**Group Members: Brian Mann, Jaulie Goe, Rebecca Glick, Ryan Ghayour**

**Enhancing Retrieval with LLM-Driven Knowledge Graphs**

A knowledge graph (KG) is a structured representation of named entities and their relationships, offering a semantically rich framework for organizing and retrieving data. They play a crucial role in structuring information, enabling more efficient data management and retrieval. KGs have recently gained in popularity as a key component of the Retrieval-Augmented Generation (RAG) architecture, an AI framework that uses information retrieval from external documents to generate responses. Integration of KGs into the RAG framework improves relevant document retrieval performance (Pan et al., 2024). Generating a knowledge graph on a corpus is non-trivial, often requiring many hours of human annotation, but LLMs have recently been adapted to improve this time-intensive task and extract relations from text (Melnyk et al., 2022). LLMs have also been used for automating KG generation by prompting an LLM to extract entities, tuples, and relations from text.

We propose that by iteratively prompting an LLM, we can generate richer knowledge graphs that better aid a RAG system in retrieval on the PubMedQA dataset than systems using a single-prompt generated knowledge graph or systems using vector databases.

There have been uses of iteratively prompting an LLM to build knowledge graphs, which have had a variety of approaches to model evaluation, such as comparing steps of the pipeline to baseline models for specific tasks (Zhang and Soh, 2024), employing human assessors to evaluate the quality of retrieved documents (Carta et al., 2023), or evaluating on reasoning benchmark datasets as opposed to retrieval benchmark datasets (Wei et al., 2023). As it stands, there has not been research focused specifically on evaluating the document retrieval capabilities of a RAG model using a KG constructed by iteratively prompting an LLM.

To evaluate our KG-RAG model, we will use the PubMedQA dataset, a retrieval benchmark dataset in the health sciences domain. We will be using LangChain to construct a RAG model with Llama, inserting the KG or vector database that we create each time. We will likely be using Llama as the LLM for knowledge graph creation as well, but we will be experimenting with various LLMs in the early stages of the project. As the bulk of our work will focus specifically on the knowledge graph creation and evaluation steps, it will be a feasible semester-long project given the available resources.

# Bibliography


[1] S. Carta, A. Giuliani, L. Piano, A. S. Podda, L. Pompianu, and S. G. Tiddia, "Iterative Zero-Shot LLM Prompting for Knowledge Graph Construction," Jul. 03, 2023, *arXiv*: arXiv:2307.01128. doi: 10.48550/arXiv.2307.01128.

[2] I. Melnyk, P. Dognin, and P. Das, "Knowledge Graph Generation From Text," Nov. 18, 2022, *arXiv*: arXiv:2211.10511. doi: 10.48550/arXiv.2211.10511.

[3] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu, "Unifying Large Language Models and Knowledge Graphs: A Roadmap," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 7, pp. 3580–3599, Jul. 2024, doi: 10.1109/TKDE.2024.3352100.

[4] J. Wei *et al.*, "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," Jan. 10, 2023, *arXiv*: arXiv:2201.11903. doi: 10.48550/arXiv.2201.11903.

[5] B. Zhang and H. Soh, "Extract, Define, Canonicalize: An LLM-based Framework for Knowledge Graph Construction," Oct. 02, 2024, *arXiv*: arXiv:2404.03868. doi: 10.48550/arXiv.2404.03868.

[6] Jin, Q., Dhingra, B., Liu, Z., Cohen, W., & Lu, X. (2019). PubMedQA: A Dataset for Biomedical Research Question Answering. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2567–2577.