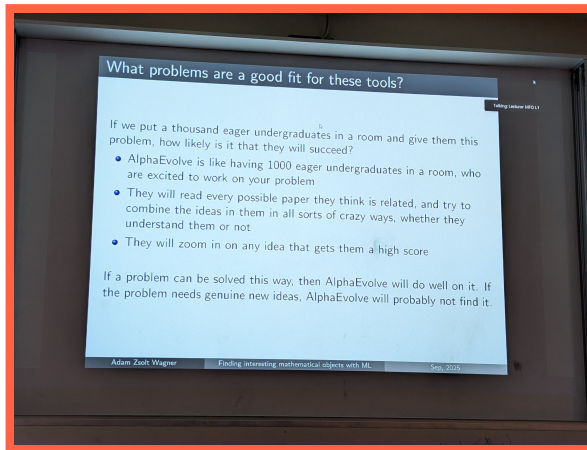


AI for the working mathematician - part 2?

Or: What is...proving with AI?

The slogan



- ▶ Above Photo from one of Adam Wagner's (Google DeepMind) slides
- ▶ AlphaEvolve = LLM-type AI to autonomously discover and refine "things"
- ▶ In general This is what LLM-type AI can do in 2025 \Rightarrow Let's exploit this!

Example of the slogan

From the
AlphaEvolve
paper:

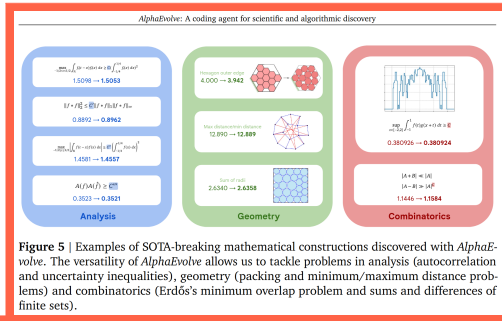
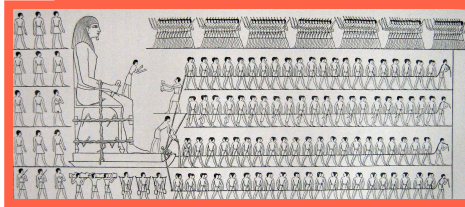


Figure 5 | Examples of SOTA-breaking mathematical constructions discovered with *AlphaEvolve*. The versatility of *AlphaEvolve* allows us to tackle problems in analysis (autocorrelation and uncertainty inequalities), geometry (packing and minimum/maximum distance problems) and combinatorics (Erdős’s minimum overlap problem and sums and differences of finite sets).



- You can do amazing things with enough people and good coordination
- Example How big does a hexagon need to be to fit 12 unit hexagons inside?

An example from my research

Think longer for this one.

I have the below sequence $a[n,m]$. I know that $b[n]=\sum_m a[n,m]$ is asymptotically $\sqrt{\pi/2} \cdot n^{1/2} \cdot (-1/2)^{n/2} \cdot n$.

For $l \geq 2$, I consider $c[n]=\sum_m (-1)^m \mod l \cdot a[n,m]$, and I believe that it has the asymptotic of $b[n]$ divided by l .

How to prove that?

(*binomial that is 0 unless k is an integer in $[0,n]$ *)
 $\text{safeBinom}[n, k] := \text{If}[\text{IntegerQ}[k] \ \&\& \ 0 \leq k \leq n, \text{Binomial}[n, k], 0];$

(*case (1): $a[n,m]=C(n,k) \cdot C(n,k-1), k=(n-m)/2$ *)
 $a[n, m] :=$
 $\text{Module}[\{k = (n - m)/2\}, \text{safeBinom}[n, k] \cdot \text{safeBinom}[n, k - 1];$



Thought for 3m 17s

Great question — and you're on the right track. Let me (i) sanity-check the exact array you wrote, (ii) pin down what b_n really is for that array, and then (iii) give a clean proof that the congruence class sums are asymptotically b_n/l for any fixed $l > 2$ (with an explicit error term).

1) What your $a_{n,m}$ counts and what b_n actually is

With

$$a_{n,m} = \binom{n}{k} - \binom{n}{k-1}, \quad k = \frac{n-m}{2},$$

```
(*case (1): a[n,m]=C(n,k)-C(n,k-1), k=(n-m)/2*)
a[n, m_] :=
Module[{k = (n - m)/2}, safeBinom[n, k] - safeBinom[n, k - 1];
```

Thought for 3m 17s

- The user has a sequence with a binomial difference. The parameters are defined, so I need to examine how the sum behaves, considering m runs over k values, and whether it results in a telescoping sum. It seems like it might relate to the central binomial coefficient difference, considering boundary conditions. I'll explore how the sum simplifies.
- I've noticed that the sum telescopes nicely, with results leading to 1, but this doesn't seem to match the asymptotics the user provided for $b[n]$. While the mapping fits for some cases, the sum structure and boundary conditions are tricky. I wonder if the behavior resembles something like Dyck paths or ballot numbers!
- The telescoping sum computation leads to a constant $b[n]$ of 1, which contradicts the user's expected asymptotics. Something seems off in how I interpreted the sum over m . Maybe they define m in a range tied to symmetric random walks or some specific constraint. I'll need to check that.
- I'm testing the number of symmetric random walks (SRW) and looking at how the difference of binomials relates. The formula seems to connect to ballot and Catalan numbers, with a reflection principle. The sum over m might give something similar to the central binomial coefficient, but with a different constraint on the path.
- I'm trying to figure out if $b[n]$ is correctly defined as the sum of $a[n,m]$ or if there's a misdefinition. The sum seems to telescope correctly, but the result doesn't match the expected asymptotic behavior (which should resemble the central binomial coefficient). Something doesn't add up; possibly a different definition for $a[n,m]$ is needed, or maybe I got the sum range wrong. I'll check with actual calculations soon.

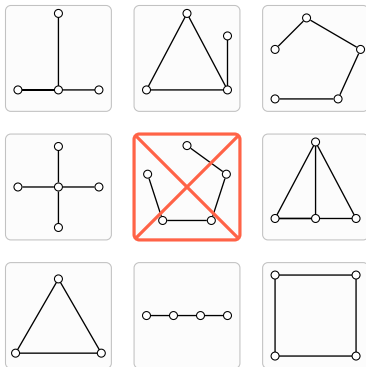
14 Evaluating asymptotics of sequence b[n]

import math



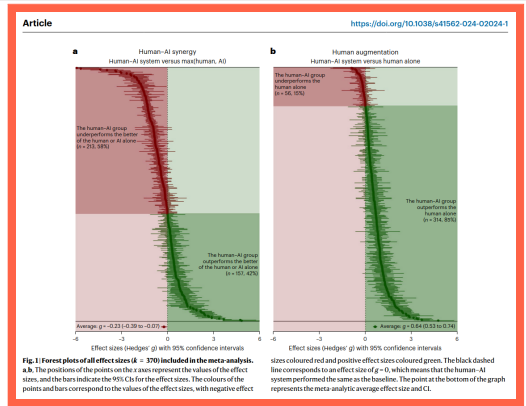
- Above I asked a question to ChatGPT on a lemma within a research paper
- Result? The proof it gave was not correct, but it had the right idea and I was able to fill in the details myself
- The point It knows the literature, so it can piece things together that you did not expect to fit together, but it does that in a bit of a wacky way

Another example: Finding proofs by counterexample-first



-
- ▶ **Setup** Encode the claim as checks; choose a finite search space and a verifier (Sage/Magma/Lean/SnapPy)
 - ▶ **AI's role** Propose families, write the generator/shrinker, and triage candidates; keep only verified failures
 - ▶ **Payoff** Either a minimal counterexample or sharper hypotheses, making the final proof short and clean

From the paper When combinations of humans and AI are useful: A systematic review and meta-analysis



- ▶ This is a big one If a human is better than the AI at a task by themselves, then teaming up often boosts performance
- ▶ As a working mathematician you can exploit and guide your own intuition
- ▶ Why? Because proofs often have literature-reading parts (which AI does well) and parts needing insight (which humans do well)

Thank you for your attention!

I hope that was of some help.