Project Proposal: Knowledge and Language Course - 2025

José Matos[1] , **Vasco Rodrigues**[1]
[1]University of Coimbra, Department of Informatics Engineering
{josematos, vascorodrigues}@student.dei.uc.pt,

1 Context

Large Language Models (LLMs) are widely known for their strong capabilities in a broad range of topics, making them very powerful for day-to-day use. However, when applied to specialized domains like science, they tend to hallucinate and generate wrong outputs. A known way of mitigating hallucinations in these domains is by utilizing Knowledge Injection (KI) as a way to ground the LLM's responses by providing it with the specialized knowledge needed to perform the requested tasks. Different ways to perform KI into LLMs exist, with different levels of complexity. Our goal is to compare different types of KI, and assess how they affect hallucinations in the LLM's response, by quantifying the rate at which they occur.

With this work we are looking to answer to the following questions:

- **RQ1** - Is there a significant difference between the tested methods of KI?
- **RQ2** - Can hallucinations be quantified, and if so, can they be used as a way to benchmark different types of KI in the scientific domain?

2 Proposed Methodology

1. Gather a "golden standard" Knowledge Graph for each document in the set of documents (either through an available dataset, or perform KG extraction) ;
2. Generate a set of questions for each document, that can be answered by having access to it;
3. Prompt the LLM with the questions, and inject document related knowledge at inference time, through various different techniques (RAG, Parametric RAG, etc.);
4. From the LLM's response, extract a KG automatically;
5. Assess hallucination in the response through a technique that leverages both KGs, like RefChecker [Hu *et al.*, 2024].

3 Tool and Data Use

The SciER dataset [Zhang *et al.*, 2024] contains both entity annotation and relation annotation for scientific papers, covering three different types of entities within these papers: Dataset, Method, Task.

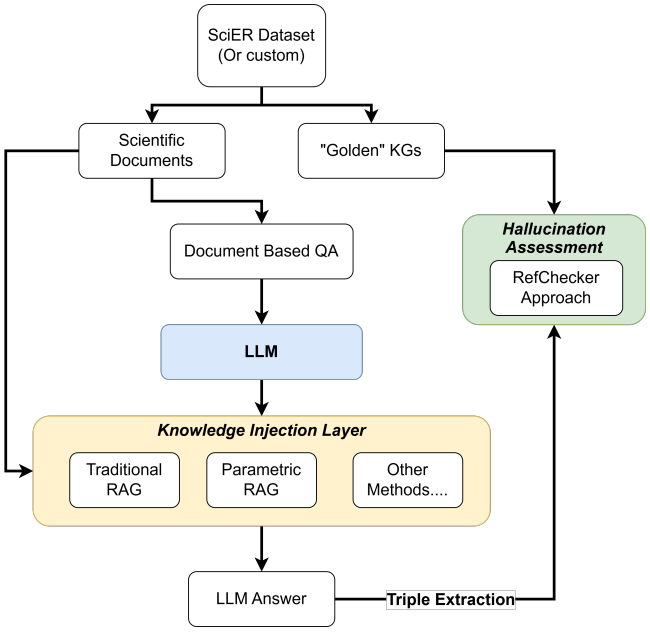

Figure 1: Diagram of proposed methodology.

For LLM use, we aim to use Ollama for local inference with models from the Qwen3 family [Yang *et al.*, 2025].

For Knowledge Injection, LangChain could be an interesting tool and one we are hoping to explore.

As for the Hallucination Assessment module, we are looking to use the RefChecker framework [Hu *et al.*, 2024], either by using the tool in the provided GitHub directory, or a similar custom approach.

4 Project Checkpoint Dates

- Tuesday, 11th of November, 11:00AM
- Tuesday, 25th of November, 11:00AM

References

[Hu *et al.*, 2024] Xiangkun Hu, Dongyu Ru, Lin Qiu, Qipeng Guo, Tianhang Zhang, Yang Xu, Yun Luo, Pengfei Liu, Yue Zhang, and Zheng Zhang. Knowledge-centric hallucination detection. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 6953–6975, Miami, Florida, USA, November 2024. Association for Computational Linguistics.

[Yang *et al.*, 2025] An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 Technical Report, May 2025. arXiv:2505.09388 [cs].

[Zhang *et al.*, 2024] Qi Zhang, Zhijia Chen, Huitong Pan, Cornelia Caragea, Longin Jan Latecki, and Eduard Dragut. SciER: An Entity and Relation Extraction Dataset for Datasets, Methods, and Tasks in Scientific Documents, October 2024. arXiv:2410.21155 [cs].