

Peer Review

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

Haoxiang Shi¹

1. Waseda University, Japan

The paper proposes a creative approach to enhancing retrieval-augmented generation with a focus on iterative refinement and surprising information extraction. However, it would benefit from a more comprehensive comparison with existing methods and clearer evaluation metrics. Addressing these issues could significantly improve the paper's contribution to long-context understanding in LLMs.

Strengths:

- The introduction of **Inner Loop Memory-Augmented Tree Retrieval (ILM-TR)** is innovative, especially with the use of **Short-Term Memory (STM)** to iteratively refine retrieval processes.
- The focus on **surprising facts** in the summarization process is intriguing and provides a novel way to enhance comprehension in long-context scenarios.
- The demonstrated improvements on benchmarks such as **M-NIAH** and **BABILong** validate the method's potential for handling long-context tasks.

Weaknesses:

- The paper lacks clarity on **summary quality metrics**, such as ROUGE or BLEU, and does not provide sufficient qualitative analysis for the "surprising facts" mechanism.
- The convergence criteria for STM-based retrieval are vaguely defined, making it difficult to assess the robustness of the proposed inner-loop process.
- Baseline comparisons are limited to RAPTOR, excluding other widely-used retrieval-augmented methods (e.g., DPR, REALM), which restricts the scope and impact of the findings.

Declarations

Potential competing interests: No potential competing interests to declare.