

# Examining the Ability of Large Language Model on Entity-Based Question Answering

Seonmin Koo<sup>1,◦</sup>, Yeongwook Yang<sup>2</sup>, JeongBae Park<sup>3,\*</sup>

<sup>1</sup>Department of Computer Science and Engineering, Korea University

<sup>2</sup>Department of Computer Science and Engineering Gangneung-Wonju National University Wonju, Republic of Korea

<sup>3</sup>Human-inspired AI Research

{fhdahd, insmile}@korea.ac.kr

## Abstract

Recently, large language models have shown remarkable performance in various tasks. Nevertheless, in knowledge-intensive tasks, large language models often have difficulty following the correct inference path, which results in incorrect responses. There are existing methods to improve the inference ability of large language models by cutting out or summarizing parts of external knowledge that may contain important information. However, these methods may contain important information in the parts lost during the reduction process. To address this, in this paper, we propose a method to improve the ability of large language models in question-answering tasks through entity-based inference without losing the original external knowledge. We explicitly emphasize that important entities in the context can be accurately focused to improve the inference ability of large language models. This provides a foundation for providing insight into the Korean question-answering ability of large language models through entity-based inference.

Keywords: Large Language Models, Relation Extraction, Question Answering, Probing

## 1. Introduction

The question-answering task aims to generate the correct response to the user’s query. With the development of large language models, they have shown high performance in various tasks including question-answering tasks, and their usability in real-world applications such as Alexa and Siri is increasing [1, 2]. Due to the constantly changing nature of knowledge and the limitations of the size of large language models, the utilization of systems that provide external knowledge to the model and infer based on it, such as search augmentation generation, is receiving more attention. In the question-answering situation, large language models are required to have the ability to generate responses based on an appropriate inference path for given external knowledge. However, large language models often have difficulty following the correct inference path, which results in incorrect responses.

In existing studies, there are various methods to improve the understanding of external knowledge of large language models. We summarize and provide a focus on the parts that can serve as evidence for the answer in the context [3, 4]. Or,

by utilizing a small language model, the complexity of each token in the prompt is evaluated and tokens with little information are removed to provide a compressed overall length, or the context is divided into chunks and compressed [5, 6].

However, these methodologies may contain important information in the cut-off part during the information summary process. Since the performance of the system cannot always be perfect, it is not possible to rule out the possibility that important information may be omitted, which may hinder the model from following the appropriate inference path or generating the correct response. Therefore, it is necessary to explore a methodology that can effectively utilize a large language model in a question-answering system by allowing the large language model to focus on important information while excluding knowledge loss.

Therefore, in this paper, we propose a method to explore the ability of a large language model in a question-answering task through entity-based inference while maintaining the original external knowledge. We explicitly emphasize that we can accurately focus on important entities in the context to improve the inference ability of the large language model. Specifically, we extract important entities from the query and

---

\*Corresponding author

perform entity extraction on sentences containing important entities in the context. For entities extracted in this way, the location of the entity is explicitly indicated so that it can be focused on.

## 2. Proposed Methods

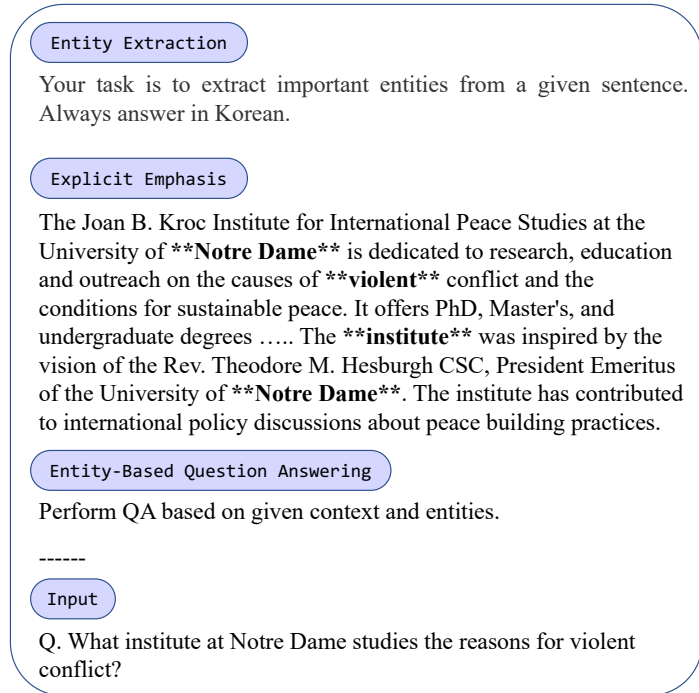


Figure 1. Example prompts for entity-based question answering. The methodology consists of Entity Extraction, Explicit Emphasis, and Entity-Based Question Answering, enabling the model to perform effective entity-based reasoning without losing the original context.

### 2.1 Entity Extraction

Figure 1 shows an example of a prompt for entity-based question answering. Give a query to a large language model and have it extract important entities. For example, if the input is a sentence like “What institute at Notre Dame studies the reasons for violent conflict?”, the model can extract ‘Notre Dame’ and ‘institute’ as important entities. These entities serve as cues for the model to focus on in the context when performing the question answering task.

In the next step, entity extraction is performed on sentences that contain important entities in the context. Since entity extraction is performed on sentences filtered by important entities rather than performing entity extraction on all sentences, it improves more query-focused inference.

Through this process, we create a foundation for focusing on key information for performing question-answering tasks from external knowledge.

### 2.2 Explicit Emphasis

As shown in ‘Explicit Emphasis’ in Figure 1, the extracted entities are explicitly emphasized in the context. ‘Notre Dame’ and ‘institute’, which are important entities in the query, are explicitly emphasized, and ‘violent’, which is one of the entities extracted from sentences containing important entities, is explicitly emphasized. Through this process, we enable the large language model to effectively utilize the part that it focuses on to answer the query without losing the original external knowledge for inference.

### 2.3 Entity-Based Question Answering

The original external knowledge is reconstructed into external knowledge that emphasizes important entities and provided to the model. The model can perform the question-answering task more effectively through entity-based reasoning by explicitly providing the important parts in the context.

For example, important entities such as ‘Notre Dame’, ‘institute’, etc. can be explicitly marked as **\*\*important entity\*\***. These methods can guide the correct inference path while avoiding the possibility of information loss by truncating or summarizing existing external knowledge.

## 3. Conclusion

Recently, large language models have shown excellent performance in various tasks. However, in knowledge-intensive tasks, they often fail to follow the correct inference path and generate inaccurate responses. Existing methods try to improve the inference ability of the model by removing or summarizing some of the external knowledge, but there is a risk of losing important information in the process. To solve this problem, in this paper, we propose an entity-based inference method as a way to improve the question answering ability of large-scale language models without losing external knowledge. We argue that the inference ability of the model can be enhanced by explicitly emphasizing important entities in the context. This approach provides a foundation for providing insights into the Korean question answering ability of large-scale language models through entity-based inference.

## Acknowledgements

This work was supported by Institute for Information communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (RS-2024-00398115, Research on the reliability and coherence of outcomes produced by Generative AI). This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2021R1A6A1A03045425). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1C1C2004868).

## Reference

- [1] C. Qu, L. Yang, C. Chen, M. Qiu, W. B. Croft, and M. Iyyer, “Open-retrieval conversational question answering,” *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, pp. 539–548, 2020.
- [2] K. Basu, “Conversational ai: Open domain question answering and commonsense reasoning,” *arXiv preprint arXiv:1909.08258*, 2019.
- [3] S. Ko, H. Cho, H. Chae, J. Yeo, and D. Lee, “Evidence-focused fact summarization for knowledge-augmented zero-shot question answering,” *arXiv preprint arXiv:2403.02966*, 2024.
- [4] M. A. Arefeen, B. Debnath, and S. Chakradhar, “Lean-context: Cost-efficient domain-specific question answering using llms,” *Natural Language Processing Journal*, Vol. 7, p. 100065, 2024.
- [5] H. Jiang, Q. Wu, X. Luo, D. Li, C.-Y. Lin, Y. Yang, and L. Qiu, “Longllmlingua: Accelerating and enhancing llms in long context scenarios via prompt compression,” *arXiv preprint arXiv:2310.06839*, 2023.
- [6] W. Song, S. Oh, S. Mo, J. Kim, S. Yun, J.-W. Ha, and J. Shin, “Hierarchical context merging: Better long context understanding for pre-trained llms,” *arXiv preprint arXiv:2404.10308*, 2024.