Mathematics in the age of AI

Asvin G.*

With recent spectacular progress in machine learning and AI, it is at least conceivable that many human intellectual and creative tasks might be automatable very soon. In this essay, we explore the possible scenarios and implications for such progress within the very narrow domain of mathematics, both because there is reason to suspect that automation will proceed fastest in this domain and because it is the field the author knows best.

The likely shape of our future

We have consistently underestimated and mispredicted the rate of progress within machine learning and AI in the past decade. Tasks that once seemed to be safely out of reach of computers (comprehending human language, playing go, generating images, writing code ...) have all been automated to some degree and progress shows no signs of slowing. There are good reasons to suspect that mathematical research is next in line - mathematics is a domain where correctness is easy to verify, there is comparatively a vast amount of human generated training data and increased amounts of computation leveraged through search might have a genuine advantage over humans (as was the case in board games such as chess or go).

Given these factors and existing trends, the author believes that automation might make significant inroads into mathematical research by the end of this year (2025). While the future is still murky and uncertain, it seems worthwhile to try and extrapolate current progress - both to prepare for the coming shocks but also to shape our future in more productive directions. I expect progress to be uneven and choppy and for mathematical tastes to change in reaction to advancing capabilities.

Research in mathematics is made up of many distinct threads and the focus on what is important has shifted over time. For instance, computation was a much more important part of any mathematician's activity before the twentieth century but the introduction of computers (amongst other things) shifted the focus to a more abstract viewpoint that currently dominates large parts of research. The introduction of current AI methods into mathematics will likely have a similar shift on the interest and focus of research mathematicians.

Proof completion and automation

In the short term, the most likely target for automation is *proof completion* and *auto-formalization*. Proof completion might take several forms, the simplest of which would simply be for the machine to produce likely proofs (in a fixed natural or formal language) to a given statement. Initially, these are more likely to be applicable to relatively simple lemmas (that fill up most of a math paper)

*email: gasvinseeker94@gmail.com

but over time (perhaps a very short time) graduate to more difficult proofs, perhaps even the main statements of current mathematical papers.

On the other hand, (auto)formalization is concerned only indirectly with current mathematical research and the focus would instead be on converting natural language proofs (generated by a human or AI) into a formal language. This is straightforwardly useful for proof verification but the larger importance likely derives from such formalization efforts being the "training data" for automating proof completions. Conversely, proof completions itself provides training data for AI to get better at formalization. As such, proof completions and auto-formalizations can form a tight feedback loop for rapid progress in automating mathematical research.

The impact on mathematical research

What is likely to be the impact of AI automating away most of the lemmas one needs in the course of writing a mathematical paper (and even bigger theorems)? I predict that mathematicians, as a community, will then downgrade the emphasis on finding and writing proofs. While they will still form the backbone of mathematics, they will be relatively invisible and mathematicians will emphasize grand mathematical programs and conjectural landscapes over limited proofs of specific theorems. Moreover, *definitions* and *axiomatizations* will take center stage and be the prime mathematical currency by which we test our understanding.

A historical example might help illustrate this idea. André Weil in the early twentieth century formulated some conjectures in number theory and algebraic geometry concerning the relation between finite field arithmetic and complex topology. Over the next few decades, Jean-Pierre Serre and Alexander Grothendieck led a project (involving many other mathematicians) to prove these conjectures by rewriting the foundations of algebraic geometry in order to extend the tools of algebraic topology into algebraic geometry. The vision for what should be build was relatively clear early on and an enormous amount of work went into finding exactly the right formalizations and definitions in order to realize this vision. This naturally involved proving hundreds of little lemmas illustrating that the proposed definitions did indeed match the intended intuition and were sufficiently powerful to realize the proof of the Weil conjectures.

In our envisioned future however, such a project would be mostly carried out by AI. While a human might have the grand vision in mind and propose initial definitions based on his or her intuition, AI would very quickly try many possibilities to test the validity of the proposed definitions for the task at hand and perhaps even suggest refinements. The work of the human then would be focused on refining their intuition (using the computer experiments) to iterate on their definitions and programs, and to find possible patterns to explain in the first place. Proofs would be much closer to *natural experiments*, human mathematicians to project managers and AI would play the role sometimes played by graduate students in other scientific disciplines of carrying out an established experimental program.

However, the seeds for such grand visions are rooted in a deep understanding of existing mathematics including current proofs and this will not be significantly different in the future either. Necessarily then, a large part of our energy will also be spent in understanding the proofs generated by AI and re-conceptualizing it for human digestion. This will be far from a routine activity and involve significant creativity by itself, and hopefully will receive the appropriate attention and accolade it will deserve. As Thurston forcefully observes, "The product of mathematics is clarity and understanding. Not theorems, by themselves."

The long-term prognosis

This final section is necessarily much more speculative (even if "long-term" might only mean five years instead of one). So far, I have supposed that AI is competitive or super-human at producing proof completions but not at coming up with new definitions or new frameworks in which to find the proofs to observed conjectures and patterns. One might question this premise: isn't it possible that AI sufficiently advanced at proof completion will also be able to output (extremely) long proofs to our most difficult questions? I think this is unlikely because any proof completion is a combination of "intuitive understanding" combined with a search through the space of possible proofs and this search grows exponentially and in-feasibly more difficult in the length of the proof (even with exponentially falling costs). Improvements to proof completion then necessarily come through improvements in the "intuitive understanding" which is a direct consequence of identifying the correct concepts in which to express our mathematics.

But in the long run, it is almost certain that we will also be able to automate the process of creating such new definitions and concepts. What will become of mathematicians then and what role, if any, will they have to play in society? This question is hopelessly tied up with broader questions about the role of AI in society and the place for humans in a "post-AGI world". For now, the best we might do is look for inspiration in other domains that have had AI impact them similarly. The closest such comparison might be to chess or go.

In these games, AI is now at a point that is completely super-human and the role of humans has been reduced almost entirely to querying their computer and then trying to understand the resulting suggestions in order to improve their own understanding that can then be deployed in games between humans. Since chess and go are at their core spectator games, this has not impacted the "jobs" of elite chess/go players. Moreover, the easy access to superhuman intelligence (in the context of these games) has made it easier in many ways for people to get into the game. Will this also be the fate of mathematics?

References