The need for ethical guidelines in mathematical research in the time of generative AI

Markus Pantsar[1][0000-0001-8572-1453]

[1] RWTH Aachen University, Theaterplatz 14, 52062 Aachen Germany
Markus.pantsar@gmail.com

Abstract. Generative artificial intelligence (AI) applications based on large language models have not enjoyed much success in symbolic processing and reasoning tasks, thus making them of little use in mathematical research. However, recently DeepMind's AlphaProof and AlphaGeometry 2 applications have recently been reported to perform well in mathematical problem solving. These applications are hybrid systems combining large language models with rule-based systems, an approach sometimes called neuro-symbolic AI. In this paper, I present a scenario in which such systems are used in research mathematics, more precisely in theorem proving. In the most extreme case, such a system could be an autonomous automated theorem prover (AATP), with the potential of proving new humanly interesting theorems and even presenting team in research papers. The use of such AI applications would be transformative to mathematical practice and demand clear ethical guidelines. In addition to that scenario, I identify other, less radical, uses of generative AI in mathematical research. I analyse how guidelines set for ethical AI use in scientific research can be applied in the case of mathematics, arguing that while there are many similarities, there is also a need for mathematics-specific guidelines.

Keywords: Automated theorem proving, mathematical practice, research ethics

1 Extended abstract

The importance of artificial intelligence (AI) applications in the modern world can hardly be overstated. With the introduction of deep neural networks and transformer architectures, machine learning systems have been successful in many areas where computers were previously of limited use. A particularly exciting development has been that of generative AI, based on large language models (LLM) and more recently multi-modal large language models. However, for all their success fields like natural language and image processing, generative AI applications have been notoriously bad in tasks involving symbolic processing and reasoning. As probabilistic models, LLMs are simply used to output the most likely (or one of the most likely) tokens associated with a string of tokens. Consequently, they do not reason in a human-like fashion, which is why they can be led astray so easily in simple-seeming reasoning task.

When it comes to mathematical AI, the LLM-architectures thus seem inherently problematic. Mathematical deduction is not probabilistic. Instead of detecting patterns in data, a successful mathematical AI application has to follow rules – corresponding

in some way to the axioms and rules of proof in formal systems of mathematics. Against this background, it is hardly surprising that LLM-based AI systems have found relatively little use in mathematics. However, this may change when LLM-architectures are combined with rule-based systems for new types of hybrid systems. In this kind of system, the LLM is used to generate potential solutions to mathematical problems, which are then tested on a rule-based system. Such a hybrid approach received a big boost in 2024, when Google's DeepMind reported significant success with its *AlphaGeometry* and *AlphaProof* applications in solving problems of the International Mathematical Olympiad [1,2]. Combining the *Gemini* large language model and the rule-based theorem prover *Lean*, the applications are exactly that kind of hybrid – sometimes called *neuro-symbolic* – AI.

In this philosophical paper, I investigate the potential of such AI tools in research mathematics, and their ethical consequences. While current theorem proving tools used by mathematicians are rule-based systems and as such have limited functionality, it is conceivable that a neuro-symbolic hybrid system could provide new proofs, including new mathematical theorems, autonomously. Such an *autonomous automated theorem prover* (AATP) could be transformative to mathematical practice. In the extreme case, new mathematical proofs could be generated simply by entering a system of axioms and rules of logic as input to an AATP. In such case, the human contribution would be minimal, yet it could lead to important achievements in the mathematical community. But even in less extreme cases, AI tools could be used to replace much of what is currently valued in the work of human mathematicians. This raises important questions about the future of mathematics. In this paper, I present epistemological considerations based on such a scenario. However, I will ultimately focus on the *ethics* of using such mathematical AI applications.

The question is not whether such AI applications should or would be used. If their use means progress in mathematics – e.g., providing a proof for a new interesting theorem – it is likely that the mathematical community would not only accept but embrace the use of the new tools. Hence the challenge is to find proper scientific practices for the use of such tools, including clear ethical guidelines.

Based on the guidelines presented for AI use in scientific research by Resnik and Hosseini [3], I propose guidelines for the ethical use of generative AI tools in mathematical research, and how to communicate such use. I argue that mathematical research has distinct characteristics that necessitate mathematics-specific guidelines. In the near future, a lot is at stake for mathematical communities – and we appear to be woefully unprepared for that. In an extreme case, publication records and consequently even careers in mathematics could be created by being a skilful user of AI. If AI use is neither declared nor detected, dishonest people can use AI tools to create a misleading image of expertise. Indeed, this might be happening already, even though AATPs have not yet been publicly introduced.

References

1. DeepMind, AlphaGeometry: An Olympiad-level AI system for geometry, Google DeepMind (2024). https://deepmind.google/discover/blog/alphageometry-an-olympiad-level-ai-system-for-geometry/ (accessed October 10, 2024).
2. DeepMind, AI achieves silver-medal standard solving International Mathematical Olympiad problems, Google DeepMind (2024). https://deepmind.google/discover/blog/ai-solves-imo-problems-at-silver-medal-level/ (accessed October 10, 2024).
3. D.B. Resnik, M. Hosseini, The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool, AI Ethics (2024). https://doi.org/10.1007/s43681-024-00493-8.Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010).