RESEARCH

AI achieves silver-medal standard solving International Mathematical Olympiad problems

25 JULY 2024

AlphaProof and AlphaGeometry teams

⤳ Share

We've made great progress building AI systems that help mathematicians discover [new insights](#), [novel algorithms](#) and answers to [open problems](#). But current AI systems still struggle with solving general math problems because of limitations in reasoning skills and training data.

Today, we present AlphaProof, a new reinforcement-learning based system for formal math reasoning, and AlphaGeometry 2, an improved version of our [geometry-solving system](#). Together, these systems solved four out of six problems from this year's [International Mathematical Olympiad](#) (IMO), achieving the same level as a silver medalist in the competition for the first time.

Breakthrough AI performance solving complex math problems

The IMO is the oldest, largest and most prestigious competition for young mathematicians, held annually since 1959.

Each year, elite pre-college mathematicians train, sometimes for thousands of hours, to solve six exceptionally difficult problems in algebra, combinatorics, geometry and number theory. Many of the winners of the [Fields Medal](#), one of the highest honors for mathematicians, have represented their country at the IMO.

More recently, the annual IMO competition has also become widely recognised as a grand challenge in machine learning and an aspirational benchmark for measuring an AI system's advanced mathematical reasoning capabilities.

This year, we applied our combined AI system to the competition problems, provided by the IMO organizers. Our solutions were scored according to the IMO's point-awarding rules by prominent mathematicians [Prof Sir Timothy Gowers](#), an IMO gold medalist and Fields Medal winner, and [Dr Joseph Myers](#), a two-time IMO gold medalist and Chair of the IMO 2024 Problem Selection Committee.

> **The fact that the program can come up with a non-**

obvious construction like this is very impressive, and well beyond what I thought was state of the art.

PROF SIR TIMOTHY GOWERS,
IMO GOLD MEDALIST AND FIELDS MEDAL WINNER

First, the problems were manually translated into formal mathematical language for our systems to understand. In the official competition, students submit answers in two sessions of 4.5 hours each. Our systems solved one problem within minutes and took up to three days to solve the others.

AlphaProof solved two algebra problems and one number theory problem by determining the answer and proving it was correct. This included the hardest problem in the competition, solved by only five contestants at this year's IMO. AlphaGeometry 2 proved the geometry problem, while the two combinatorics problems remained unsolved.

See our system's IMO 2024 solutions ↗

Each of the six problems can earn seven points, with a total maximum of 42. Our system achieved a final score of 28 points, earning a perfect score on each problem solved — equivalent to the top end of the silver-medal category. This year, the gold-medal threshold starts at 29 points, and was achieved by 58 of 609 contestants at the official competition.

Score on IMO 2024 problems

42
40

TOTAL

Graph showing performance of our AI system relative to human competitors at IMO 2024. We earned 28 out of 42 total points, achieving the same level as a silver medalist in the competition.

AlphaProof: a formal approach to reasoning

AlphaProof is a system that trains itself to prove mathematical statements in the formal language Lean. It couples a pre-trained language model with the AlphaZero reinforcement learning algorithm, which previously taught itself how to master the games of chess, shogi and Go.

Formal languages offer the critical advantage that proofs involving mathematical reasoning can be formally verified for correctness. Their use in machine learning has, however, previously been constrained by the very limited amount of human-written data available.

In contrast, natural language based approaches can hallucinate plausible but incorrect intermediate reasoning steps and solutions, despite having access to orders of magnitudes more data. We established a bridge between these two complementary spheres by fine-tuning a Gemini model to automatically translate natural language problem statements into formal statements, creating a large library of formal problems of varying difficulty.

When presented with a problem, AlphaProof generates solution candidates and then proves or disproves them by searching over possible proof steps in Lean. Each proof that was found and verified is used to reinforce AlphaProof's language model, enhancing its ability to solve subsequent, more challenging problems.

We trained AlphaProof for the IMO by proving or disproving millions of problems, covering a wide range of difficulties and mathematical topic areas over a period of weeks leading up to the competition. The training loop was also applied during the contest, reinforcing proofs of self-generated variations of the contest problems until a full solution could be found.

Process infographic of AlphaProof's reinforcement learning training loop: Around one million informal math problems are translated into a formal math language by a formalizer network. Then a solver network searches for proofs or disproofs of the problems, progressively training itself via the AlphaZero algorithm to solve more challenging problems.

A more competitive AlphaGeometry 2

AlphaGeometry 2 is a significantly improved version of [AlphaGeometry](). It's a neuro-symbolic hybrid system in which the language model was based on [Gemini]() and trained from scratch on an order of magnitude more synthetic data than its predecessor. This helped the model tackle much more challenging geometry problems, including problems about movements of objects and equations of angles, ratio or distances.

AlphaGeometry 2 employs a symbolic engine that is two orders of magnitude faster than its predecessor. When presented with a new problem, a novel knowledge-sharing mechanism is used to enable advanced combinations of different search trees to tackle more complex problems.

Before this year's competition, AlphaGeometry 2 could solve 83% of all historical IMO geometry problems from the past 25 years, compared to the 53% rate achieved by its predecessor. For IMO 2024, AlphaGeometry 2 [solved Problem 4](#) within 19 seconds after receiving its formalization.

Illustration of Problem 4, which asks to prove the sum of ∠KIL and ∠XPY equals 180°. AlphaGeometry 2 proposed to construct E, a point on the line BI so that ∠AEB = 90°. Point E helps give purpose to the midpoint L of AB, creating many pairs of similar triangles such as ABE ~ YBI and ALE ~ IPC needed to prove the conclusion.

New frontiers in mathematical reasoning

As part of our IMO work, we also experimented with a natural language reasoning system, built upon [Gemini](#) and our latest research to enable advanced problem-solving skills. This system doesn't require the problems to be translated into a formal language and could be combined with other AI systems. We also tested this approach on this year's IMO problems and the results showed great promise.

Our teams are continuing to explore multiple AI approaches for advancing mathematical reasoning and plan to release more technical details on AlphaProof soon.

We're excited for a future in which mathematicians work with AI tools to explore hypotheses, try bold new approaches to solving long-standing problems and quickly complete time-consuming elements of proofs — and where AI systems like Gemini become more capable at math and broader reasoning.

managed the overall project.

Related posts

View all posts

TECHNOLOGIES

Gemini breaks new ground: a faster model, longer context and AI agents

TECHNOLOGIES

Our next-generation model: Gemini 1.5

The model delivers dramatically enhanced performance, with a...

We're introducing a series of updates across the Gemini...

14 MAY 2024

15 FEBRUARY 2024

TECHNOLOGIES

Introducing Gemini: our largest and most capable AI model

Making AI more helpful for everyone

6 DECEMBER 2023

RESEARCH

AlphaGeometry: An Olympiad-level AI system for geometry

Advancing AI reasoning in mathematics

17 JANUARY 2024

RESEARCH

AlphaZero: Shedding new light on chess, shogi, and Go

In late 2017 we introduced AlphaZero, a single system that...

RESEARCH

Competitive programming with AlphaCode

Solving novel problems and setting a new milestone in...

6 DECEMBER 2018 →

8 DECEMBER 2022 →

Follow us

About ⌄

Learn more ⌄

Sign up for updates on our latest innovations

✉ Email address   Sign up

I accept Google's Terms and Conditions and acknowledge that my information will be used in accordance with Google's Privacy Policy.

Google

About Google   Google products   Privacy   Terms   Manage cookies