Review of: "Limitations of and Lessons from the Learning of Large Language Models"

Tibor Tajti

Potential competing interests: No potential competing interests to declare.

The goal of the article is to argue for theoretical limitations of logical reasoning that can be learned by large language models due to their restriction to investigating a relatively short sequence of tokens.

The main achievements are:

It connects the limitations of large language models to the Curry-Howard correspondence, which establishes an isomorphism between proofs in logic and programs in a functional typed lambda calculus.

It argues that while intuitionistic logic maps to a version of lambda calculus that can be carried out locally, classical logic requires non-local relations that cannot be learned by large language models.

In terms of originality and contribution to the field, the paper aims to provide a foundational theoretical understanding of the principal limitations of large language models for logical reasoning, which it argues has been lacking. The application of the Curry-Howard correspondence seems like a novel approach.

Technically, the explanations of lambda calculus, typed lambda calculus, the Curry-Howard correspondence, and lambda-mu calculus seem sound. The argument linking these technical details to the limitations of large language models follows logically.

The methods used are appropriate - applying the theoretical tools of proof theory and programming languages semantics to understand logical reasoning capabilities.

Some important related works are not referred to in the article, e.g., I do not see any references cited in the paper that directly relate previous applications of the Curry-Howard correspondence to analyzing capabilities of large language models. Without references to prior related work combining Curry-Howard with analyses of computational reasoning abilities, the paper does not fully establish the originality of its theoretical approach.

The writing is clear and easy to follow for a technical paper in the field. The explanations are easy to understand. The only minor issues are occasional typos or grammatical errors.

No data is presented, so reproducibility does not apply. However, the theoretical framework and example are clearly explained to allow evaluation of the argument.

In terms of strengths, the paper provides a novel foundational theoretical understanding of limitations. The application of

the Curry-Howard correspondence is an interesting approach. However, some weaknesses are that the argument is not demonstrated empirically and the conclusions cannot be directly falsified.

The language, readability, and grammar are very good, with only minor occasional typos or errors.

The paper provides an interesting theoretical contribution to understanding limitations of large language models. With some empirical demonstration or discussion of falsifiability, it could be improved for publication in a journal.

There is a strong statement in the paper saying: "It is pretty safe to say that nobody really understands how large language models produce their results."
Although large language models (LLMs) like GPT are complex and sometimes considered "black boxes," the claim that nobody understands how they produce results is an overstatement. The foundational principles, such as transformer architecture and training processes, are well understood. Ongoing research in AI interpretability and explainability is gradually demystifying their operations. While individual predictions can be intricate, the AI community has a solid grasp of these models' overall functionality and is making strides towards greater transparency and understanding.

The conclusions identify intrinsic design constraints rather than making falsifiable empirical claims. While not proven through experiments, they seem to be valid extrapolations from the underlying formal models accurately portrayed.

The only possible issue is the lack of direct empirical validation, but the paper acknowledges this and aims for foundational theoretical understanding rather than testing.

Overall, given the technical presentation and logical flow of reasoning, the main conclusions drawn about limitations of LLMs for classical logic versus intuitionistic logic seem scientifically sound based on the arguments provided in the paper. The theoretical contribution and insights are novel and useful.