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

Salman Panahy

Potential competing interests: No potential competing interests to declare.

The skeleton of the main argument of the paper is as follows:

1. In LLM, the meaning of the tokens is fixed by the translation of training text to token sequences and vice versa. These input sequences are of a certain fixed length.
2. This length limits the ability of the model torefer non-locally to distant parts of the input stream.
3. To extend the Curry-Howard correspondence between intuitionistic logical proofs and functional programming to classical logic proofs, one needs to add non-local control operations to their calculus (Lambda-Mu calculus) or use continuation passing style in their Lambda calculus.
4. Either of the above techniques allows for non-local jumps, and to implement them in a program working with sequences of tokens, requires addresses of the beginning and the end of non-local operations to be present in the string in question.
5. This makes the string possibly too long for an LLM input, given the fixed length of its input.

From a review of literature about the Transformer architecture and its differences from Recurrent Neural Networks (RNN), I am not convinced that step 2 holds. If the author believes that it holds, they need to give more evidence for it. From my understanding, transformers are designed partly to address the problem of losing information in long inputs. If the author's argument is that even this much improvement is not sufficient, then they need to clarify that. In general, it would be helpful to clarify if the argument is leveled against the techniques used specifically in LLM, or against any type of technique using Neural Networks, or against any type of program working with sequences of tokens.

Given that steps 3 to 5 are valid, in order to demonstrate that this theoretical argument has any practical tooth, it would be very useful to show that some classical reasoning in mathematics, if coded using lambda-mu calculus or lambda calculus using continuation passing style, cannot be realized by sequences below a certain length. Although even this much does not prove that all of the formalizations of the reasoning in question cannot be realized. It seems to me that the formalization of a proof (propositional or predicate calculus) and the unit of tokenization (letters, words, or atomic sentences) are two important parameters here.

The above considerations aside, the article tackles a novel and underdeveloped field of research. We need to have a much better understanding of the functionality of machine learning algorithms from a mathematical, logical, and philosophical point of view.