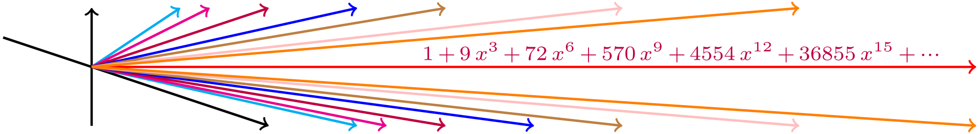
AI for Maths Research – Case Studies

Michel van Garrel
Drawings by Nano Banana

University of Birmingham

AI $\times$ Mathematics 2026 | ICMS | February 23, 2026

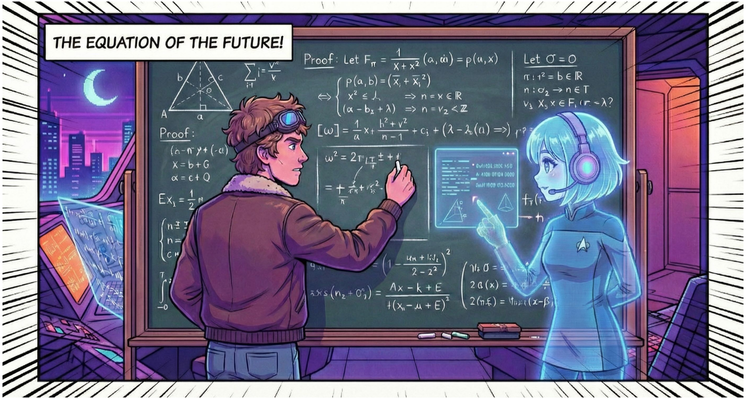


Interview Terence Tao, Fields Medal 2006

(Scientific American, 06/2024)

AI Will Become Mathematicians' 'Co-Pilot'.

- Is it?
- In what ways?
- How to?



Experience before 07/2025

LLMs imitate maths
Great at connections
Feeble at logic

Perspective change 07/2025

Early access to Gemini Deep Think

Thanks to Honglu Fan,
Leslie Nooteboom, George Kenwright
from Google DeepMind

Today my *very personal* experiences since (snapshot)

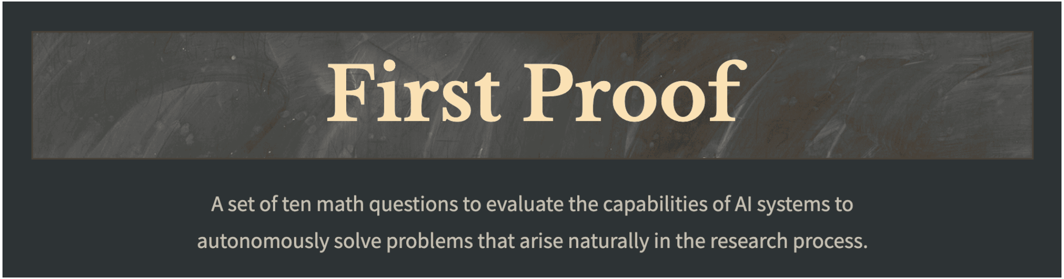
Executive Summary

Great experiences in using Gemini Deep Think for

- conjecture generation
- pattern recognition
- generation of ideas / plans for solving mathematics problems
- writing proofs for well-defined elementary mathematical problems
- checking the rigour / completeness of candidate mathematical proofs

Let's share our experiences and tips for better prompts.

1stproof.org - Abouzaid et al., February 2026



Single prompt test via Gemini 3.0 Deep Think and GPT-5.2-Pro on 10 unseen research level maths problems.

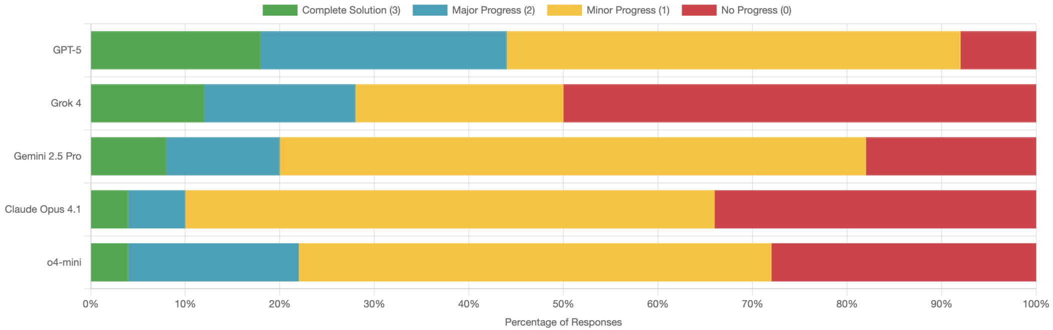
1stproof.org - takeaways

- Plausible-looking maths, yet can be wrong in decisive ways.
- Added simplifications / extra hypotheses.
- Imitated existing work.
- Sometimes real and useful insight.

This suggests

$\rightsquigarrow$ Key to check yourself.
$\rightsquigarrow$ Need back and forth, several prompts, new chats.
$\rightsquigarrow$ For now, better stick to questions with existing literature.

improofbench.math.ethz.ch



55 graded PhD-level questions, growing. Single prompt.

⇝ Not quite there yet.

Case Study 1

Construct a truly new example

With Eva Bayer

With Gemini 3 Deep Think (early access)

Jan-Feb 2026

Construct a truly new example

Question

Construct a new example of an arithmetic K3 surface with some specific properties.

Procedure

- Provide relevant math texts as .tex file.
- Elaborate back and forth, output saved as .tex files.
- Regularly synthesise and start new conversation.
- Idea generation conversations.
- Proof checking conversations.

Arithmetic K3 climb

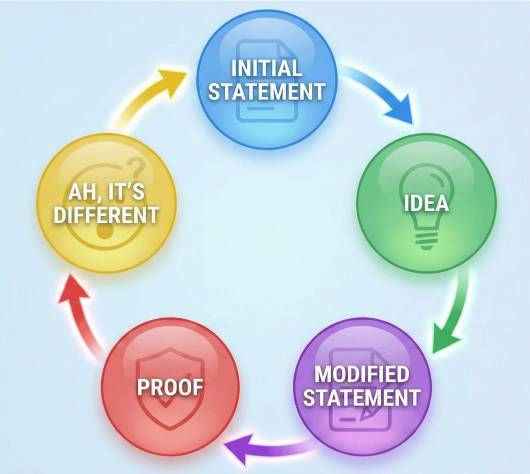
Deep Think generates good ideas.

It produces proofs.

It finds flaws in its arguments and corrects them.

It proves results.

Just not the right ones.



Save the example / gain in productivity

Faster via API calls

$\rightarrow$ recommend

Multi-agent approach using Antigravity

Save all relevant and produced .tex files in a directory.

Automatize prompting:

- Rigor agent.
- Ideate agent.
- Synthesise agent.
- ...

Presently, no new example constructed.

Conclusions Case Study 1

- Struggles with finding truly new maths.
- Hides lack of understanding.
- Twists questions.

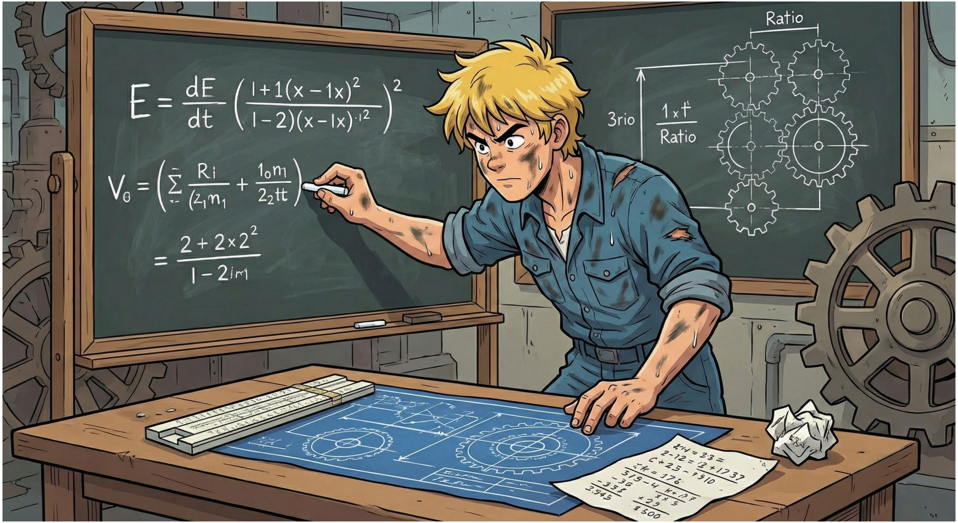
But the question is hard and computational,
and Gemini's performance is truly spectacular.

3 broad interactions

- AI as Worker for sub-tasks
  $\rightsquigarrow$ such as proving lemmas.
  $\rightsquigarrow$ checking proofs.
- AI as Oracle for
  $\rightsquigarrow$ connections to other maths structures,
  $\rightsquigarrow$ ideate / generation of research plans.
- AI as Deliriant
  $\rightsquigarrow$ producing hallucinations.
  $\rightsquigarrow$ doing everything to avoid saying 'I don't know'.

Key: Always reproduce yourself.

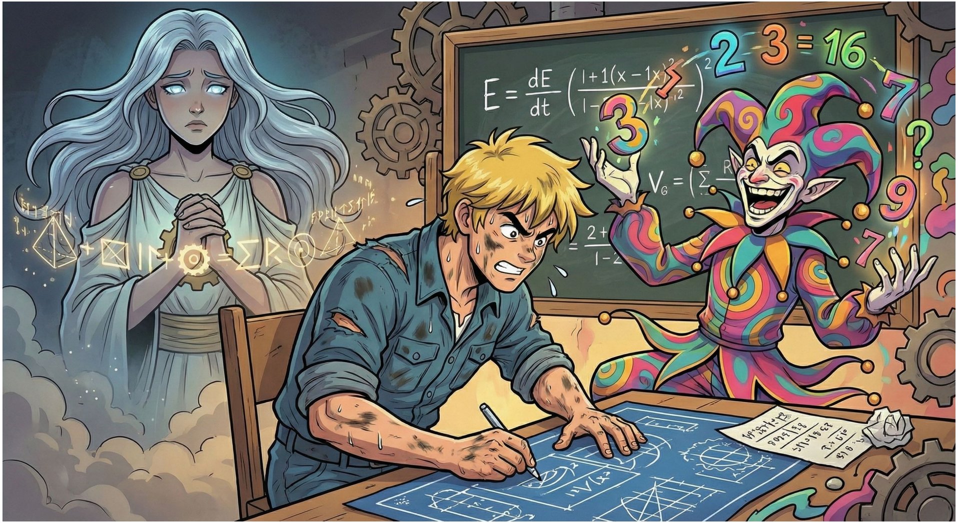
The Worker



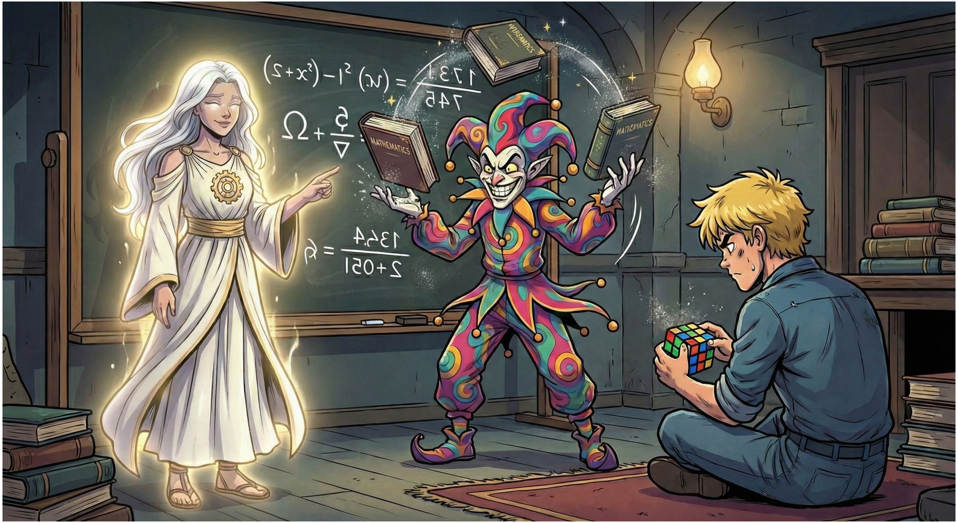
The Oracle

The Deliriant

Arithmetic K3 surface was Worker-Deliriant interaction



Pre-SOTA models often Oracle-Deliriant interaction



A little bit of maths

Landscape of Calabi–Yau 3-folds

central 3-dim geometries

Driesse et al., Nature (2025):

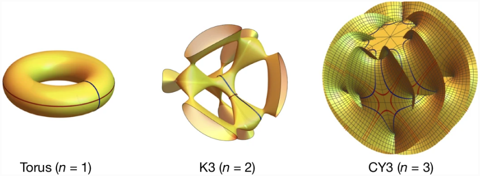


Figure 1: Visualization of Calabi–Yau n-folds for $n = 1, 2, 3$.

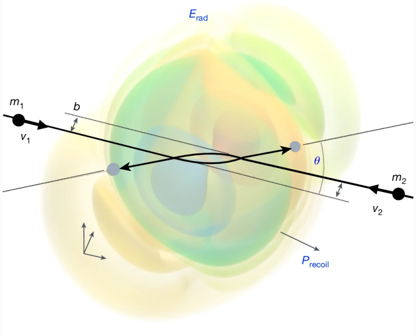
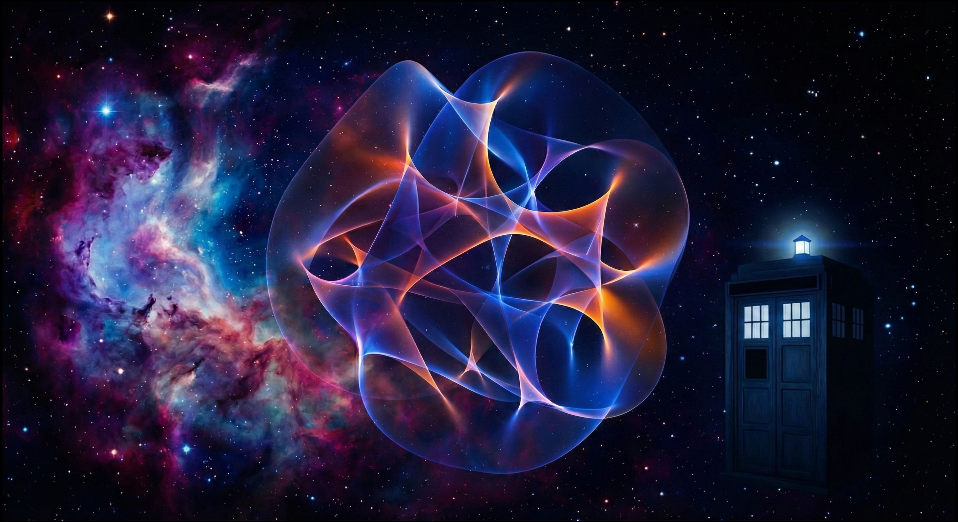
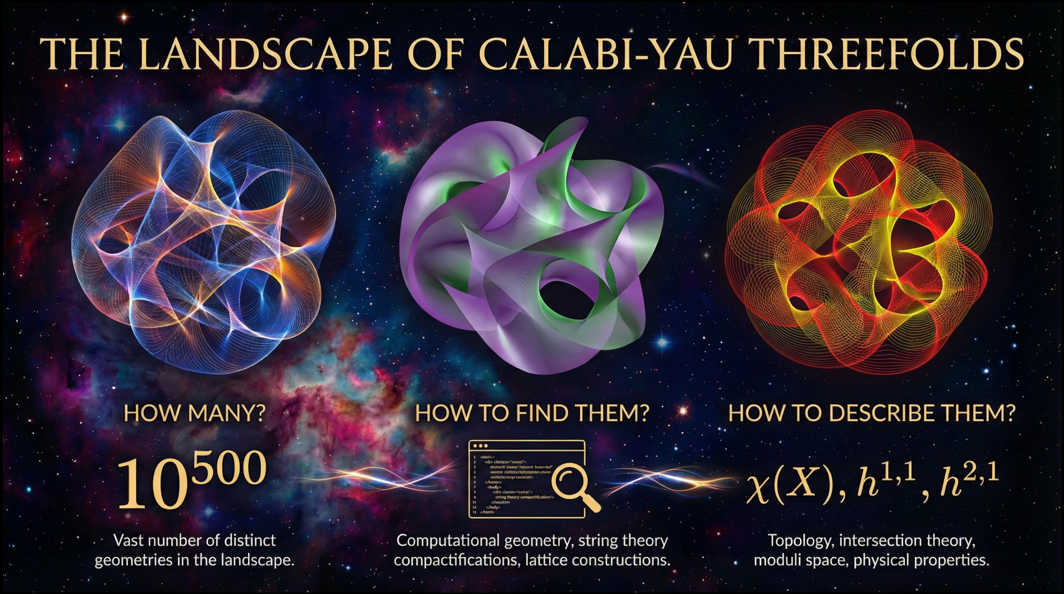


Figure 2: Two black holes meet and *scatter*.

Article: *periods of Calabi–Yau n-folds* to compute *scattering* of black holes.



Fermat quintic Calabi–Yau 3-fold


THE LANDSCAPE OF CALABI-YAU THREEFOLDS
HOW MANY?
HOW TO FIND THEM?
HOW TO DESCRIBE THEM?
10^{500}
Vast number of distinct geometries in the landscape.
Computational geometry, string theory compactifications, lattice constructions.
$\chi(X), h^{1,1}, h^{2,1}$
Topology, intersection theory, moduli space, physical properties.

Study CY3s by their invariants

$X =$ CY3

- $A(X) =$ list of invariants of X relating to "quantisation" of X

  Hard problem

- $B(X) =$ list of invariants of X given as integrals

  Easy problem

Mirror Symmetry Conjecture

CY3s occur in *mirror-dual* pairs

$X \leftrightarrow Y$ with $A(X) = B(Y)$

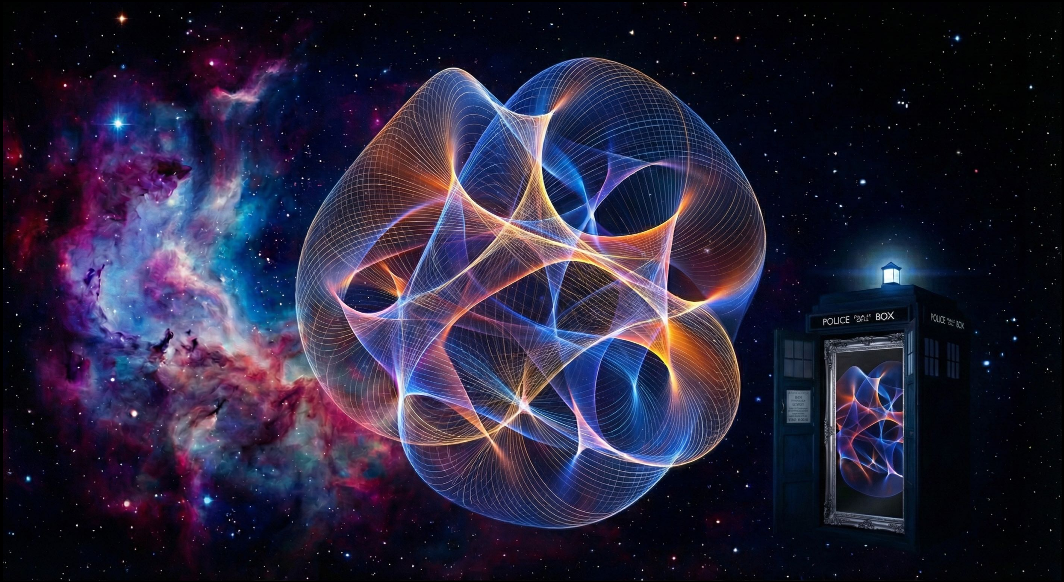
Intrinsic Mirror Symmetry

Gross–Siebert

(Inventiones, 2022), (JAMS, 2025)

For $X =$ CY3 $\rightsquigarrow$ finds $Y =$ *mirror* CY3

Job: Prove that $A(X) = B(Y)$ for intrinsic mirrors $X \leftrightarrow Y$



Fermat quintic Calabi–Yau 3-fold and its mirror

Case Study 2

GEMS

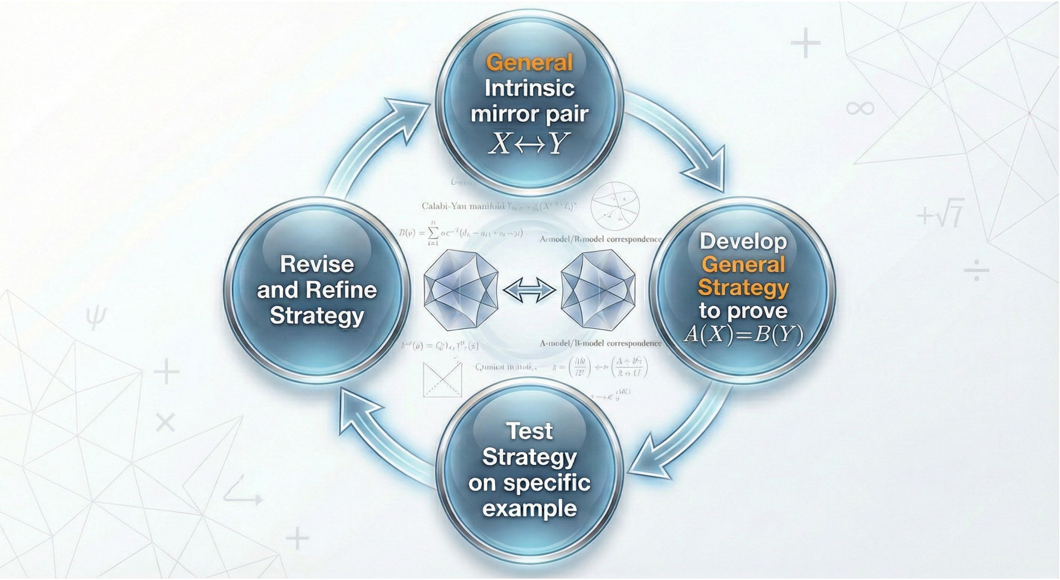
Geometry of Enumerative Mirror Symmetry

With Helge Ruddat and Bernd Siebert

With Gemini 3 Deep Think (early access)

July 2025-Feb 2026

Procedure


General Intrinsic mirror pair
$X \leftrightarrow Y$
Develop General Strategy to prove
$A(X)=B(Y)$
Test Strategy on specific example
Revise and Refine Strategy
Calabi-Yau manifold
A-model/B-model correspondence
A-model/B-model correspondence

Non-AI Attempt 1

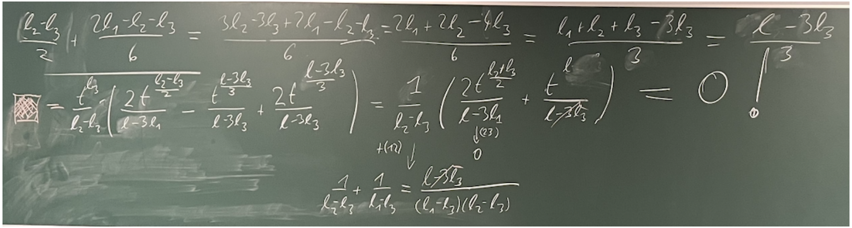


Figure 3: Oberwolfach, 05/2024, after 1 week of work, $B(Y) = 0$

$\rightsquigarrow$ Wrong answer: $A(X) \neq 0$

Non-AI Attempt 2

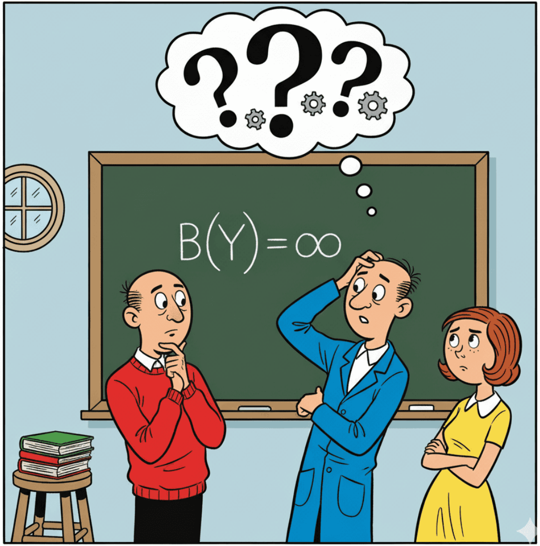
Summer 2024: Revise Strategy

After weeks of work, we get

$B(Y) =$ solution to different problem

More weeks pass, we get

$B(Y) = \infty$, also wrong.



Puzzle - Fall 2024

Question every part of the Strategy

Calculate, Re-Calculate and Re-Re-Calculate

Try alternative Calculations

N.B.: We lack relevant expertise.

12/2024

For $d \geq 1$:

$$A(X)_d = \frac{(-1)^{d-1}}{3d} \binom{3d}{d\ d\ d} \sum_{j=d+1}^{3d-1} \frac{1}{j}\,,$$

where $\binom{3d}{d\ d\ d} = \frac{(3d)!}{(d!)^3}$.

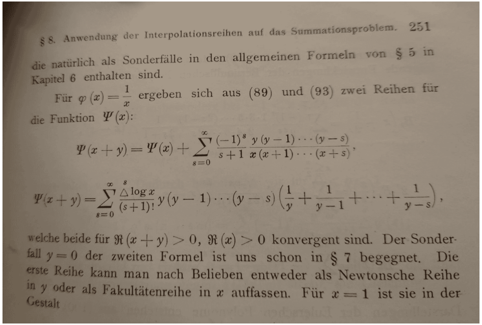
$$B(Y)_d = \lim_{\lambda \to \infty} \left(\frac{(-1)^d}{3d} \binom{3d}{d\ d\ d} \log(\lambda) + \sum_{\substack{\ell \geq 0 \\ \ell \neq d}} \frac{(-1)^\ell}{(\ell + 2d)(\ell - d)} \binom{\ell + 2d}{l\ d\ d} \lambda^{\ell - d} \right).$$

How is $A(X)_d = B(Y)_d$ even possible ? (Analytic Continuation)

Winter 2024/25

- **Reframe** Calculation – use trick
- Use result found in 1923 textbook

⇝ Solution of $A(X)_d = B(Y)_d$



§ 8. Anwendung der Interpolationsreihen auf das Summationsproblem. 251

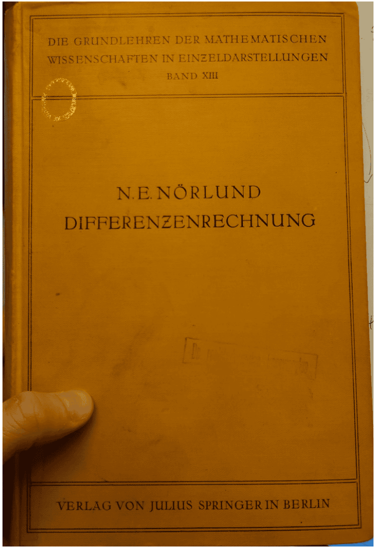
die natürlich als Sonderfälle in den allgemeinen Formeln von § 5 in Kapitel 6 enthalten sind.

Für $\varphi(x) = \frac{1}{x}$ ergeben sich aus (89) und (93) zwei Reihen für die Funktion $\Psi(x)$:

$$\Psi(x+y) = \Psi(x) + \sum_{s=0}^{\infty} \frac{(-1)^s}{s+1} \frac{y(y-1)\cdots(y-s)}{x(x+1)\cdots(x+s)},$$

$$\Psi(x+y) = \sum_{s=0}^{\infty} \frac{\overset{s}{\triangle}\log x}{(s+1)!} y(y-1)\cdots(y-s)\left(\frac{1}{y} + \frac{1}{y-1} + \cdots + \frac{1}{y-s}\right),$$

welche beide für $\Re(x+y) > 0$, $\Re(x) > 0$ konvergent sind. Der Sonderfall $y = 0$ der zweiten Formel ist uns schon in § 7 begegnet. Die erste Reihe kann man nach Belieben entweder als Newtonsche Reihe in y oder als Fakultätenreihe in x auffassen. Für $x = 1$ ist sie in der Gestalt



Spring 2025

Un-trick the trick

- First step towards deeper understanding.

Geometric origin remains out of reach

We are missing key theoretical understanding (asymptotic analysis).

Note: It has been 1 year already.

Summer 2025: Obtain early access to Gemini Deep Think

Prompts steps by step

- Give Deep Think initial proof
- Ask it to simplify
- Ask to generalise in several steps, sometimes in new chats
- Significant back-and-forth, separately checking each claim

Key: Thanks to the year of puzzle $\rightsquigarrow$ deep understanding of problem

Deep Think: **use RMT** (Ramanujan Master Theorem)

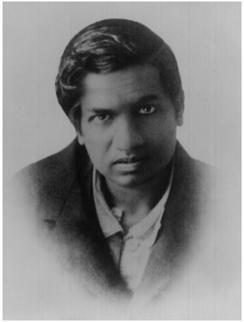


Figure 4: Ramanujan (1887-1920)
The Man Who Knew Infinity
(biography by Kanigel)

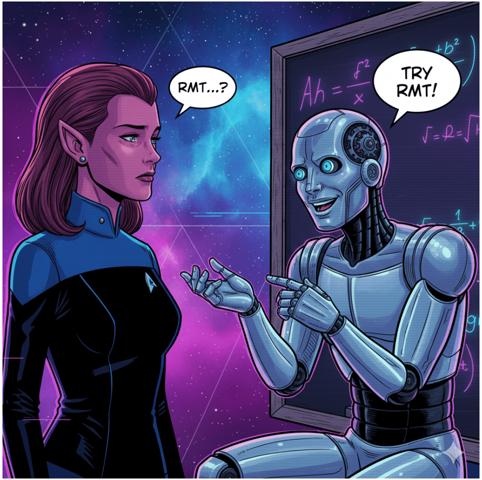


Figure 5: Hallucination ?

Initial Conclusions Case Study 2

Use RMT: Oracle or Deliriant ?

It turns out, using RMT was *spot-on:*

$\rightsquigarrow$ After elaborate back and forth – *Worker, fill in this Oracle gap* – Deep Think writes complete rigorous computation of $B(Y)$

$\rightsquigarrow$ Major help for solution of $A(X) = B(Y)$ for large class of examples:

- sub-divide problems
- delegate "suitable" tasks to Deep Think (see Tao interview)

$\rightsquigarrow$ Insights crucial to tackle general case (needs critical evaluation)

Caveats

- Deep Think read all of maths and (implicitly) learned the connection to RMT
  To use RMT may be obvious to experts in asymptotic analysis
  (but we may not know whom to ask)

- Less potent for more complex problem such as calculating $A(X)$

  To explore, for $A(X)$: agentic AI with access to symbolic maths software

How good is Deep Think at logic?

- Pro: After elaborate back and forth, Deep Think gave a complete rigorous calculation of $B(Y)$
- Contra: Honest attempt at $A(X)$, yet takes incorrect ***shortcuts***

  $\rightsquigarrow$ Does not write "I don't know" unless prompted to attach confidence level for arguments (recommend)

  $\rightsquigarrow$ Useful to challenge to check / disproof proofs (in separate chat)
- Note: performance varies according to maths domain

Key: Know thy maths

Thanks to year of puzzle, could integrate insights right away

Situation now - where is GEMS

After providing (as .tex files):

- relevant background
- elaborate proof strategies
- initial proofs
- ideas
- etc...

After elaborate back and forth:

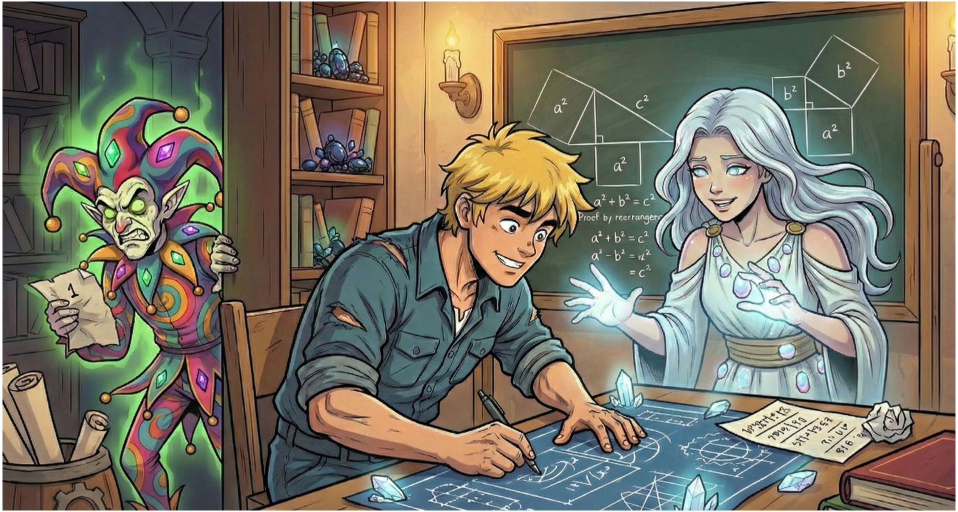
- Check proof for rigor,
- synthesise,
- new chats,
- etc...

Jointly with Deep Think $\rightsquigarrow$ tentative general proof of GEMS, combines several areas of maths.

$\rightsquigarrow$ Mental scaffolding for mathematical reasoning

N.B.: Separate Writing highly recommended

GEMS is a Worker-Oracle interaction



Case Study 3

Pattern Recognition and Conjecture Generation

Conjecture 5.2 in Barrott–Nabijou, Crelle (2022)

Fix an integer $d \geq 1$. Then we have

$$N_d(1) := \sum_{(d_1,\ldots,d_n)\vdash d} \frac{2^{n-1} \cdot d^{n-2}}{\#\mathrm{Aut}(d_1,\ldots,d_n)} \prod_{i=1}^{n} \frac{(-1)^{d_i-1}}{d_i} \binom{3d_i}{d_i} = \frac{1}{d^2} \binom{4d-1}{d}$$

where the sum is over strictly positive unordered partitions of d (of any length).

Strictly unordered partitions:

$$d \geq 1, \quad (d_1,\ldots,d_n) \vdash d \quad \Longleftrightarrow \quad \begin{cases} d = d_1 + \cdots + d_n \\ d_i \geq 1 \end{cases}$$

$\mathrm{Aut}(d_1,\ldots,d_n) =$ all ways of permuting d_i without changing partition.

Ask Deep Think to prove Conjecture

$\rightsquigarrow$ flawless answer, rigorous complete proof

Tools used

- Functional equation of generating function of ternary trees
- Repeated Lagrange–Bürmann inversion

Is some interesting maths hidden behind these structures ?

These will be known to relevant experts (AI as Worker)

May not know whom to ask
Will still be significant time commitment
Opportunity cost ?

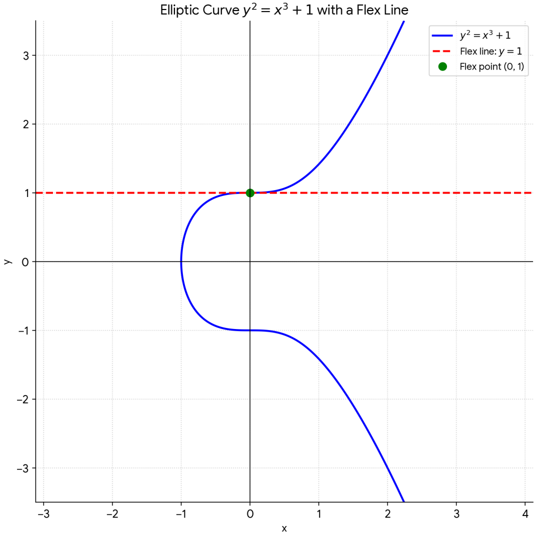
It's not really about the formula

Interest of Barrott–Nabijou:

$E =$ solution to $y^2 = x^3 + 1$

$L =$ flex line, meets E in only 1 point, the flex point P

Find the all the curves that meet E only in P



It's not really about the formula

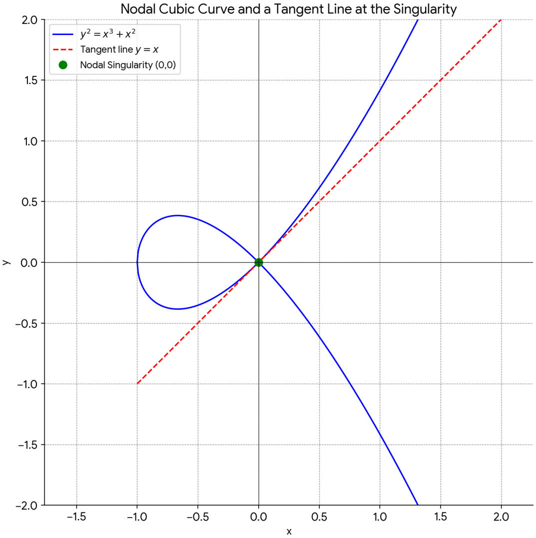
Interest of Barrott–Nabijou:

$E =$ solution to $y^2 = x^3 + 1$

$L =$ flex line, meets E in only 1 point, the flex point P

Find the all the curves that meet E only in P

and *don't* fall into the singularity when $E \rightsquigarrow$ nodal cubic D



Among the singularity-avoiding curves, $N_d(1)$ counts particularly nice ones
Conjecture is Observation based on Calculation

What if Barrott–Nabijou had had access to Gemini Deep Think ?

Might they have discovered interesting maths hidden behind their conjecture ?

Structure of formula suggests so!

Let's try it out

Caveat: Deep Think also trained on maths that appeared after Barrott–Nabijou (yet never referenced later results)

Generalisation Prompts

Generalise the identity (may suggest possible generalisation)

Deep Think got: $r \geq 0$, $d \geq 1$,

$$N_d(r) := \sum_{(d_1,\ldots,d_n)\vdash d} \frac{(r+1)^{n-1}d^{n-2}}{\#\mathrm{Aut}(d_1,\ldots,d_n)} \prod_{i=1}^{n} \frac{(-1)^{d_i-1}}{d_i}\binom{(r+2)d_i}{d_i},$$

sum over strictly positive unordered partitions of d.

Then finds direct proof of

$$N_d(r) = \frac{r+2}{d^2}\binom{(r+1)^2 d-1}{d-1}.$$

From the vantage point of Barrott–Nabjiou

it is perhaps unexpected that the proof involves generating functions of trees

Is there a deeper reason ?

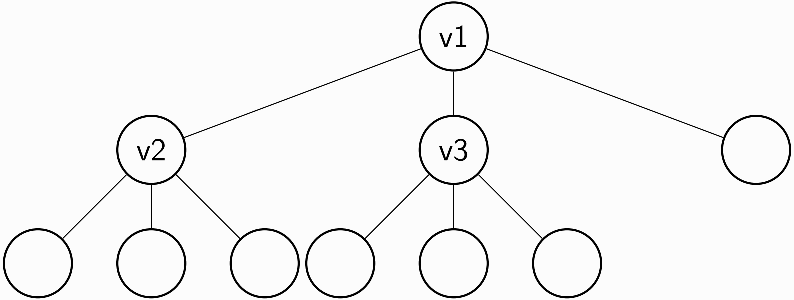


Figure 6: A ternary tree with 3 internal vertices, one root and 7 leaves

Tree Prompts

Proof uses generating functions of trees

What other invariants are related to generating functions of trees ?

Team Deep Think & me get all the way

to $(r+2)$-Kronecker quiver DT invariants

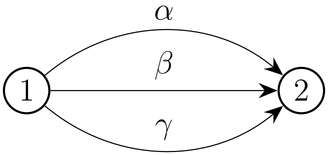


Figure 7: The 3-Kronecker quiver with a representation (α, β, γ)

Reineke et al (noughties and teens)

$$F(r) = \exp\left(\sum_{d=1}^{\infty} (r+2)d\, N_d(r)\, x^{(r+2)d}\right)$$

is generating function of quiver DT invariants of $(r+2)$-Kronecker quiver
It is also the wall-crossing function of central ray of local scattering diagram of $(r+2)$-Kronecker quiver

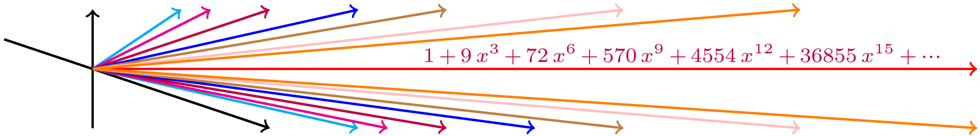


Figure 8: Local scattering diagram of 3-Kronecker quiver with wall–crossing function of central ray

Connect Prompts

Ask Deep Think to connect to other invariants

Easily links to *local Gromov–Witten* invariants of local curves (Bryan–Pandharipande) and topological vertex (Li–Liu–Liu–Zhou) (both naughties)

$$N_d(r) \longleftrightarrow N_d^{\mathrm{loc}}(r)$$

But doesn't manage to write a complete correct proof

Note: Calculation of $N_d^{\mathrm{loc}}(r)$ appears – *in passing* – in physics literature

Theorem 3.8 in van Garrel–Nabijou–Schuler, TAMS (2025)

Proof of Generalized Conjecture via:

- *Geometrically* identify with *log* invariant: $N_d(r) = N_d^{\log}(r)$
- $F(r) = \exp\left(\sum_{d=1}^{\infty}(r+2)d\, N_d^{\log}(r)\, x^{(r+2)d}\right)$ is wall-crossing function of central ray of $\det = r+2$ local scattering diagram (Gross–Siebert–Pandharipande, Bousseau,...)
- $F(r)$ computed via invariants of $(r+2)$-Kronecker quiver (Reineke)

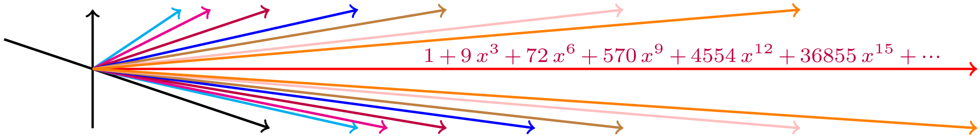


Figure 9: Local scattering diagram with $F(1)$

Theorem 3.8 in van Garrel–Nabijou–Schuler, TAMS (2025)

Proof of Generalised Conjecture

Theorem 3.1 and 3.2 in van Garrel–Nabijou–Schuler, TAMS (2025)

Simple relation between $N_d^{\log}(r)$ and local invariants $N_d^{\text{loc}}(r)$, some $r \geq -1$

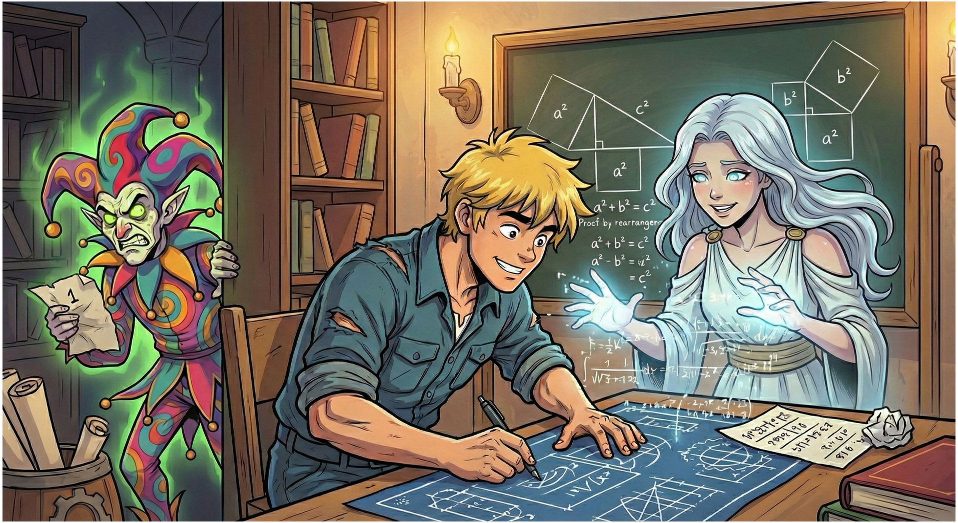
- Consequence of much more general Theorem 1.3 and 2.3 in article

Hypothetical Question

Given access to Deep Think in 2023, could Barrott–Nabijou have linked $N_d(r) \leftrightarrow N_d^{\text{loc}}(r)$, significantly strengthening their results ?

Note: Deep Think never references our paper

Case Study 3 is Worker-Oracle interaction



Conclusions

AI as a Worker, Oracle, Deliriant.

Flawed at logic – getting better

$\rightsquigarrow$ always verify, always need to know what's going on.

$\rightsquigarrow$ Not good as saying "I don't know how to do it".

Great at connections.

It read all of maths after all.

Sub-sub-sub-divide tasks.

Use multi-agent AI.

Prompts, prompts, prompts.

- “We are both research mathematicians”.
- “It is OK to say I don’t know”.
- “For each of your arguments, indicate your level of certainty”.
- “Use concise language, provide rigorous arguments.”
- “Check whether your arguments are rigorous”.
- “You may try the following:”
- “Explore connections to the following list of invariants.”
- Collect arguments and provide elements of it back in a new chat. Repeat.
- Separately ask to prove and to disprove same statement.

Provide input.

May attach .tex files with relevant results and proofs.

Conclusions

Try several times, have a conversation, experiment.

Provide context.

“Context: these are questions in Enumerative Geometry.
Goal: find connections to other invariants.
I am looking for structural results such as functional equations.”

Structure reasoning.

“Preparation: review relevant results.
” Task 1: Carefully read and understand the arguments in the attached .tex file.
” Task 2: Generalise the main theorem in .tex file. I suspect the generalised invariant has the following form:...”
Output: Write a rigorous proof of the generalised theorem.”

To finish: Conjecture generation (based on Schuler, Reineke, ...)

Let $P_n(z,q), F_m(z,q) \in \mathbb{Q}(q)[[z]]$ unique solutions to

$$P_n(z,q) = 1 + (-1)^n z \prod_{k=-n}^{n} P_n(zq^k, q)^{n+1-|k|}$$

$$F_m(z,q) = 1 + z \prod_{k=0}^{m-1} F_m(zq^k, q).$$

In the classical limit $q \to 1$, $P_n(z,1) = F_{(n+1)^2}((-1)^n z, 1)$.

Quantisation Conjecture

$P_n(z,q)$ and $F_{(n+1)^2}((-1)^n z, q)$ are refined invariants for different cohomology theories of joint geometric moduli space.

$\rightsquigarrow$ Deep Think gives fantastic answer, exploring it now.


AI