DeepMind's AlphaGeometry would win Math Olympiads Silver Medal

Augmented large language models take another step towards human-like reasoning

Wolfgang Spahn

2024-01-31

The AI world got off to a strong start in January 2024. Google DeepMind published in Nature [1] an AI system that solves complex geometry problems at a level beyond most humans. It outperformed the current silver medallist in the International Mathematical Olympiad.
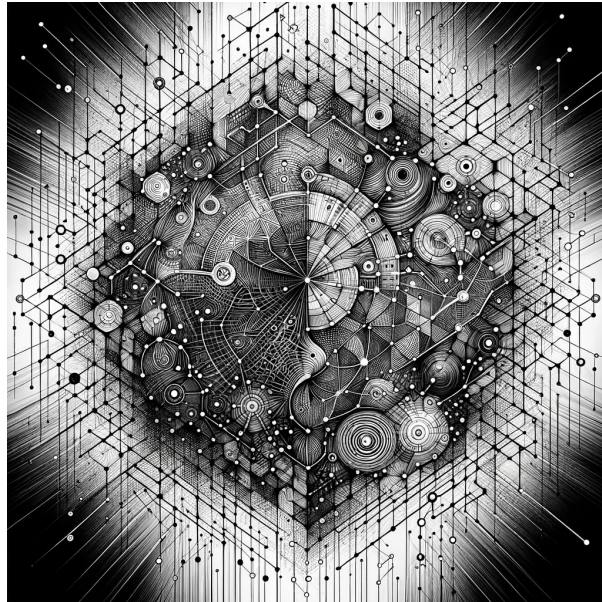

Figure 1: As AI imagines to do Math (by GPT-4)

This combination of a large language model and a symbolic deduction engine could solve 25 out of 30 problems, where state-of-the-art systems achieve 10 out of 30 mathematical tasks. It is a nice illustration of the recent AGI trend of combining neural machine learning with classical AI for hard logic.

This trend has emerged in 2023, as the amazing ability to generate responses that seem remarkably human-like in their adaptability and creativity comes at the cost of wrongly extrapolating or even hallucinating answers. The year 2023 has seen various attempts to add factual accuracy or logical coherence.

In an approach called augmented large language models, researchers try to follow the findings of cognitive psychology, popularised by the psychologist Daniel Kahneman in his book "Thinking, Fast and Slow". These terms refer to two different ways in which the human brain thinks. Fast thinking occurs automatically and quickly, with little or no effort. It is efficient for processing familiar information and situations, but can be prone to bias and error, particularly in complex or unfamiliar situations. Slow thinking directs attention to the more effortful mental activities, such as complex reasoning and deliberative decision-making. It's

---

[1] https://www.nature.com/articles/s41586-023-06747-5

more logical and methodical, but it's also slower and requires more mental energy. In blended AI cognitive systems, LLMs play the role of fast thinkers and symbolic engines are the slow thinkers.

DeepMind used existing classical symbolic AI engines on 100 million synthetic theorems and their proofs to train an LLM to generate proposols for solving unknown complex geometry problems, which are then validated and finalised by a symbolic engine.

For our digital AI-driven tutors, we use a similar combination of classical supervisor logic combined with machine learning, not to solve maths problems, but to teach students how to solve reasoning problems. Because we are targeting pre-university higher education, the logic and maths required is much less sophisticated. The challenge is how to come to a realistic pedagogical dialogue applying well specified didactic principles. But the approach is the same, combining artificial fast (neural) and slow (symbolic) thinking to ensure correctness together with human like dialog capabilites.

2023 had brought breakthroughs that were unimaginable a few years ago. Bringing this potential into a real system, discussing the risks involved and finding ways to mitigate them has been very exciting. I am sure that 2024 will continue at the same pace.