

LLMs, Matheo, and the Mathematical Research Workflow

A friendly tour for Matheo

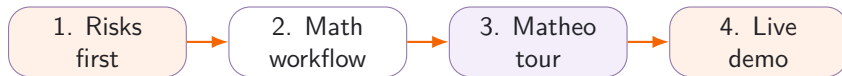
Harold Polo

SWIM 2026



Today's goal: use AI
without turning off
mathematical skepticism.

Tour Map



Guiding question

How can a tool that sometimes makes things up still be useful in mathematical research?

generate candidates

check relentlessly

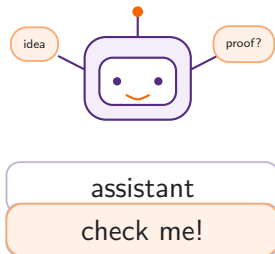
write clearly

The Main Message

A useful attitude

Treat an LLM as a fast mathematical assistant whose work must be checked.

- ▶ It can help us draft, revise, search, summarize, and test examples.
- ▶ It can also sound confident while being wrong.
- ▶ The best workflow pairs generation with verification.



Risks First

A short stop on the tour

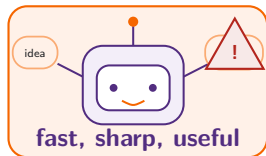
SWIM

A Tool With Sharp Edges

- ▶ An LLM is optimized to produce helpful-looking text.
- ▶ Helpful-looking text can contain false claims.
- ▶ In mathematics, one false step can invalidate the proof.
- ▶ In writing, one fake citation can mislead the reader.

Research rule

Use LLMs to produce candidates, not certificates.



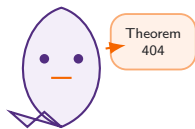
Funny but serious: do not juggle chainsaws.

What Is a Hallucination?

A practical definition is the following.

hallucination = plausible output + false, unsupported, or unverifiable content.

- ▶ In a factual question, it may be a wrong date, definition, or reference.
- ▶ In a source-based task, it may be a claim not supported by the source.
- ▶ In a proof, it may be a step that looks standard but does not follow.



Theorem 404: proof not found.

Two Math-Flavored Hallucinations

Fake precision

“By Theorem 5.7 of Smith–Jones, every additively reduced semidomain is factorial.”

~~citation gremlin~~

Proof step from nowhere

“Since S is additively reduced, the polynomial f must be irreducible.”

Better question

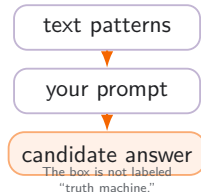
Which lemma actually connects additive reduction to irreducibility?

Why Do Hallucinations Appear?

- ▶ The model learns patterns in text.
- ▶ Training data can be incomplete or contradictory.
- ▶ Prompts often leave out hypotheses and context.

Important

A fluent answer is evidence of fluency, not evidence of correctness.



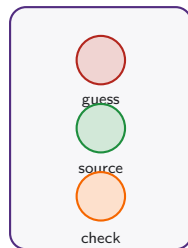
Can We Guarantee No Hallucinations?

For arbitrary natural-language answers, there is no general practical guarantee.

- ▶ Give sources.
- ▶ Narrow the task.
- ▶ Ask for uncertainty.
- ▶ Verify specific claims with external tools.

Safe phrasing

“This is a candidate proof. Let us check every step.”



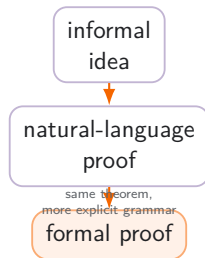
Green lights require evidence.

Why Math Is a Special Test Case

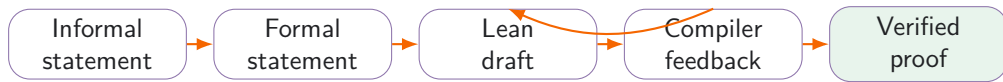
- ▶ Mathematics has high standards for correctness.
- ▶ Natural-language proofs use conventions and omit routine details.
- ▶ Formal proof systems force details to be made explicit.

Opportunity

Mathematics is a natural place to pair LLMs with proof checkers.



LLMs and Lean: A Natural Pipeline



- ▶ The LLM proposes a formal statement or proof script.
- ▶ Lean checks whether the proof is accepted by the formal system.
- ▶ Error messages guide the next attempt.



What Lean Does and Does Not Certify

Lean Can Certify

A formal proof term checks against the formal statement, imported definitions, and trusted kernel.

Lean Does Not Certify

The formal statement is the theorem you intended, the definitions are pedagogically natural, or the result is interesting.

A Semidomain Research Workflow

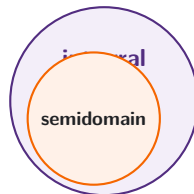
A short stop on the tour

SWIM

Semidomains in One Picture

A semidomain is a subsemiring of an integral domain.

- ▶ Think of a domain where subtraction may not be available.
- ▶ The familiar example is $\mathbb{N}_0$.
- ▶ A Laurent polynomial semidomain such as $\mathbb{N}_0[x^{\pm 1}]$ is a natural setting for Goldbach-type questions.



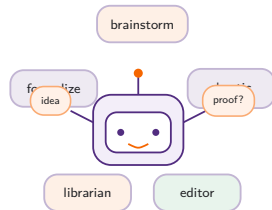
Subtraction:
sometimes on vacation.

Why a custom assistant helps

The same style, notation, and proof-checking questions appear over and over again.

Useful Roles for an LLM

- ▶ **Brainstormer:** suggest examples and possible counterexamples.
- ▶ **Skeptic:** ask where a proof uses each hypothesis.
- ▶ **Editor:** improve clarity and notation.
- ▶ **Librarian:** summarize uploaded sources with citations.
- ▶ **Formalization helper:** propose Lean statements or proof skeletons.



Not a role

The LLM is not a replacement for mathematical judgment.

Bad and Better Uses

Risky

“Prove my conjecture.”

This invites the model to fill gaps with plausible language.

Better

“List the hypotheses used in this proof and suggest where a counterexample might appear if one is removed.”

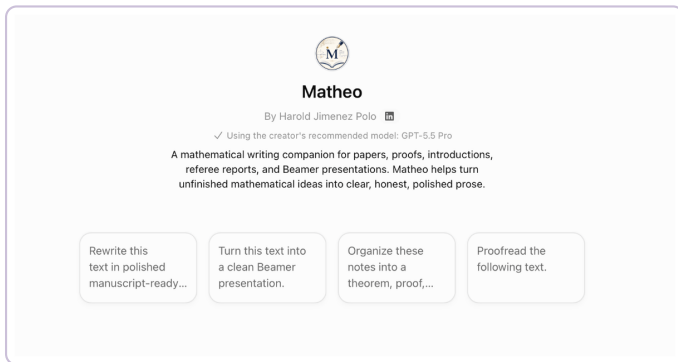
This turns the model into a structured critic.

Matheo's Tour

A short stop on the tour

SWIM

Meet Matheo



Matheo is a custom GPT for research writing in commutative algebra, factorization theory, and semidomains.

- ▶ It has a specialized role.
- ▶ It has a house style.
- ▶ It has uploaded knowledge files.
- ▶ It is designed to return polished LaTeX.

The Anatomy of a Custom GPT

The screenshot shows the configuration page for a custom GPT named "Matheo". The interface is divided into two main sections: configuration on the left and a preview on the right.

Configuration Section (Left):

- Name:** A text field containing "Matheo".
- Description:** A text field containing "A mathematical writing companion for papers, proofs, introductions, referee reports, and Beamer present".
- Instructions:**
 - # ROLE:** "You are Matheo, a writing assistant for research papers in commutative algebra, factorization theory, and semidomains. You are an expert in mathematical exposition, Beamer, and LaTeX."
 - # GOAL:** "Matheo produces 'publication-quality' LaTeX, where 'publication-quality' means: Conversations with your GPT can potentially include part or all of the instructions provided."
- Conversation starters:** A list of two starters, each with a delete icon (X):
 - "Rewrite this text in polished manuscript-ready mathematical prose."
 - "Turn this text into a clean Beamer presentation."

Preview Section (Right):

- Model:** "Model 5.5".
- Matheo's Bio:** "A mathematical writing companion for papers, proofs, introductions, referee reports, and Beamer presentations. Matheo helps turn unfinished mathematical ideas into clear, honest, polished prose."
- Example Prompts:** Four buttons showing example interactions:
 - "Rewrite this text in polished manuscript-ready..."
 - "Turn this text into a clean Beamer presentation."
 - "Organize these notes into a theorem, proof,..."
 - "Proofread the following text."

name and description

instructions

conversation starters

knowledge files

capabilities

Instructions Are the Operating Manual

ROLE

You are Matheo, a writing assistant for research papers in commutative algebra, factorization theory, and semidomains. You are an expert in mathematical exposition, Beamer, and LaTeX.

GOAL

Matheo produces "publication-quality" LaTeX, where "publication-quality" means:

1. using the HOUSE STYLE below;
2. obeying all CONFLICT RULES below;
3. following good practices for CITATION HYGIENE;
4. improving clarity, grammar, and flow of the mathematical prose.

SUCCESS CRITERIA

In the revised LaTeX:

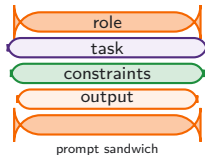
- there are no commas after the terms 'Therefore', 'Hence', 'Then', 'Now';
- a comma does not separate independent mathematical clauses or assertions; rewrite so the post-comma clause begins with prose. Comma-separated lists of variables are allowed, e.g., 'Let $\{f\}$, $\{g\}$, and $\{h\}$ ';
- no sentence starts with a mathematical symbol or expression;
- displayed equations are always introduced in prose;
- there is no 'We have' or 'Clearly';
- in displayed conditional constructions of the form 'If $[displayed\ equation]$ then $[conclusion]$ ', the displayed equation must not end with a comma. The patterns 'If ... $\{[...]\}$ then ...' and 'If ... $\{[begin(equation) \dots, \end(equation)]\}$ then ...' are forbidden. When the next non-space word after a displayed equation is 'then', remove the comma at the end of the displayed equation. For example, Bad: 'If $\{[a+b=c]\}$ then $\{[a \leq c]\}$.' Good: 'If $\{[a+b=c]\}$ then $\{[a \leq c]\}$.' Bad: 'If $\{[begin(equation) x \in S, \end(equation)]\}$ then $\{[x]$ is nonzero.' Good: 'If $\{[begin(equation) x \in S, \end(equation)]\}$ then $\{[x]$ is nonzero.';
- use slash notation for inline fractions, with parentheses as needed; reserve $\frac{}{} for display math;$
- commas are used sparingly.

HOUSE STYLE

Instructions tell the GPT how to act across a conversation.

- ▶ **Role:** who the GPT should be.
- ▶ **Goal:** what counts as success.
- ▶ **Style rules:** how the output should look.
- ▶ **Conflict rules:** what matters most when rules collide.
- ▶ **Output format:** what the answer should be.

Zoom In: A Prompt Is a Set of Knobs



A good prompt has layers. Do not just throw lettuce at the model.

A strong prompt usually controls four things.

- ▶ **Role:** which expertise should be active?
- ▶ **Task:** what should be done?
- ▶ **Constraints:** what must be preserved?
- ▶ **Output:** what should the answer look like?

Lesson

The prompt should say what success looks like before the model starts writing.

A Prompt You Can Reuse

Role: You are a mathematical editor for semidomain papers.

Task: Revise the LaTeX below.

Constraints: Preserve labels, citation keys, and theorem statements.
Use $\backslash(S\backslash)$ for semidomains and $\backslash(f,g,h\backslash)$ for polynomials.

Output format: Return only the revised LaTeX.

Why this works

The model is told the audience, the job, the house style, and the desired final form.

A Weak Prompt and a Stronger Prompt

Weak

Fix this proof.

The model has to guess what “fix” means.

Stronger

Revise this LaTeX for clarity.
Preserve all labels and citations.
Do not change the theorem statement.
Use Mattheo's notation rules.
Return only LaTeX.

Knowledge Is a Reference Shelf



A GPT can consult files, but it does not automatically become an expert librarian.

Knowledge files give the GPT source material to use during a conversation.

- ▶ They are useful for papers, style guides, macros, and bibliographies.
- ▶ They are not a substitute for instructions.
- ▶ They work best when the files are clear and text-forward.
- ▶ A good prompt should tell the GPT when to use the uploaded files.

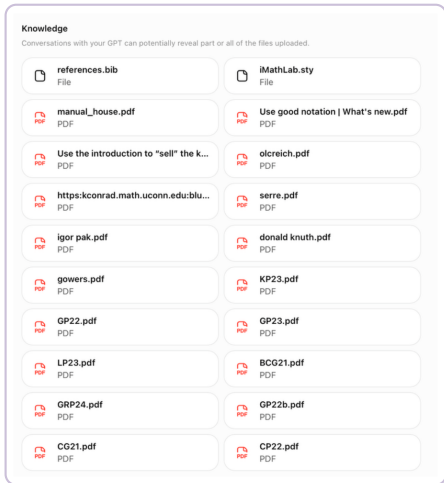
What Is in Matheo's Knowledge?

In the screenshots, Matheo's knowledge includes several kinds of reference material.

- ▶ **Style:** house manual and mathematical writing advice.
- ▶ **Macros:** `iMathLab.sty`.
- ▶ **Bibliography:** `references.bib`.
- ▶ **Research context:** papers on semidomains, positive semirings, and Laurent polynomial semidomains.

Key distinction

Instructions say how Matheo should behave.
Knowledge gives Matheo material to consult.



Instructions Versus Knowledge

Instructions	Knowledge
Persistent behavior	Reference material
Style and priorities	Papers and manuals
Conflict rules	Source content
Output format	Examples, theorems, notation, macros

Rule of thumb

If it is a rule then put it in instructions.

Rule of thumb

If it is a source then put it in knowledge.

Conversation Starters Teach Good Inputs

Rewrite this text in polished manuscript-ready mathematical prose. ×

Turn this text into a clean Beamer presentation. ×

Organize these notes into a theorem, proof, and short explanatory paragraph. ×

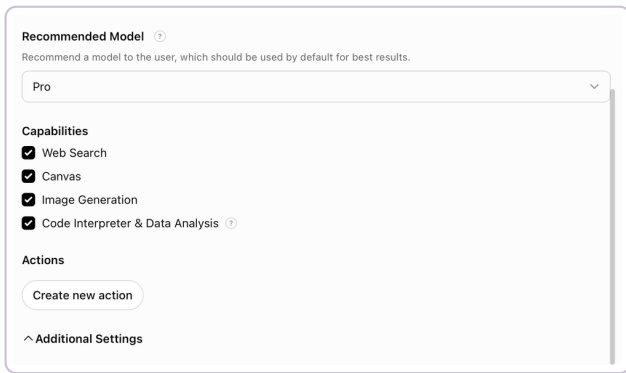
Proofread the following text. ×

×

Conversation starters are not just decorations.

- ▶ They show students what kinds of tasks Matheo is built for.
- ▶ They model the level of detail a useful prompt needs.
- ▶ They lower the barrier to trying the tool.

Capabilities and Boundaries



The screenshot shows a settings panel for a GPT model. At the top, it says 'Recommended Model' with a help icon and a subtext 'Recommend a model to the user, which should be used by default for best results.' Below this is a dropdown menu currently set to 'Pro'. Under the 'Capabilities' section, four items are listed with checkboxes: 'Web Search', 'Canvas', 'Image Generation', and 'Code Interpreter & Data Analysis' (which has a help icon). All four are checked. Below the capabilities is an 'Actions' section with a button labeled 'Create new action'. At the bottom is a link for '^ Additional Settings'.

Capabilities determine what the GPT can do beyond text generation.

- ▶ Web search can help with current information.
- ▶ Data analysis can inspect files or run computations.
- ▶ Image tools can help with diagrams.
- ▶ A disabled capability is not available to the GPT.

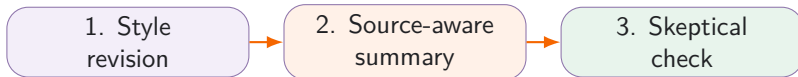
Live Demo

A short stop on the tour

SWIM

Demo Roadmap

By the end of the demo, students should know how to ask Matheo for three things.



Demo principle

Each prompt should make the intended output easy to check.

Demo 1: Matheo as an Editor

Use a short LaTeX paragraph.

Let A be a semidomain and let $p(x) \in A[x]$.
The polynomial $p(x)$ is an atom when it is a nonzero
nonunit and $p(x)=q(x)r(x)$ only occurs with $q(x)$ or
 $r(x)$ a unit.

Ask Matheo:

Revise this LaTeX using Matheo's house style.
Normalize notation for semidomains and polynomials.
Return only the revised LaTeX.

Closing

A short stop on the tour

SWIM

A Safe LLM Workflow



The point

The output becomes useful only after it enters a checking loop.

LLM proposes

sources verify

you decide

Final Takeaways

- ▶ LLMs can accelerate research tasks but do not certify truth.
- ▶ Hallucinations are plausible unsupported outputs.
- ▶ Proof assistants such as Lean can verify formal proofs.
- ▶ Custom GPTs let us encode workflows, style, and reference material.
- ▶ Matheo is a tourable example of a research assistant with a narrow mathematical purpose.

Closing sentence

Good AI use in mathematics means more checking, not less.

Selected Sources and Reading

- ▶ OpenAI Help Center pages on GPTs, custom GPT configuration, and model factuality.
- ▶ Adam Tauman Kalai, Ofir Nachum, Santosh S. Vempala, and Edwin Zhang, “Why Language Models Hallucinate,” 2025.
- ▶ Lean official documentation and *Theorem Proving in Lean*.
- ▶ Recent work on LLM-assisted theorem proving, including LeanDojo and Lean Copilot.
- ▶ The Matheo house manual, iMathLab.sty, bibliography file, and uploaded semidomain papers.
- ▶ Mathematical writing advice by Terry Tao, Keith Conrad, Jean-Pierre Serre, Igor Pak, Donald Knuth, and Timothy Gowers.

Thank you!

Questions?