Qeios

Peer Review

Review of: "Enhancing Long Context Performance in LLMs Through Inner Loop Query Mechanism"

Han Xiao[1]

1. University of California, Santa Barbara, United States

This paper provides an innovative method to process very long documents. While traditional RAG systems retrieve the most relevant information based on the query, the proposed ILM-TR method mimics how humans tackle complex questions by storing information in “short-term memory” and iteratively building up understanding through multiple retrieval cycles. The research question is relevant and important, and the authors have proposed an interesting solution.

Strengths:

The proposed method is innovative and intended to overcome limitations in the traditional system. The dual-summary approach (regular summary + surprising facts) provides a clever way to capture overlooked information that can be potentially important.

Weaknesses:

The authors mentioned that the key limitation is time consumption, and hence the authors should provide a more comprehensive discussion on the efficiency of the proposed method, including time consumption as well as hardware requirements.

There is a lack of rigor in the definition and measure of “surprise” - “the informational value of a communicated message depends on the degree of surprise in its contents”, however, the authors should elaborate on what constitutes “surprise” more specifically in the context of text summarization. It is also beneficial to provide examples of “surprising information”.

The authors should also discuss potential error amplification in the iterative retrieval systems. The ILM-TR method depends cumulatively on previous steps, and once incorrect information enters, it may be

doi.org/10.32388/BIOTIT

reinforced since the system may preferentially retrieve additional information that confirms this error through confirmation bias. The authors may also describe a mechanism for error detection or correction.

Declarations

Potential competing interests: No potential competing interests to declare.