Technische Hochschule Brandenburg
University of Applied Sciences

ACL 2025 VIENNA JULY 27 - AUGUST 1

Mitigating Hallucination by Integrating Knowledge Graphs into LLM Inference – a Systematic Literature Review

Robin Wagner, Emanuel Kitzelmann, Ingo Boersch – Brandenburg University of Applied Sciences

Background

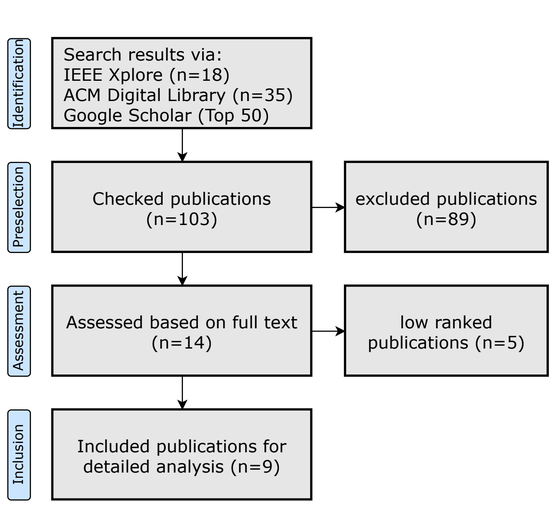
Large Language Models (LLMs) demonstrate strong performance on various natural language tasks but tend to **hallucinate** plausible but factually incorrect answers, reducing their applicability in sensitive domains such as medicine. **Retrieval-Augmented Generation (RAG)** with **Knowledge Graphs (KG)** can help mitigate hallucinations by providing the LLM with necessary context.

Research Questions

- How can KGs be integrated into LLM inference to mitigate hallucinations?
- What is the structure of the integrated KGs and where do they come from?
- To what extent does the integration of KGs improve the quality of LLM answers?
- What other advantages does the integration of KGs have?
- What challenges arise when integrating KGs?

Methodology

A **systematic literature review** was conducted on methods integrating KGs into LLM inference with the clear aim of mitigating hallucinations. Search strings were constructed to find relevant work with **IEEE Xplore**, **ACM Digital Library** and **Google Scholar**. Publications had to be **English** primary literature, **peer-reviewed or cited more than 50 times**, and had to cover a method for **KG integration into LLM inference as a main topic**. Each selected publication was assigned a **score based on nine questions** about method clarity, applicability and evaluation. The nine highest-scoring publications were selected for in-depth analysis and synthesis.



Approaches

Approach	Extraction from Input	Entry into KG	Traversing the KG	Final Context
Fang et al., 2024	Entity, relation	Semantic similarity (entity, relation)	Relation	N/A
Luo et al., 2023	Entity, relation path	Directly via entity	Relation path	Reasoning paths
Guo et al., 2024	Entity, number of hops, question variants	Directly via entity	Iterative selection of the most relevant relation up to predicted hop depth	Verbalized triples
Sun et al., 2023	Entities	Directly via entities	Iterative selection of the most relevant relation until LLM terminates	Reasoning paths
Kim et al., 2024	N/A	Semantic similarity (question concept)	All adjacency relations, random walk	Reasoning paths
Zhu et al., 2024	Patient features, diseases	Semantic similarity (features, diseases)	All adjacency relations of connected diseases	Patient features, diseases mentioned, diseases found with definition, description, info triplets
Xu et al., 2024	Entities	Semantic similarity (entities)	All adjacency relations	Verbalized triples
Ye et al., 2024	Two entities	Directly via entity	Relation path from one entity to the other	Naive answer, reasoning path
Kang et al., 2024	Entities	Subgraph creation	Iterative inclusion of entities with high information gain in subgraph	Classification, subgraph

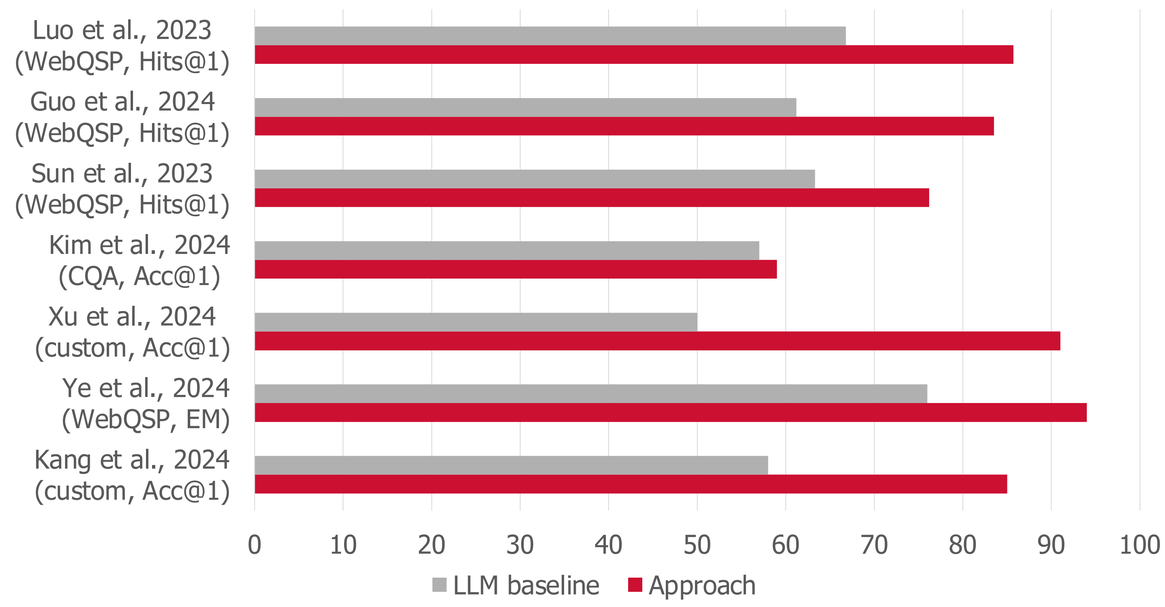
All approaches start with the **extraction of entities from the input** with an LLM. Some methods also extract additional insights, like a relation path to apply to the KG or a hop number for later traversal. Some approaches use extracted **entities directly as starting points** in the KG while others use **vector similarity to find semantically similar entities**. **KG traversal varies strongly** by method. Lightweight approaches apply an extracted relation path from the starting entity or extract adjacent nodes based on the KG structure. More computationally intensive approaches traverse the KG step-by-step by letting an LLM decide which adjacent relation or entity is most relevant to the question. **Prompt Engineering** is crucial to insert derived triples or reasoning paths as context in the LLM prompt for answering the query.

Used Knowledge Graphs

Most approaches use popular, publicy accessible KGs like **Freebase**, **WikiData** (general knowledge) or **ConceptNet** (semantic word relationships) or **PrimeKG** (disease knowledge). Some approaches **constructed their own domain-specific KG**, modelling a car manual, traditional Chinese folklore or public complaints. KGs tend to have a **simple structure**. Some use classes or specify constraints for certain relations, but none are based on formal, e.g., description logics.

Benchmarks

The analyzed publications use various benchmarks to demonstrate the factuality and reasoning improvement achieved through their approach. Most use **Knowledge Base Question Answering benchmarks** which evaluate systems that answer natural-language questions based on a given knowledge base, including **WebQuestionsSP** (WebQSP), **ComplexWebQuestions** and **SimpleQuestions**. Three studies created their **own benchmarks** by commissioning test subjects to formulate questions or by extracting questions from defined databases and websites. The benchmark scores show that KG integration improves the performance of LLMs for different types of questions, with **KBQA improvements ranging from 4% to 320%**. Improvements in benchmarks with complex questions imply, that providing an LLM with reasoning paths from a KG not only improves its factual grounding but also its reasoning capabilities.



Discussion & Conclusion

LLM hallucinations can be mitigated through KG integration in various ways, with most analyzed approaches relying on **shallow traversal** and **semantic similarity**. Further advantages of KG integration include **improved reasoning** capability through reasoning paths, **efficient inclusion of new knowledge** without having to retrain the LLM, and **improved explainability of results** thanks to explicit KG knowledge. Two major challenges of KG integration are described in the analyzed publications: **Incorrect traversal** when the question demands a long reasoning chain or when an LLM agent for step-by-step KG traversal is offered too many adjacency relations at once, and **computational complexity** due to several LLM requests before the generation of the final response. While the approaches mostly rely on relatively shallow traversal and semantic similarity, we advocate for **integrating LLMs with symbolic reasoners** to improve inference quality and explainability. This could include translation of natural language queries into formal query languages (e. g. SPARQL, Cypher), deeper exploitation of graph schema (e. g. property constraints) and ontological reasoning based on logical axioms (e. g. transitivity, subclasses).

References

- Fang, Y., Chen, Y., Jiang, Z., Xiao, J., & Ge, Y. (2024). Effective and Reliable Domain-Specific Knowledge Question Answering. *2024 IEEE International Conference on E-Business Engineering (ICEBE)*, 238–243.
- Guo, T., Yang, Q., Wang, C., Liu, Y., Li, P., Tang, J., Li, D., & Wen, Y. (2024). KnowledgeNavigator: Leveraging large language models for enhanced reasoning over knowledge graph. *Complex & Intelligent Systems*, *10*(5), 7063–7076.
- Kang, J., Pan, W., Zhang, T., Wang, Z., Yang, S., Wang, Z., Wang, J., & Niu, X. (2024). Correcting Factuality Hallucination in Complaint Large Language Model via Entity-Augmented. *2024 International Joint Conference on Neural Networks (IJCNN)*, 1–8.
- Kim, Y., Kang, E., Kim, J., & Huang, H. H. (2024). *Causal Reasoning in Large Language Models: A Knowledge Graph Approach* (No. arXiv:2410.11588).
- Luo, L., Li, Y.-F., Haffari, G., & Pan, S. (2024). *Reasoning on Graphs: Faithful and Interpretable Large Language Model Reasoning* (No. arXiv:2310.01061).
- Sun, J., Xu, C., Tang, L., Wang, S., Lin, C., Gong, Y., Ni, L. M., Shum, H.-Y., & Guo, J. (2024). Think-on-Graph: Deep and Responsible Reasoning of Large Language Model on Knowledge Graph. *Proceedings of the International Conference on Learning Representations (ICLR 2024).*
- Xu, J., Zhang, H., Zhang, H., Lu, J., & Xiao, G. (2024). ChatTf: A Knowledge Graph-Enhanced Intelligent Q&A System for Mitigating Factuality Hallucinations in Traditional Folklore. *IEEE Access*, *12*, 162638–162650.
- Ye, W., Zhang, Q., Zhou, X., Hu, W., Tian, C., & Cheng, J. (2024). Correcting Factual Errors in LLMs via Inference Paths Based on Knowledge Graph. *2024 International Conference on Computational Linguistics and Natural Language Processing (CLNLP)*, 12–16.
- Zhu, Y., Ren, C., Wang, Z., Zheng, X., Xie, S., Feng, J., Zhu, X., Li, Z., Ma, L., & Pan, C. (2024). EMERGE: Enhancing Multimodal Electronic Health Records Predictive Modeling with Retrieval-Augmented Generation. *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 3549–3559.