The Agentic Researcher: Accelerating Optimization and Learning Research with AI

Sebastian Pokutta

Technische Universität Berlin
and
Zuse Institute Berlin

pokutta@math.tu-berlin.de
@spokutta

INRIA

July 22nd, 2026 · Online (AI4Science seminar)

What is this talk about?

Introduction

A personal, highly-biased, and incomplete take of what AI can do in Mathematics et al.

What is this talk about?

Introduction

A personal, highly-biased, and incomplete take of what AI can do in Mathematics et al.

Why? AI methods increasingly shape how scientific discovery is performed, configured, and accelerated.

What is this talk about?

Introduction

A personal, highly-biased, and incomplete take of what AI can do in Mathematics et al.

Why? AI methods increasingly shape how scientific discovery is performed, configured, and accelerated.

Today: Brief Overview + Tech Demo

What is this talk about?

Introduction

A personal, highly-biased, and incomplete take of what AI can do in Mathematics et al.

Why? AI methods increasingly shape how scientific discovery is performed, configured, and accelerated.

Today: Brief Overview + Tech Demo

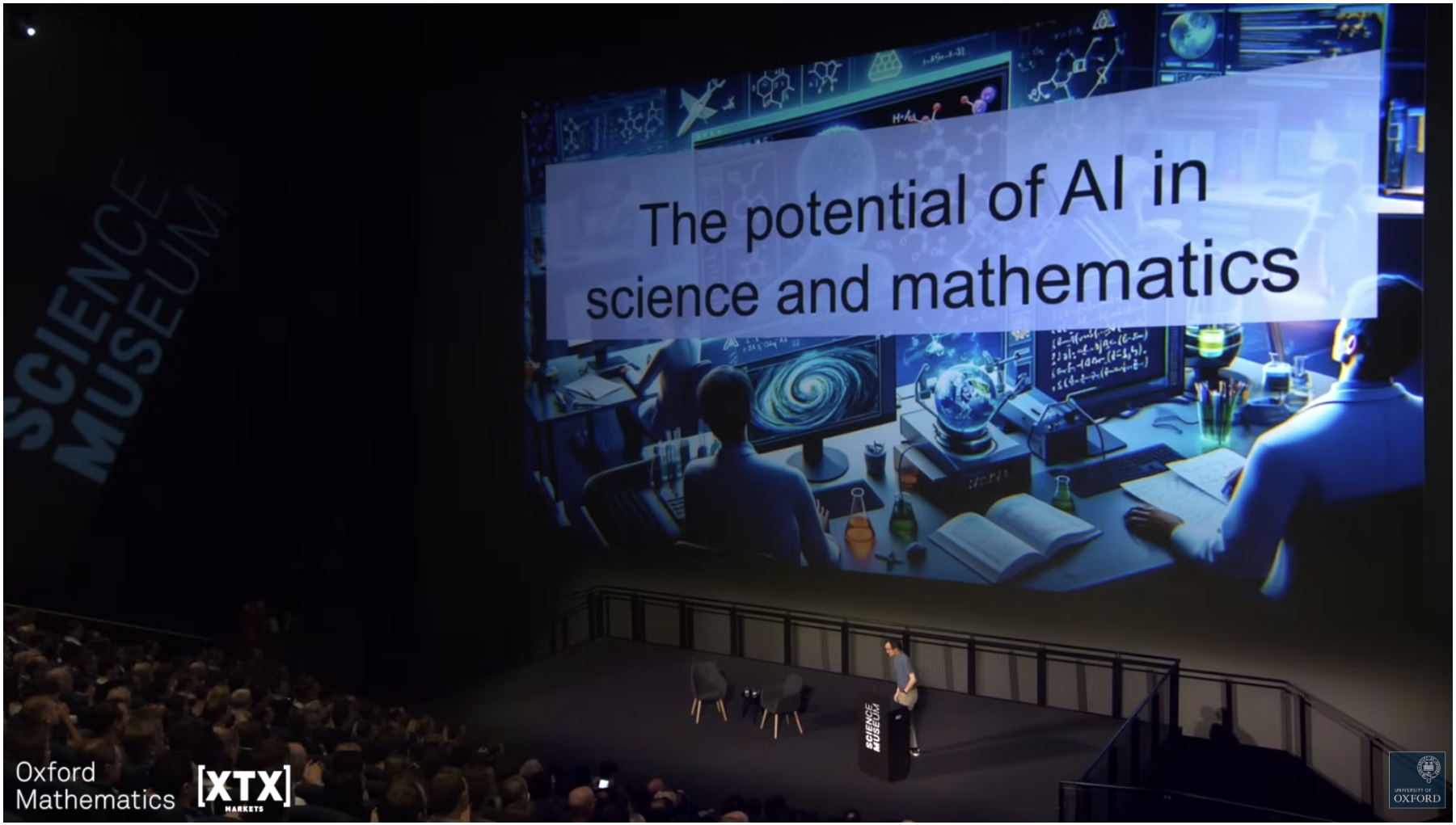
Outline

- A bit of high-level perspective
- Agentic AI Researcher framework
- Live Tech Demo

(Hyperlinked) References are not exhaustive; check references contained therein.

What is this talk about?

AI and Mathematics et al



[The Potential for AI in Science and Mathematics - Terence Tao]

What is this talk about?

AI and Mathematics et al

Various levels of co-creation. [Haase and Pokutta, 2026]

- Digital Pen: basically like autocorrect, bibtex lookup, etc. “2000s”
- AI Task Specialist: ChatGPT, Claude, Gemini, etc. 2022 - 2025
- AI Assistant: Agents with integrated tools, verification, etc. 2025 -
- AI Co-creator: Fully integrated, autonomous, co-creator 2027(??) -

What is this talk about?

AI and Mathematics et al

Various levels of co-creation. [Haase and Pokutta, 2026]

- Digital Pen: basically like autocorrect, bibtex lookup, etc. “2000s”
- AI Task Specialist: ChatGPT, Claude, Gemini, etc. 2022 - 2025
- AI Assistant: Agents with integrated tools, verification, etc. 2025 -
- AI Co-creator: Fully integrated, autonomous, co-creator 2027(??) -

Broadly. Two categories: (a) co-creation agent and (b) tool in larger system.

What is this talk about?

AI and Mathematics et al

Various levels of co-creation. [Haase and Pokutta, 2026]

- Digital Pen: basically like autocorrect, bibtex lookup, etc. “2000s”
- AI Task Specialist: ChatGPT, Claude, Gemini, etc. 2022 - 2025
- AI Assistant: Agents with integrated tools, verification, etc. 2025 -
- AI Co-creator: Fully integrated, autonomous, co-creator 2027(??) -

Broadly. Two categories: (a) co-creation agent and (b) tool in larger system.

Capabilities are impressive but unstable:

- SOTA models achieve post-PhD level scores on benchmarks, yet in day-to-day use make trivial (logical) errors.
- No hard verification of results and randomness across runs.
- Prompting and scaffolding are still a challenge.
- Availability of tools for verification etc crucial.

What is this talk about?

Mathematics with computers is not new

Various high-profile examples from the past.

- Four Color Theorem: massive computer-based case checking

[Appel and Haken, 1977, Robertson et al., 1997]

- Kepler Conjecture / Hales' Theorem: extensive computer verification

[Hales et al., 2017]

- Classification of Finite Simple Groups: Formal verification with Lean/Coq
- Boolean Pythagorean Triples Problem: A spectacular 200TB SAT-solver proof.

[Heule et al., 2016]

What is this talk about?

Mathematics with computers is not new

Various high-profile examples from the past.

- Four Color Theorem: massive computer-based case checking

[Appel and Haken, 1977, Robertson et al., 1997]

- Kepler Conjecture / Hales' Theorem: extensive computer verification

[Hales et al., 2017]

- Classification of Finite Simple Groups: Formal verification with Lean/Coq
- Boolean Pythagorean Triples Problem: A spectacular 200TB SAT-solver proof.

[Heule et al., 2016]

Crucial role in computational mathematics / scientific computing

- Finite Elements
- Numerical Simulations
- Optimization
- Engineering
- ...

The Agentic Researcher

A Practical Guide to AI-Assisted Research in Mathematics and Machine Learning

joint work with: Max Zimmer, Nico Pelleriti, Christophe Roux

ICML AI4RESEARCH Workshop 2026 (oral)
https://arxiv.org/abs/2603.15914

Partially supported by ExC MATH+ Project EF-LiOpt-3

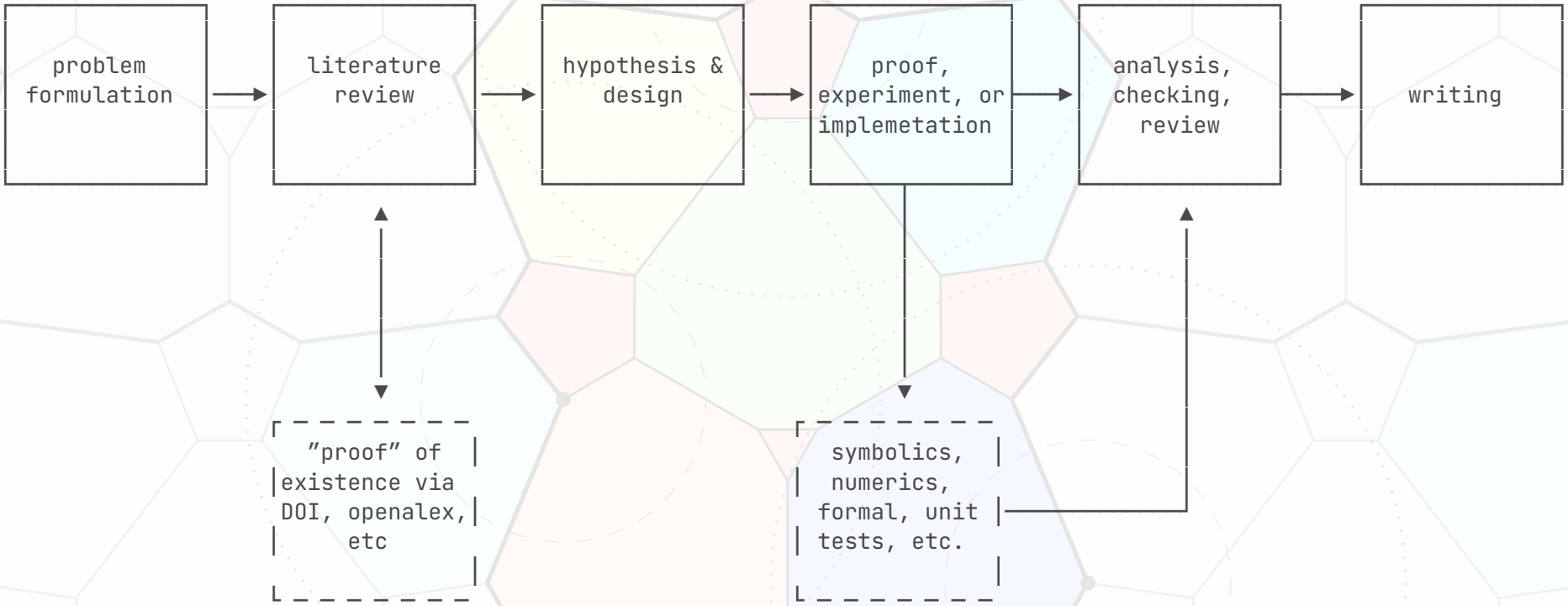
Agent AI in Mathematics

[Zimmer et al., 2026]

AI in the Scientific Discovery Workflow

The agentic researcher framework

AI systems can be used at almost any stage of the scientific workflow.

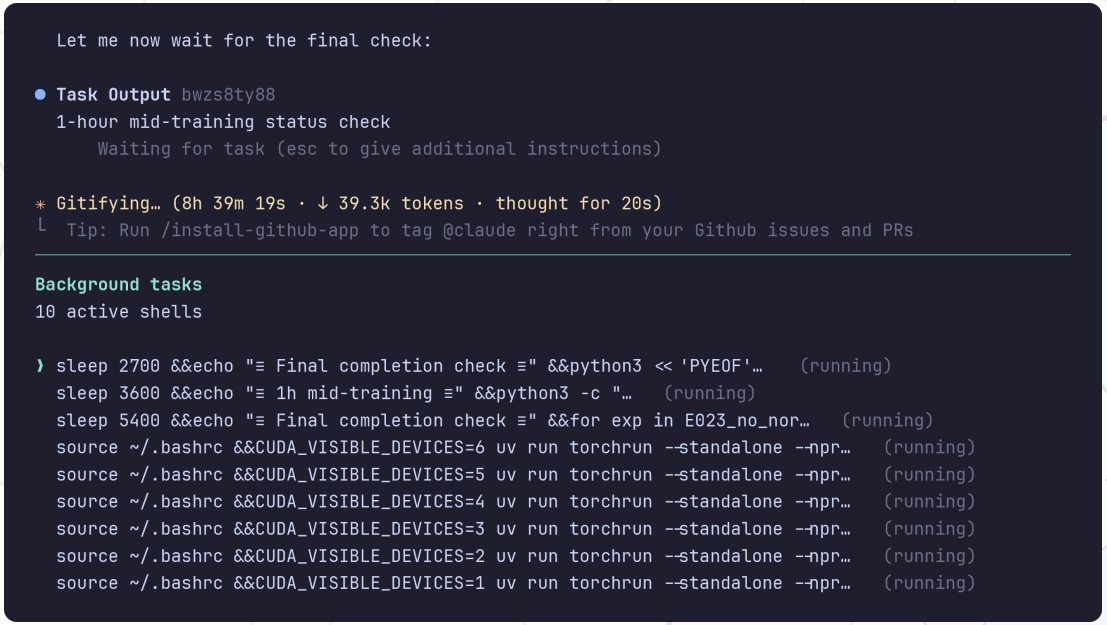


(workflow simplified. for illustration only.)

Note. The researcher needs to be in charge and responsible(!!) throughout.

Building a practical tool

The agentic researcher framework



Screenshot from a run of the framework.

- Born out of the MATH+ project *Agentic AI in Mathematics*.
- Refined over roughly 1.5 years of day-to-day research use.
- Sandboxed CLI agents with project instructions, tools, Git, LaTeX, and GPU-backed experiments.
- Puts a premium on verification: record everything, verify citations, and verify before claiming.

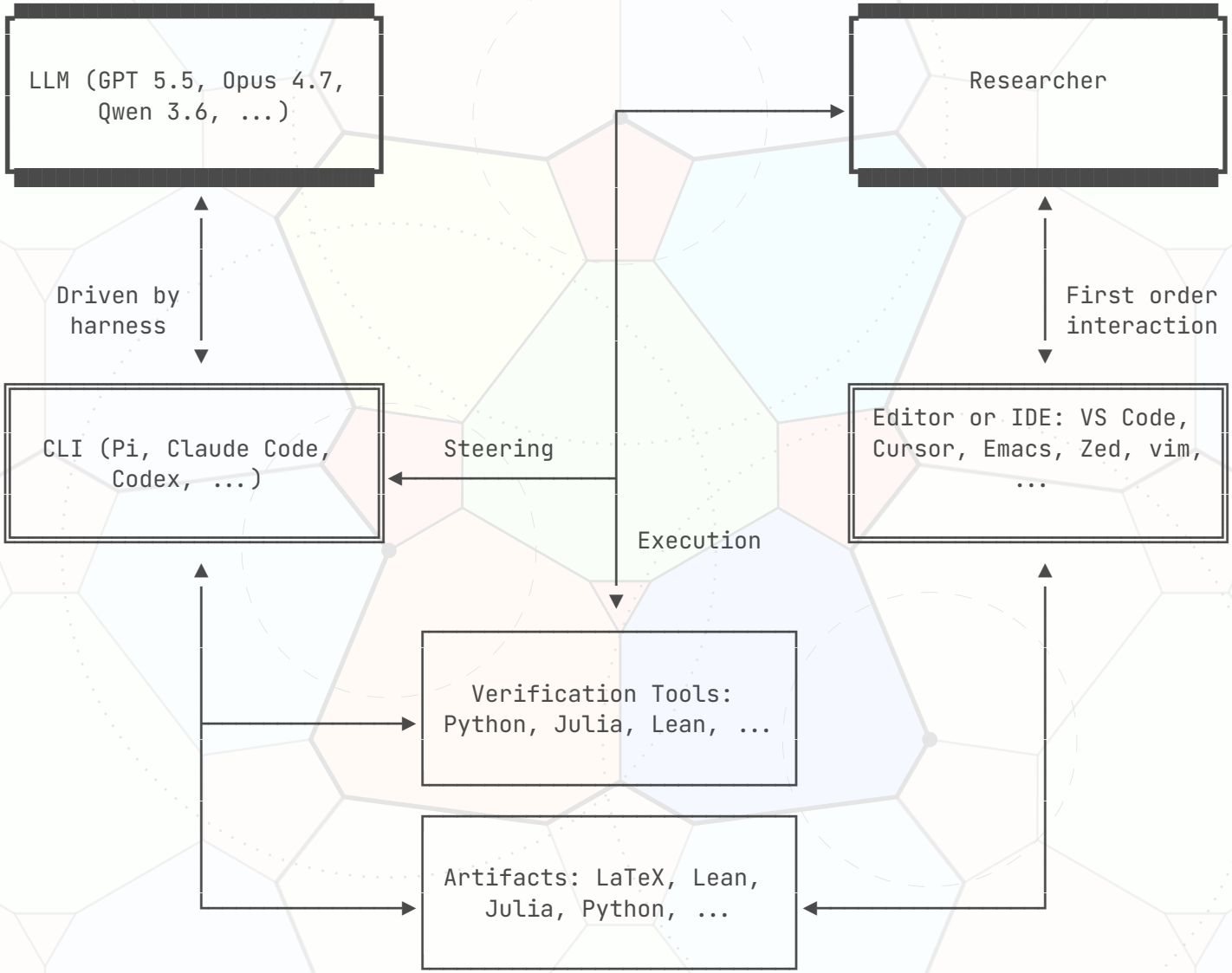
[blog post] [arXiv] [GitHub]

Does it works? Several real-world use-cases in paper

1. Convergence lower bounds for Frank-Wolfe on uniformly convex sets
2. Multi-Variable Dual Tightening for Boscia.jl for MINLPs
3. Weight Reconstruction in LLM pruning
4. ...

Interacting with the System

The agentic researcher framework



Frank-Wolfe Beyond 1/t Convergence

preprint

https://arxiv.org/abs/2604.28006

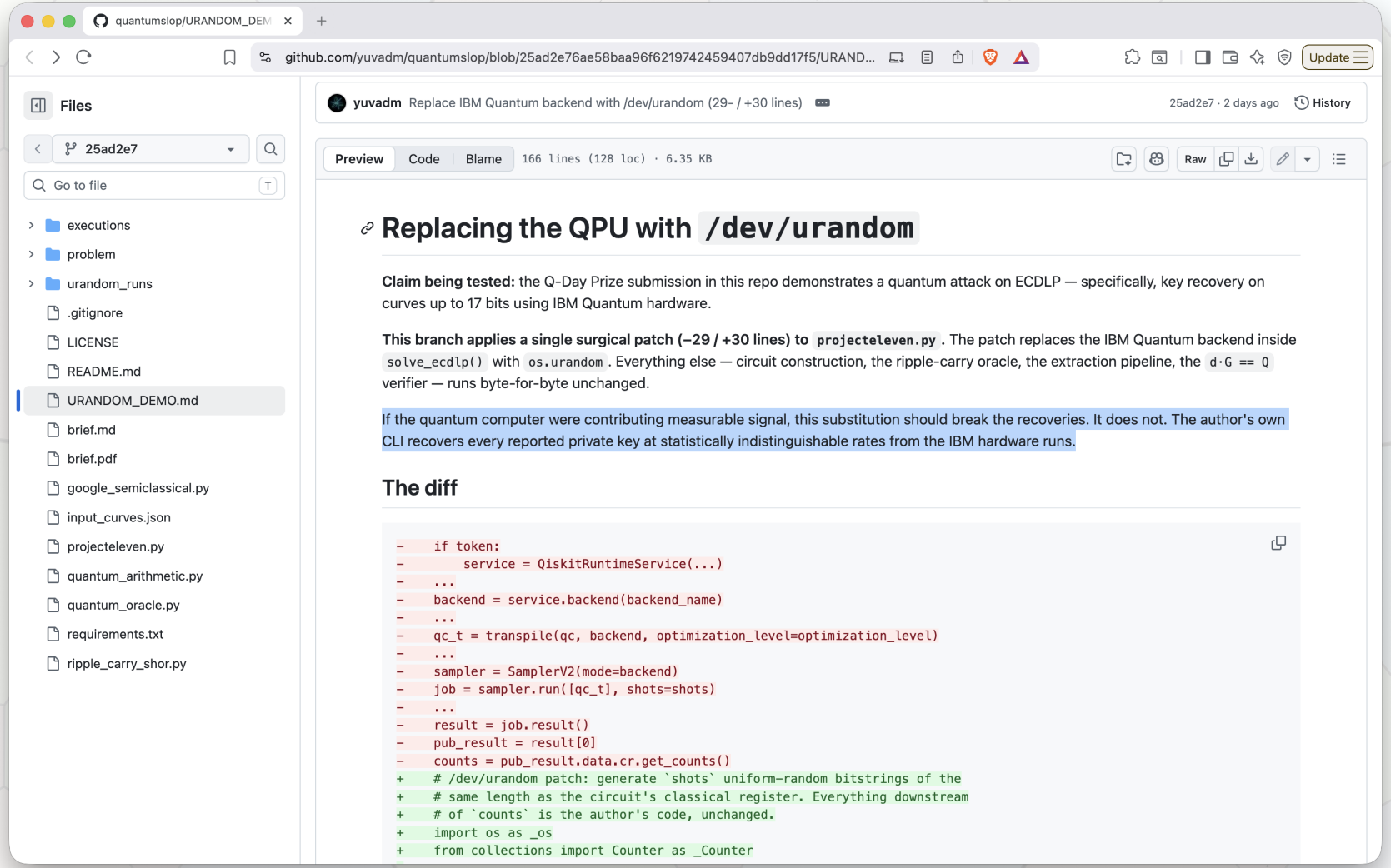
Partially supported by ExC MATH+ Project EF-LiOpt-3

Agent AI in Mathematics

[Pokutta, 2026]

Why Verification is Crucial

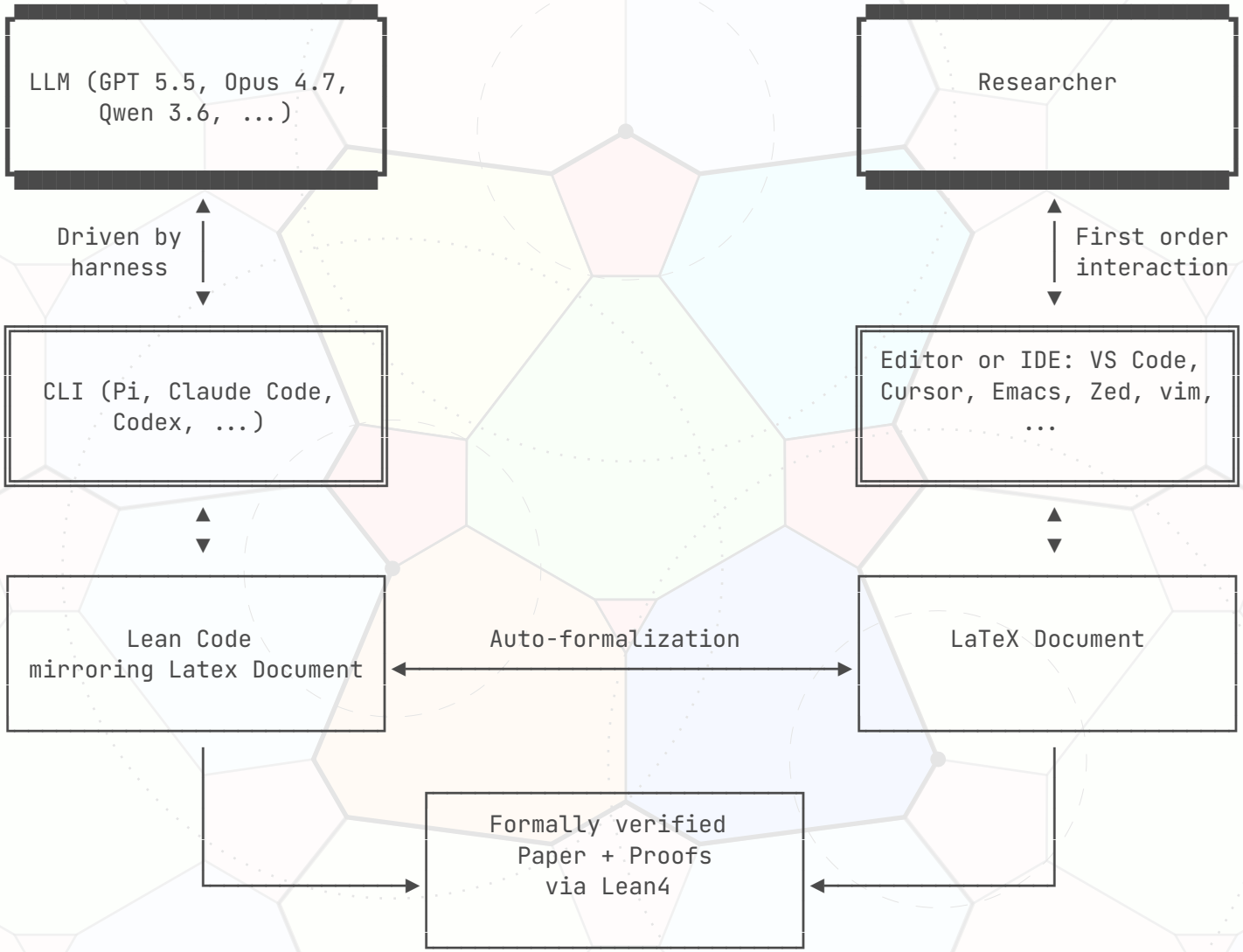
The agentic researcher framework



[GitHub Repository]

End-2-End Formal Verification with Lean 4

The agentic researcher framework



When generation is cheap, verification is everything.

[Pokutta, 2026]

Final Remarks

1. AI can be used in various ways in modern mathematical research workflows (actual discovery, verification, etc); beyond simple black-box prompting
2. Fully-automatic discovery of new mathematics might be possible in the future but relies on *strong* verification approaches
3. The agentic harness seems to be key: how does the agent receive feedback on its work, how is it guided, and which tools are available?
4. Empirically: the human-in-the-loop is crucial to guide the search

The promises of AI4MATH are great but need to go beyond simple black-box prompting “your favorite Erdős problem” (which then turns out having a solution that is already known) ...

Thank you!

Zimmer, M., Pelleriti, N., Roux, C., and Pokutta, S. (2026). The Agentic Researcher: A Practical Guide to AI-Assisted Research in Mathematics and Machine Learning. Preprint. Available at https://arxiv.org/abs/2603.15914.

Pokutta, S. (2026). Frank-Wolfe Beyond 1/t Convergence. Preprint. Available at https://arxiv.org/abs/2604.28006.

References I

K. I. Appel and W. Haken. Every planar map is four-colorable. *Illinois Journal of Mathematics*, 21(3):429–490, 1977. Computer-assisted proof of the four color theorem (first major such proof).

J. Haase and S. Pokutta. Human–AI Cocreativity: Exploring synergies across levels of creative collaboration. In J. C. Kaufman and M. Worwood, editors, *Generative Artificial Intelligence and Creativity*, chapter 16, pages 205–221. 1 2026. doi: 10.1016/B978-0-443-34073-4.00009-5.

T. C. Hales, M. Adams, G. Bauer, D. T. Dang, J. Harrison, H. L. Truong, C. Kaliszyk, V. Magron, S. McLaughlin, N. T. Thang, N. Q. Truong, T. Nipkow, S. Obua, J. Pleso, J. Rute, A. Solovyev, et al. A formal proof of the kepler conjecture. *Forum of Mathematics, Pi*, 5:e2, 2017. doi: 10.1017/fmp.2017.1. Flyspeck project: complete formal verification of the Kepler conjecture.

M. J. H. Heule, O. Kullmann, and V. W. Marek. Solving and verifying the boolean pythagorean triples problem via cube-and-conquer. In *Theory and Applications of Satisfiability Testing – SAT 2016*, volume 9710 of *Lecture Notes in Computer Science*, pages 228–245. Springer, 2016. Computer-intensive proof generating a multi-terabyte certificate.

S. Pokutta. Frank-Wolfe Beyond 1/t Convergence. *preprint*, 4 2026. doi: 10.48550/arXiv.2604.28006.

N. Robertson, D. P. Sanders, P. Seymour, and R. Thomas. The four colour theorem. *Journal of Combinatorial Theory, Series B*, 70(1):2–44, 1997. Shorter and fully computer-verified proof reducing the number of configurations.

M. Zimmer, N. Pelleriti, C. Roux, and S. Pokutta. The Agentic Researcher: A Practical Guide to AI-Assisted Research in Mathematics and Machine Learning. *preprint*, 3 2026. doi: 10.48550/arXiv.2603.15914.