KNOWLEDGE GROUNDING IN LANGUAGE MODELS: AN EMPIRICAL STUDY

Martin Fixman[♠]**,Tillman Weyde**[♠]**,Chenxi Whitehouse**[♠]**,Pranava Madhyastha**[♠]
♠ City St. George's, University of London

ABSTRACT

Large language models, which exploded in quality and prevalence, have the propensity to produce hallucinations which presents a critical challenge where precision and correctness are crucial. Retrieval-Augmented Generation (RAG) has been proposed as as solution to this problem, but still presents problems as it's unclear when a large language model chooses to generate answers using the context provided by RAG over the knowledge on its parametric memory. We explore the *knowledge grounding* of large language models to understand the source chosen by the model when presented with a prompt that contains a context with information that contradicts its parametric knowledge.

Our findings show that smaller models and Seq2Seq are biased towards choosing knowledge from the context over larger and Decoder-only models. Retrieved information about the *perplexity* of an answer is used to create a predictor of the source of an answer.

This is an example of an ICLR2025 reference (Huang et al., 2023).

REFERENCES

REFERENCES

Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions, 2023.