What can't models ever do?

Dongjae Lee

2025.05.21

Abstract

The answer to 'What can models do now?' is changing very rapidly. Lots of the previous research problems have been easily solved by just improving models. This rapid progress complicates the search for meaningful research directions. Rather than focusing on current limitations, we must address fundamental constraints of models.

Recent years have marked the era of language models and AI. AI is becoming essential across all fields, driven by significant performance improvements. In certain domains, AI now surpasses human capabilities. It looks like advances in AI can solve all the existing problems.

Nevertheless, greater intelligence alone does not solve every problem. Language models are central to intelligent software but require auxiliary tools, such as MCPs (Model Context Protocols), for accessing web, temporary memory and so on. A sandbox for code execution is an example of such tool.

Still, these tools do not resolve all issues. Language models remain prone to errors, especially as information and requirements increase and complexity grows. Such errors are minor in chatbots because users can easily correct them. In contrast, in agent systems, errors can have real-world consequences before human correction.

To mitigate this, I propose integrating rule-based AI like SMT (Satisfiability Modulo Theory) solvers. Declarative requirements can be used to verify that model actions meet constraints, improving system safety. This enables ongoing verification, which is vital for agent systems. Symbolic methods, once seen as competitors to neural networks, may now become essential for verifying neural-based agents.