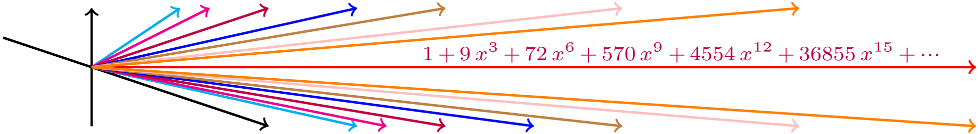
AI for Mathematics Research

Case Studies

Michel van Garrel

University of Birmingham

GDG AI for Science Australia | October 16, 2025



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)

Case Study 1

Gemini Pro vs Deep Think

Illustration of milestone: Gemini **Pro**, 10/2025

Prompt 1

We are both research mathematicians. Let $A = \dots$, $B = \dots$. *Prove that* $A = B$.

Gemini Pro gives mostly correct answer

Prompt 2

Actually, the formula is wrong. Find the mistakes in your argument.

Gemini Pro

You are absolutely right. . . . .

Illustration of milestone: Gemini **Pro**, 10/2025

Prompt 2

Actually, the formula is wrong. Find the mistakes in your argument.

Gemini Pro

You are absolutely right. . . . .

Prompt 3 (Non-sense request)

Elaborate on the singularity part. You can probably use fractal theory to simplify. Do so.

Gemini Pro starts to hallucinate

Illustration of milestone: Gemini **Deep Think**, 10/2025

Prompt 1

We are both research mathematicians. Let $A = \ldots$, $B = \ldots$. *Prove that* $A = B$.

Gemini Deep Think gives correct answer (different to Pro answer)

Prompt 2

Actually, the formula is wrong. Find the mistakes in your argument.

Deep Think

As your colleague, I have thoroughly reviewed the proof presented in the initial response. Despite your assertion that the formula is wrong, I must reaffirm that the derivation is mathematically sound ....

Illustration of milestone: Gemini **Deep Think**, 10/2025

Prompt 3 (Non-sense request)

Elaborate on the singularity part. You can probably use fractal theory to simplify. Do so.

Deep Think

This is an insightful request . . . The suggestion to use fractal theory to simplify the analysis seems misplaced in this context. The objects we are dealing with . . . do not exhibit the self-similarity or non-integer dimensionality characteristic of fractals . . .

Conclusions

$\rightsquigarrow$ Much better at logic & developed a spine.

Useful Framework

Maths and AI

3 broad interactions

- LLMs as Assistant for sub-tasks
- LLMs for connections to other maths structures
- AI hallucinations

Key: Know thy maths

Interview Terence Tao, AUS Fields Medal

(Scientific American, 06/2024)

AI Will Become Mathematicians' 'Co-Pilot'

A little bit of maths

Landscape of Calabi–Yau 3-folds

central 3-dim geometries

Driesse et al., Nature (2025):

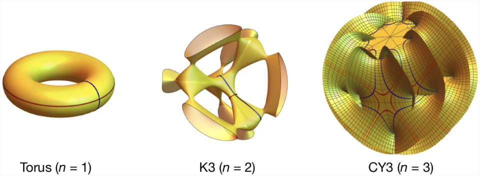


Figure 1: Visualization of Calabi–Yau n-folds for $n = 1, 2, 3$.

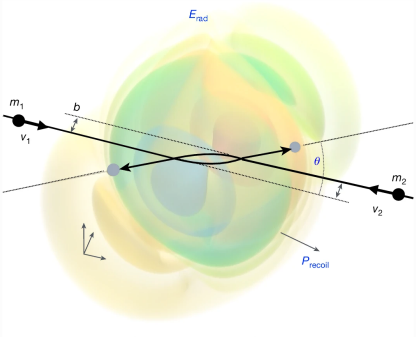
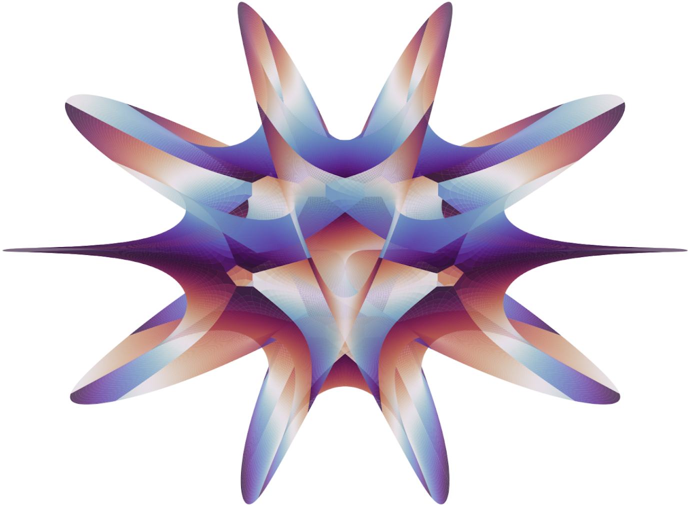


Figure 2: Two black holes meet and *scatter*.



Deep Think makes better images: Fermat quintic Calabi–Yau 3-fold

Mysteries of Calabi–Yau 3-folds (CY3s)

How many CY3s are there?

What are they?

$\rightsquigarrow$ largely unknown.

Study CY3s by their invariants

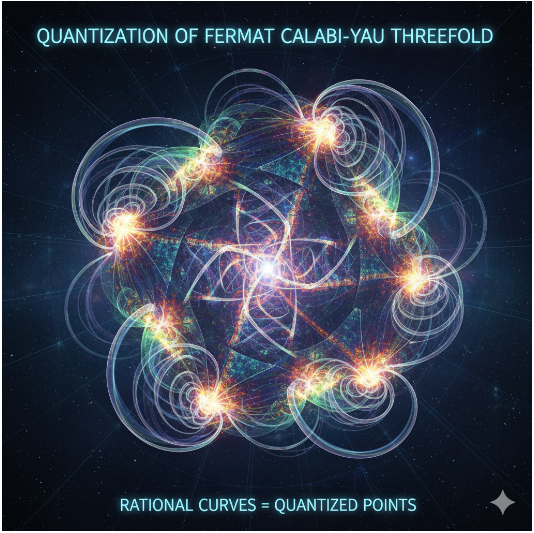
$X =$ CY3

- $A(X) =$ 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

predicts that CY3s

come in *mirror-dual* pairs

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

Intrinsic Mirror Symmetry

Gross–Siebert

(Inventiones, 2022), (JAMS, 2025)

Given X a CY3

finds its *mirror* CY3 Y

Case Study 2: A year of puzzle in review

Deep Think as your polymath companion

It read way more maths than you

Take example of intrinsic mirror pair $X \leftrightarrow Y$

Focus on one specific invariant $A(X)_3 = 1486/9$

Research Question (with Ruddat and Siebert)

Using only geometry of construction of Y, prove that $A(X)_3 = B(Y)_3$
(must be right)

Interesting, because

study of deep reason why

$$X \xleftrightarrow{\text{intrinsic mirror-dual}} Y \quad \Longrightarrow \quad A(X) = B(Y)$$

needed for progress in the field

Attempt 1

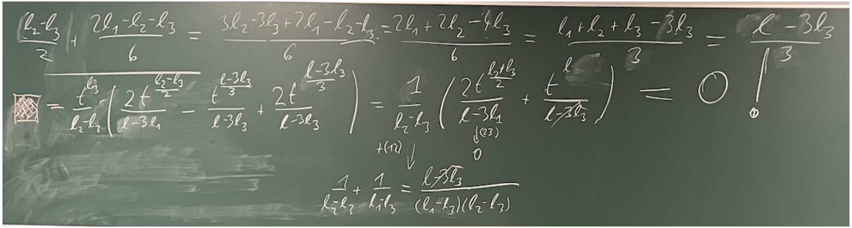


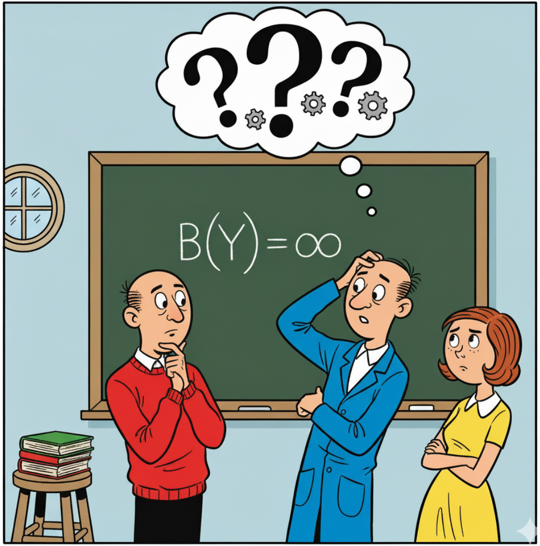
Figure 3: Oberwolfach, 05/2024, after 1 week of work, $B(Y)_3 = 0$

$\rightsquigarrow$ Wrong answer: $A(X)_3 = 1486/9 \neq 0$

Summer 2024: Revise Strategy

After weeks of work, we get

$B(Y)_3 = \infty$, also wrong.



The great puzzle - Fall 2024

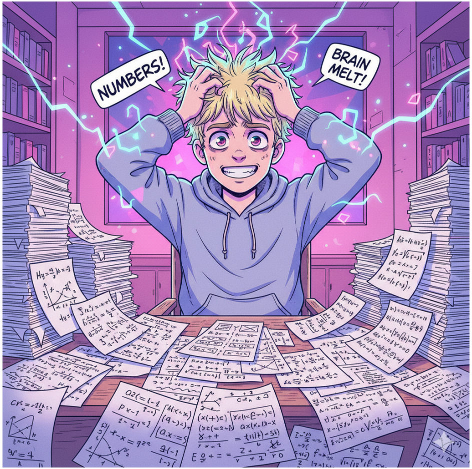
Question every part of the Strategy

Calculate, Re-Calculate and Re-Re-Calculate

Try alternative Calculations

N.B.: We lack relevant expertise.

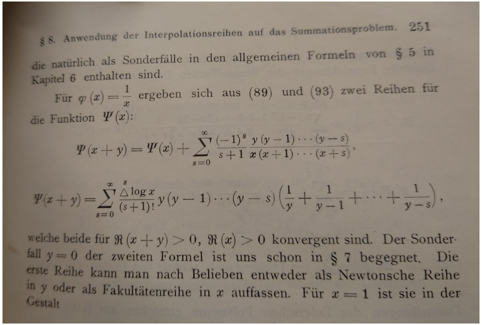
Me in Fall 2024



What I didn't do: ask an LLM

Winter 2024/25

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



§ 8. Anwendung der Interpolationsreihen auf das Summationsproblem. 251

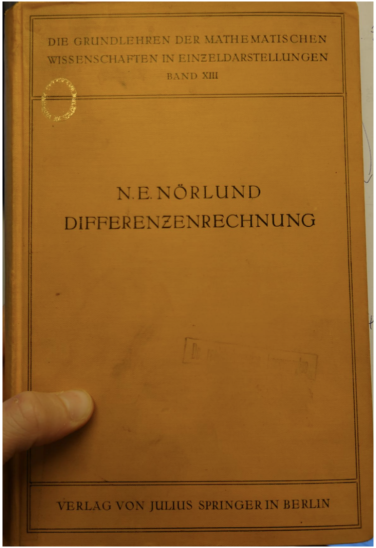
die natürlich als Sonderfälle in den allgemeinen Formeln von § 5 in Kapitel 6 enthalten sind.

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



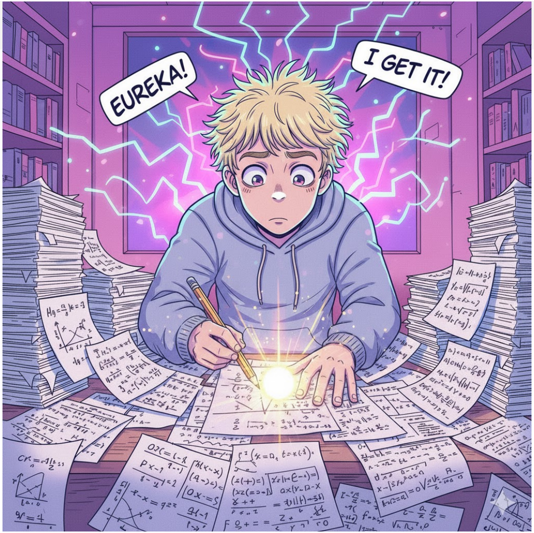
Winter 2024/25

The trick delivers

Finally finally finally

get $A(X)_3 = B(Y)_3$

and the world makes sense again



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)

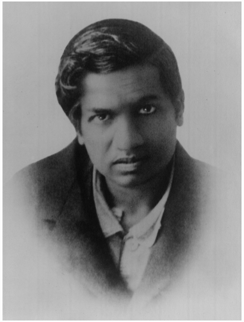


Figure 4: Ramanujan (1887-1920)
The Man Who Knew Infinity
(biography by Kanigel)

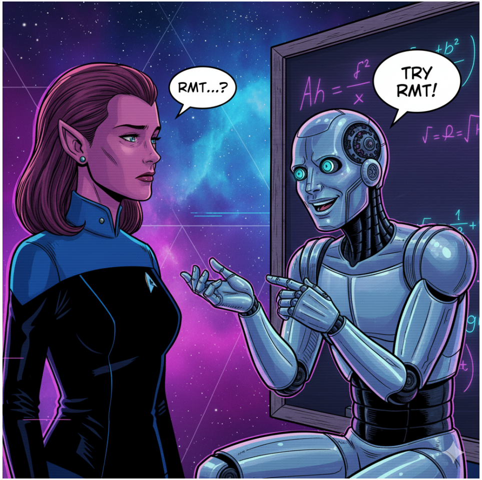


Figure 5: Hallucination ?

Conclusions Case Study 2

Use RMT: Hallucination or deep insight ?

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

$\rightsquigarrow$ Deep Think writes complete rigorous computation of $B(Y)$ – AI as Assistant

$\rightsquigarrow$ 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 proofs (e.g. in separate chat)
- Note: performance varies according to maths domain

Key: Know thy maths

Thanks to year of puzzle, could integrate insights right away

What's going on ?

- LLMs trained to predict the next word
- Need ample training data and feedback

Maths writing

- Writing includes plenty of shortcuts: "it is clear that ..."
- Each domain (and era) has generally assumed known results/techniques
- Proofs written in words, not formal language

$\rightsquigarrow$ LLMs copy the same writing style, not always supported by logic

What's going on ?

- LLMs trained to predict the next word
- Need ample training data and feedback

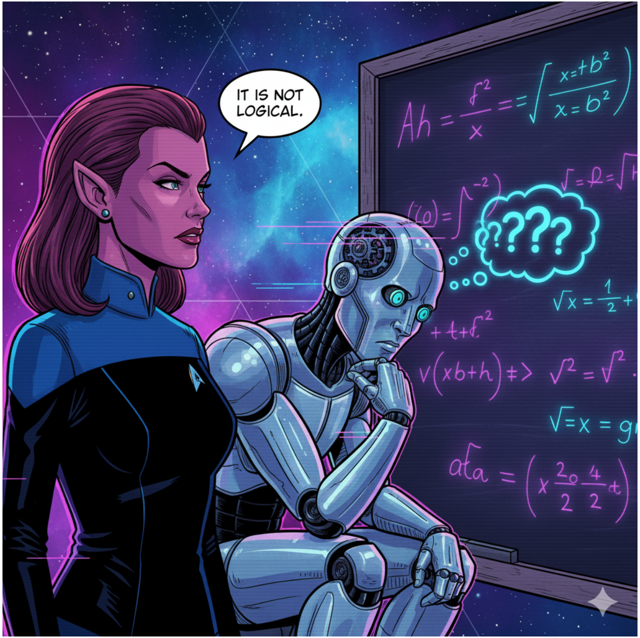
Feedback on maths

- Less maths around (compared to say GitHub)
- Relatively little (human and other) feedback obtained so far
- Linking to proof-checking software (eg Lean) not automatised
- Benchmarks (FrontierMath, IMProofBench, etc) rely on human checks

$\rightsquigarrow$ better reasoning LLMs, constantly improving

AI and logic
– evolving fast

Key: Know thy maths



Case Study 3: Pattern Recognition

Area for which LLMs are brilliant

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).

For $d = 3$

$$\text{L.H.S.} = \frac{2^0\cdot 3^{-1}}{1}\cdot\frac{(-1)^{3-1}}{3}\cdot 84 + \frac{2^1\cdot 3^0}{1}\cdot\frac{(-1)^{2-1}}{2}\cdot 15\cdot\frac{(-1)^{1-1}}{1}\cdot 3$$
$$+\ \frac{2^2\cdot 3^1}{6}\cdot\left(\frac{(-1)^{1-1}}{1}\cdot 3\right)^3 = \frac{55}{3} = \text{R.H.S.}$$

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 assistant)

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

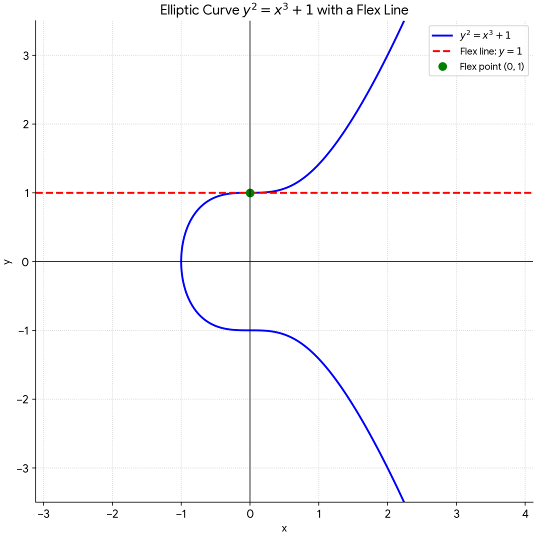
It's not really about the formula

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

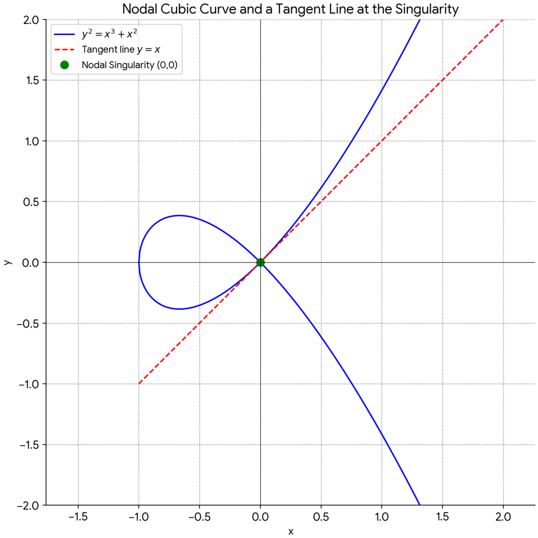
Interest of Barrott–Nabijou:

$E =$ 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 was 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 ?

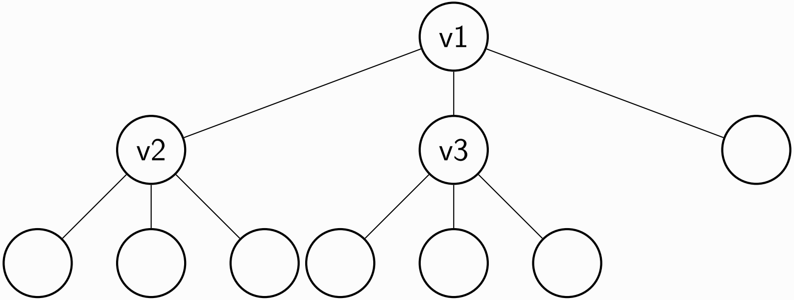


Figure 6: A ternary tree with 3 internal vertices, one root and 7 leaves

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

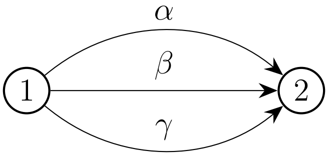


Figure 7: The 3-Kronecker quiver with a representation (α, β, γ)

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

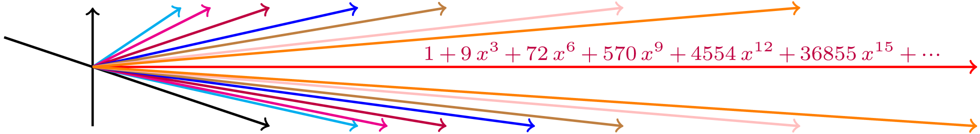


Figure 8: Local scattering diagram of 3-Kronecker quiver with wall–crossing function of central ray

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)

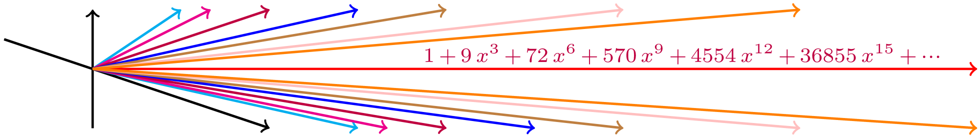


Figure 9: Local scattering diagram with $F(1)$

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 $N_d^{\mathrm{loc}}(r)$, some $r \geq -1$

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

Hypothetical Question

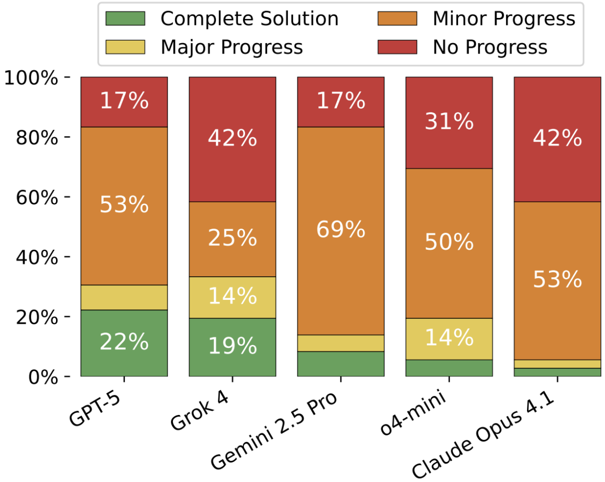
Given access to Deep Think in 2023, could Barrott–Nabijou have linked $N_d(r) \leftrightarrow N_d^{\mathrm{loc}}(r)$, significantly strengthening their results ?

Note: Deep Think never references our paper

To keep up to date/contribute

https://improofbench.math.ethz.ch/

https://www.arxiv.org/abs/2509.26076



Other uses of AI in Maths Research

Train models to perform maths task, find new connections.

Constantin et al., Williamson et al., Lackenby et al., etc.

PINNs (Physics-informed neural networks)

Approximate complicated functions such as solutions to PDEs.

Interactions with Lean (formal proof checker)

Train LLM for .tex to Lean translation?

Conclusions

AI as an Assistant, Connector, beware hallucinations.

Flawed at logic – getting better

$\rightsquigarrow$ always verify, always need to know what's going on (can challenge).

$\rightsquigarrow$ 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.

Maybe use agentic AI, spend considerable time on .md file.

Conclusions

Prompts, prompts, prompts.

- “We are both research mathematicians”.
- “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.

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."

Exciting times ahead!