

A Severe Misalignment of AI in Mathematics

Over the last few months, the mathematical capabilities of LLMs have improved dramatically, to the point that they can solve major outstanding problems in many fields of mathematics. However, the push by AI companies to solve mathematical problems as a benchmark is detrimental to the science of mathematics, and to the mathematical community. The goals of the AI companies and the goals of the mathematical community are severely misaligned. We see these as part of broader alignment issues impacting other scientific and creative professions, as well as the whole of society.

Research mathematics deals with understanding basic structures of shapes, numbers, and natural phenomena. Over the course of generations, it has built a large corpus of sophisticated ideas, methods, abstractions, and other tools to comprehend the mathematical landscape. In turn, modern technologies and sciences are based on mathematical tools.

Famous problems have often served as landmarks and lighthouses against which one can measure an improved understanding of this landscape. Solving one of these problems has been a certain sign of new insights and interesting methods, which would then be studied by a community of mathematicians, through a long and arduous process of talks, discussions, simplifications. At the end of this process, one will ideally find a textbook presentation of the results suitable for any graduate or even undergraduate student to study. Some of the mathematical ideas pursue their journey even further to become, decades or centuries after, tools that are understood and used by the whole population.

The mathematical community functions, in many ways, as a miniature version of humanity. It consists of individuals using a wide variety of different approaches, joined by core values. The most precious resources of our profession are students and ideas, and these we nurture with great care. We feel responsible to let them grow to their full potential, until they can live a life of their own in the mathematical world. For students we often suggest problems with the core intention of developing skills making them well-positioned for advances in research and elsewhere. Our ideas we disseminate in talks, private discussions and careful writeups, connecting them to the previous ideas of others. These processes invariably take time and are based on human interaction.

In recent months, the success of AI in solving major mathematical problems has made headlines even outside mathematical circles. But solving problems is only a tool and proxy for achieving the primary goal of conceptual understanding and insight. Forgetting this in the world of AI may turn the tool against the primary goal. Indeed, the mass production at faster and faster pace of "true/false" statements could destroy fertile ground instead of breathing life into new ideas.

Often these solutions are announced in a rush, leaving no time for a proper writeup, the isolation of new methods and ideas, and citing relevant previous work of others. As in all creative professions, this raises severe attribution and plagiarism questions. Moreover, without the willing mathematicians who must take care of their development and integration into the mathematical canon, AI-conceived ideas would never become fully alive and the crucial human transmission chain between mathematicians would be lost.

We are witnessing a general threat to intellectual work, with misalignment between the outcome of the use of AI and its initial purpose. In many fields and activities, years of training have traditionally served not only to produce a final answer or product, but also to develop understanding and the ability to formulate new questions and ideas. However, building on a vast body of previous human work, AI systems are becoming increasingly capable of producing the results of such work directly, and these goals cease to align. The issues the mathematical community faces now are similar to issues that other scientific and creative professions are facing, and indicate issues that all of humanity might face: how to make sure that, as AI changes the way work is done, we do not lose sight of what that work was meant to achieve in the first place.

AI offers the potential of enhancing and accelerating genuine mathematical study and understanding. Mathematics as a profession will need to adapt to these changes in several ways. However, whether these changes ultimately benefit the field or have a destructive effect will in large part be determined by the decisions of the humans in control of this new technology.

These issues must be addressed urgently, in the mathematical community, by the companies developing these technologies and, more broadly, by a society that will confront similar problems in many other forms of intellectual work.

Artur **Avila** (Fields Medal 2014)
Manjul **Bhargava** (Fields Medal 2014)
Caucher **Birkar** (Fields Medal 2018)
Pierre **Deligne** (Fields Medal 1978)
Yu **Deng** (Fields Medal 2026)
Simon **Donaldson** (Fields Medal 1986)
Hugo **Duminil-Copin** (Fields Medal 2022)
Alessio **Figalli** (Fields Medal 2018)
Martin **Hairer** (Fields Medal 2014)
June **Huh** (Fields Medal 2022)
Maxim **Kontsevich** (Fields Medal 1998)
Elon **Lindenstrauss** (Fields Medal 2010)
Pierre-Louis **Lions** (Fields Medal 1994)
James **Maynard** (Fields Medal 2022)
Curtis **McMullen** (Fields Medal 1998)
Shigefumi **Mori** (Fields Medal 1990)
Ngô Bảo Châu (Fields Medal 2010)
Andrei **Okounkov** (Fields Medal 2006)
Peter **Scholze** (Fields Medal 2018)
Stanislav **Smirnov** (Fields Medal 2010)
Terence **Tao** (Fields Medal 2006)
Maryna **Viazovska** (Fields Medal 2022)
Cédric **Villani** (Fields Medal 2010)
Wendelin **Werner** (Fields Medal 2006)
Efim **Zelmanov** (Fields Medal 1994)