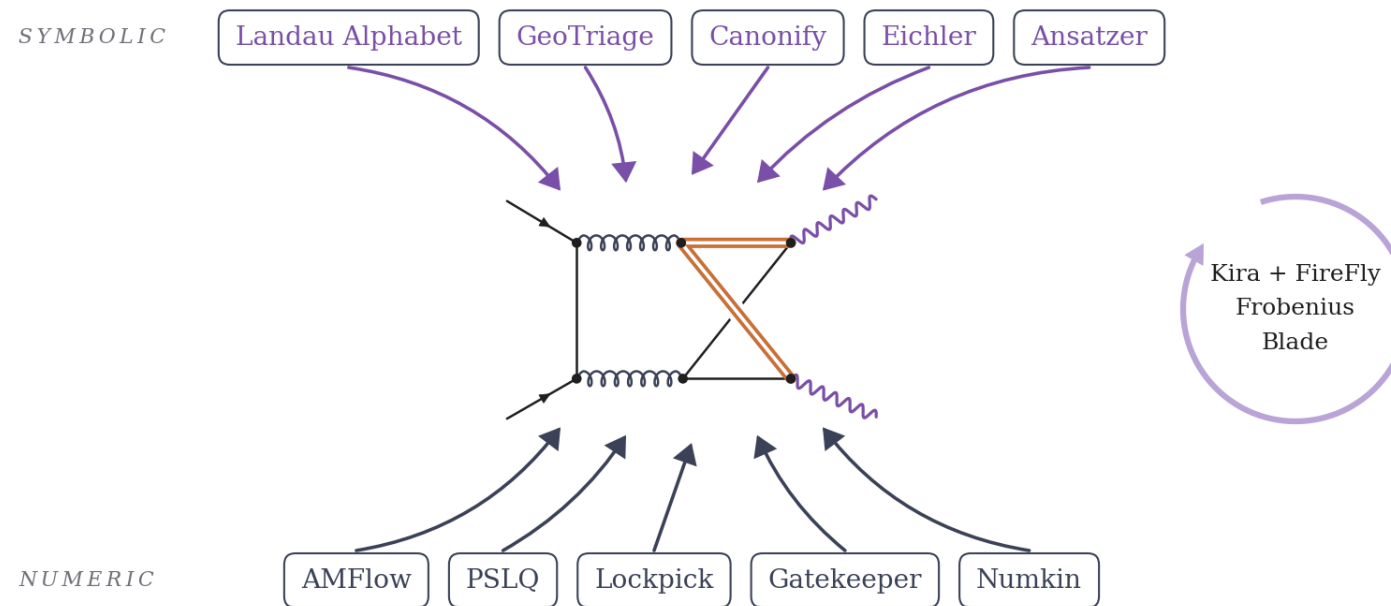
Log-position packets (paper Fig. 2)

BootLoops and exact computation

Matthew D. Schwartz · 2 October 2026

Exact computation with independent checks

- Symbolic constraints and certified numerics solved 30 frontier integrals.



Symbolic and numerical tools (BootLoops)

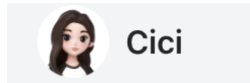
[Toolkit and documentation](#) [Overview preprint](#) [Watson result](#)

AI: Conquering One Field after Another

The beginning
of the end:

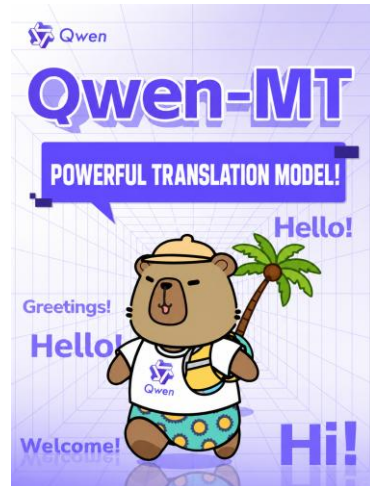
2022

- Chat client



2023

- Translation



2024

- Coding



2025

- Mathematics



2026

- Theoretical
Physics?



Agent

LLM

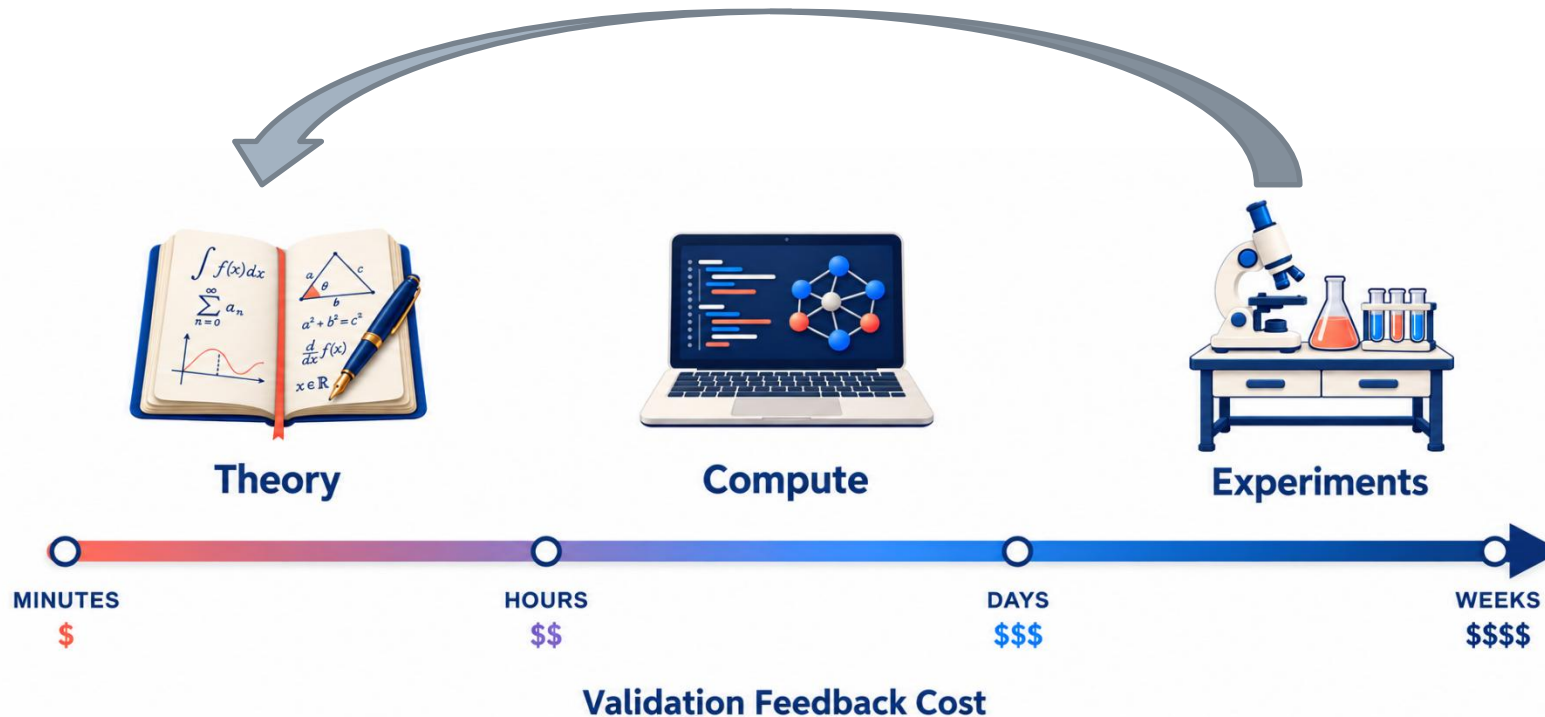
AI

Mathematics

Physics

Why the physics pig on the wall?

- Formal proofs provide fast, reusable checks.
- Physical models ultimately face experiment.
- Feedback cost shapes the pace of AI progress.



Outline

I. Background

II. Progress in Mathematics

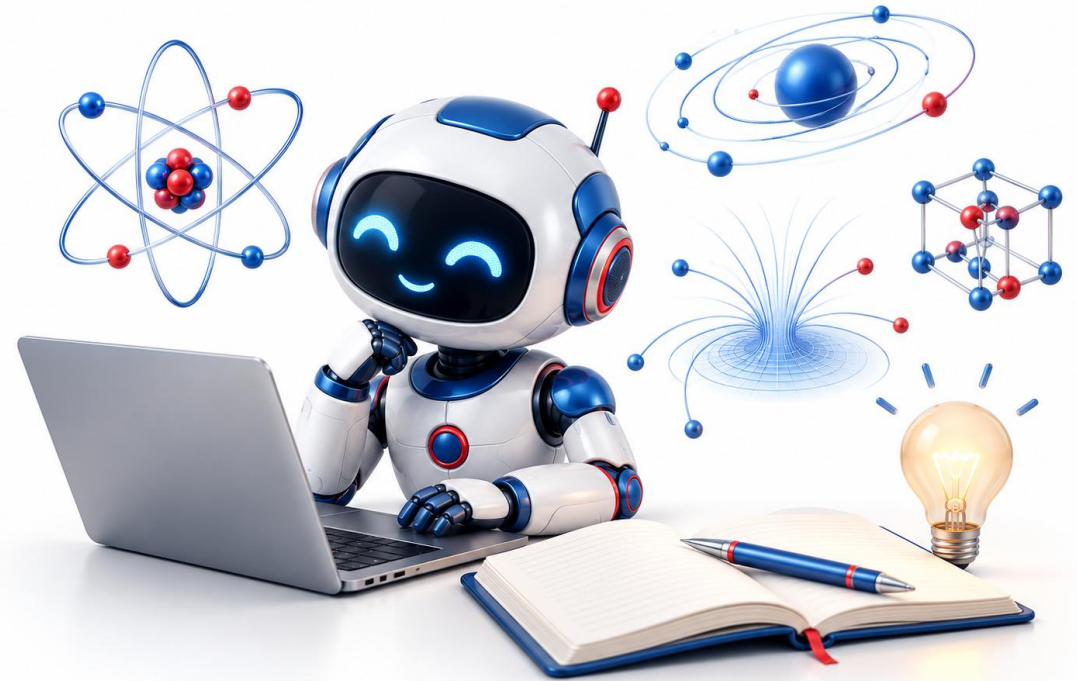
III. Progress in Physics

IV. AI-era research

What can AI do?

Execution at scale, with expert direction

- Derive formulas and write research code.
- Use established tools and propose candidate ideas.



Human judgment remains essential

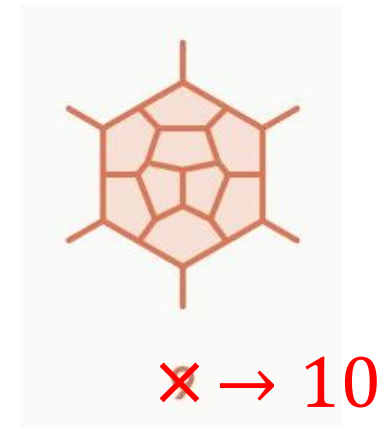
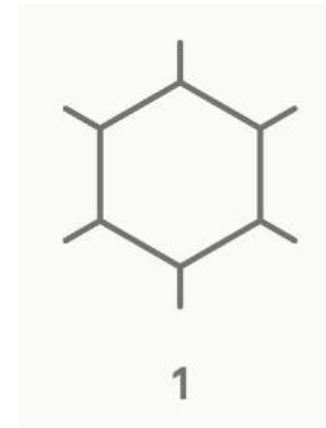
Experts choose questions and validate physical assumptions.

Six-gluon amplitudes at ten loops

Song He and collaborators · Work in progress

A new method reaches ten loops

- AI-assisted work established the nine-loop benchmark.
- A new route reportedly reaches both nine and ten loops.



Earlier one- and nine-loop illustrations

Manuscript not yet public

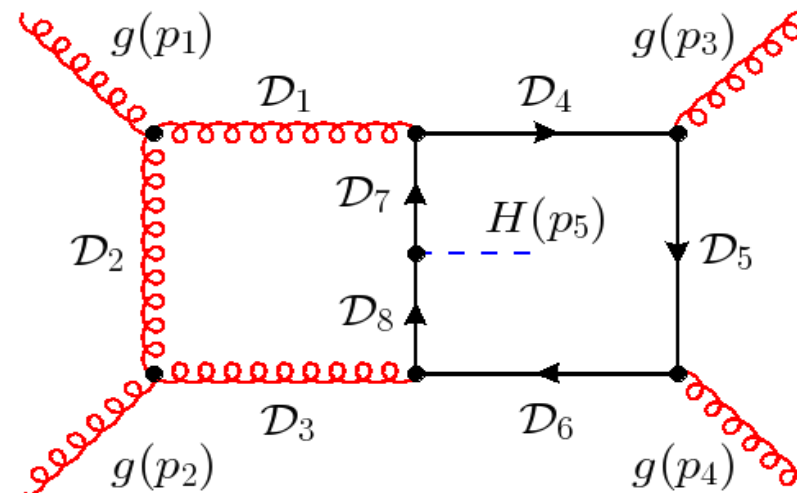
8 → 9 → 10 loops

Beyond BootLoops

Our group · 1 October 2026

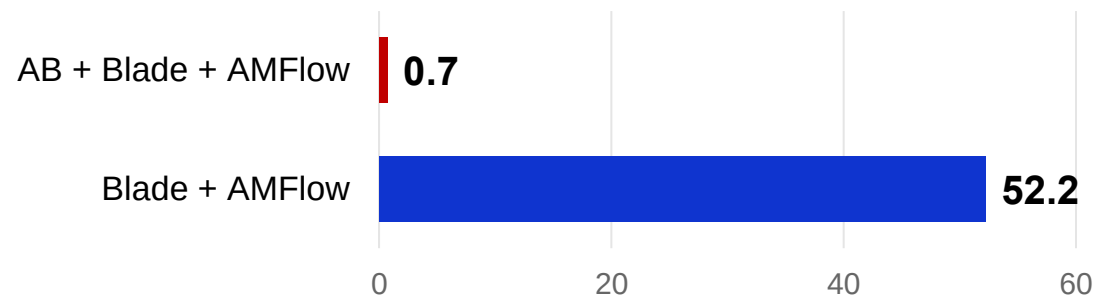
Analyticity bootstrap of Feynman integrals

- Analytic regularity constrains the reduction ansatz.
- Fewer IBP samples accelerate differential-equation construction.



Double-pentagon family (paper Fig. 3)

DE construction (CPU hours)



52.2 → 0.7 CPU h

Future of physics

Stronger tools, deeper questions

- Explore richer models and more demanding calculations.
- Research organization, review and credit must adapt.

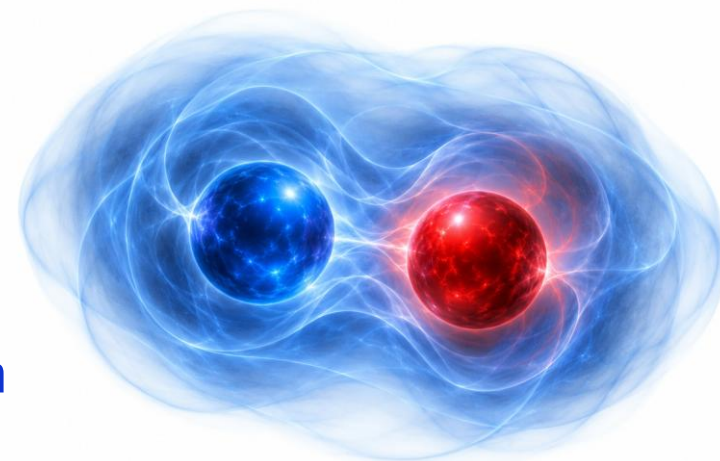
**Physics continues.
Research methods change.**



Quarkonium physics

Theoretical developments

- AI can accelerate theoretical calculations.
- Our question: establish or challenge NRQCD factorization for production.
(Several versions obtained; flaws identified upon review)



Next year: a research goal

$$\sigma(H) = \sum_n \hat{\sigma}_n \langle O_n^H \rangle$$

Short-distance coefficients $\times$ long-distance matrix elements

GDP: Survival skills in the AI era

Core Judgment

- My near-term view: AI as a supervised research collaborator
- Research with AI resembles supervising a capable graduate student

GDP is all you need!

G (Gou Tong) — Communication: discuss ideas and assign clear tasks

D (Dong Cha) — Insight: choose sound routes and catch errors

P (Pin Wei) — Taste: decide which questions deserve the work

Test-driving research paradigm is coming! Build your workflow!

Thank you!