

When machines begin to discover:

Mathematics, AI, and the future of human reason

How does the seemingly ubiquitous presence of AI transform mathematical discovery, proof and verification? AI-based technologies certainly facilitate mathematical discoveries, but this is not without risk, and the philosophical, ethical and institutional consequences of closer human–AI collaboration must be considered. These recent developments can be better understood by considering the longer history of the mechanisation of mathematics.



Dr Michael Friedman
Research Associate,
Mathematical Institute,
University of Bonn



Dr Scott Harper
Assistant Professor,
School of Mathematics,
University of
Birmingham

This machine “will be desirable to all who are engaged in computations [...] For it is unworthy of excellent men to lose hours like slaves in the labour of calculation, which could be safely relegated to anyone else if the machine were used”. This is how Leibniz describes the advantages of his calculating machine in a 1685 manuscript titled *Machina arithmetica*.

Mathematics is usually considered to have a special status among the natural sciences; on the one hand it is considered as a domain consisting of abstract, necessary, universal and objective truths, independent of time and space, and on the other hand, it is a cultural-historical practice carried out by human beings.

The human mathematician therefore creates or discovers such truths, invents proofs and concepts, and develops new subdomains while solving complex problems, exploring patterns and abstractions, drawing analogies, and reinterpreting results. In this sense, mathematics can be considered a universal, objective domain whose body of knowledge is created by humans. However, as the above quote from Leibniz suggests, starting in the second half of the 17th century, one finds various mechanisation efforts of mathematical practices, seen for example with Pascal's Pascaline or with Leibniz's stepped reckoner.

Similar remarks continue well into the 19th century. Schopenhauer underlines

in his 1851 *Parerga und Paralipomena* a parallel devaluation of human mathematical activity exactly because it is machine-executable: “That the lowest of all mental activities is arithmetic is proven by the fact that it is the only one that can also be executed by a machine, just as now in England this kind of calculating machine is already in frequent use for the sake of convenience”.

In the 20th century, computers started to play a role in solving abstract mathematical problems, by checking solutions for a potentially large (but always finite) number of cases. The most famous example of this is the 1976 proof of the Four Colour Theorem, which asserts that only four colours are needed to colour any map so that neighbouring regions do not match.

However, even when this mechanisation became more and more sophisticated, the mathematician's unique role of mathematical invention, proof finding and conceptual judgment was very much preserved. On the contrary, in the 21st century, and especially during the last couple of years, such a role and position of the mathematician begins to unravel, and several examples illustrate this shift.

AI-based mathematics: several examples

OpenAI, the company behind ChatGPT, overturned a well-known conjecture that had stumped mathematicians for 80 years. Place nine coins on the floor and

try to make as many pairs as possible exactly a ruler's length apart. A three-by-three grid gives 12 such pairs, not far off the best possible 18. In 1946, Paul Erdős, one of the most prolific mathematicians of the 20th century, asked, in a precise way, whether grid layouts are always close to best possible, no matter how many coins you have. In May 2026, an internal OpenAI model showed they are not, using tools not just from geometry but, surprisingly, from the algebraic study of numbers. The Fields Medallist Tim Gowers called this “a milestone in AI mathematics”.

Others followed. In July, the mathematician Levent Alpöge working at Anthropic used Claude Fable 5 to find a counterexample to the Jacobian Conjecture, which proposes that a certain sort of mathematical function should never send two inputs to the same output. Alpöge gave one that does (see Figure). The next month, OpenAI published solutions to ten open problems across a range of mathematical areas.

Who checks this work? Some results are disproofs: someone said “this always happens” and the machine replied “it fails here”. The counterexample to the Jacobian Conjecture is of this kind, and once found, it is easy to check. Proofs are harder to check as mathematicians demand a rigorous argument for any assertion, and the solution to Erdős's problem was checked by a team of well-known mathematicians. Increasingly,



Visualisation of a counter example to the Jacobian Conjecture, found in July 2026 by mathematician Levent Alpöge working at Anthropic using Claude Fable 5

though, the checker is also a computer. In proof formalisation, an argument is rewritten as code in a precise technical language, and if the code runs with no errors, the proof has no gaps. Each of OpenAI's ten solutions was released with such a certificate, written in the language Lean.

AI-based mathematics: shifts, promises and risks

So there is no doubt that essential mathematical practices – finding proofs, checking arguments, solving problems – are changing. The reactions of the mathematical community span from considering AI-based technologies as collaborators, to expressing wonder, astonishment or even anxiety in front of such technologies.¹

Practically speaking, one may ask, how do mathematical communities engage with and navigate these technological and disciplinary changes? What are the philosophical implications and how one can assess such developments historically?

The Collaborative Research Grant project “On Proofs and Partnerships: How AI, Big Data, and Proof Assistants are Transforming Mathematical Practices” (2026–2027), headed by the University of St Andrews and the University of Bonn, aims to answer these questions.² One of the project's main aims is to establish how such developments represent both continuity with, and rupture from, earlier attempts to mechanise mathematical

reasoning. In our two recent workshops (the first at the University of Bonn, June 2026 and the second at De Morgan House, London, August 2026) we have examined this trajectory with practising mathematicians and historians and philosophers of mathematics, contextualizing current developments within the longer arc of debates about the relationship between human cognition and computational aids in mathematical discovery. Alongside historical research, the project aims to contribute to current debates about the nature of scientific reasoning in the age of AI and the proper response due to such changes.

How should the mathematical community react to such developments? The recently published Leiden declaration attempted to answer this question.³ The declaration defends transparency, proper attribution, and research autonomy. It warns that AI-based technologies can obscure sources, distort incentives, weaken peer review, and concentrate power in the high-tech industry. It calls for disclosure of tool use, human authorship and accountability, as well as robust regulation of commercial AI.

But several questions arise:

- How should authorship and epistemic credit be allocated when a result depends on a distributed chain of human and AI-collaboration?
- What institutional arrangements would ensure that libraries of

mathematical texts, training data, and agents remain publicly governed and globally accessible rather than becoming proprietary research infrastructure?

- What political obligations follow when universities and individual mathematicians collaborate with AI firms whose systems may support surveillance, military projects, or the privatisation of public knowledge?
- And last but very certainly not least: may ‘AI authorship’ become a mechanism for corporate appropriation and erasure of former human labour?

Such questions are of essential importance, and the role of the humanities and the mathematical community is increasingly crucial in formulating a reaction to this ever-growing proprietary infrastructure.

References

¹ Friedman M (2026). Mathematical Anxiety in the Age of AI: Opacity, Role Inversion, and the Changing Place of the Mathematician. *Computers in Human Behavior: Artificial Humans*. doi.org/10.1016/j.chbah.2026.100386.

² The project is led by two PIs: Michael Friedman (University of Bonn) and Deborah Kent (University of St Andrews).

³ Alper J, et al (2026). Leiden Declaration on Artificial Intelligence and Mathematics. *Zenodo*. <https://leidendeclaration.ai>