Tutorial Paper

Unifying Large Language Models and Knowledge Graphs for Question Answering: Recent Advances and Opportunities

Chuangtao Ma
Aalborg University
Aalborg, Denmark
chuma@cs.aau.dk

Yongrui Chen
Southeast University
Nanjing, China
yongruichen@seu.edu.cn

Tianxing Wu
Southeast University
Nanjing, China
tianxingwu@seu.edu.cn

Arijit Khan
Aalborg University
Aalborg, Denmark
arijitk@cs.aau.dk

Haofen Wang
Tongji University
Shanghai, China
haofen.wang@tongji.edu.cn

ABSTRACT

Large language models (LLMs) have demonstrated remarkable performance on several question-answering (QA) tasks because of their superior capabilities in natural language understanding and generation. On the other hand, due to poor reasoning capacity, outdated or lack of domain knowledge, expensive retraining costs, and limited context lengths of LLMs, LLM-based QA methods struggle with complex QA tasks such as multi-hop QAs and long-context QAs. Knowledge graphs (KGs) store graph-based structured knowledge which are effective for reasoning and interpretability since KGs accumulate and convey explicit relationships-based factual and domain-specific knowledge from the real world. To address the challenges and limitations of LLM-based QA, several research works that unify LLMs+KGs for QA have been proposed recently. This tutorial aims to furnish an overview of the state-of-the-art advances in unifying LLMs with KGs for QA, by categorizing them into three groups according to the roles of KGs when unifying with LLMs: **(1)** KGs as background knowledge, **(2)** KGs as reasoning guidelines, **(3)** KGs as refiners and validators. The metrics and benchmarking datasets for evaluating the methods of LLMs+KGs for QA are presented, and domain-specific industry applications and demonstrations will be showcased. The open challenges are summarized and the opportunities for data management are highlighted.

1 MOTIVATION AND RELEVANCE

Question answering (QA) is essential in natural language processing, machine learning, information retrieval, and data management areas with a wide range of applications such as web search, open-domain QA, text and knowledge base querying, fact checking, customer service assistants, and chatbots, among others. The recent pre-trained language models (PLMs) and LLMs have shown strong performance in QA tasks, but they are incapable of handling complex QA due to their limited reasoning ability, lack of up-to-date or domain knowledge, and hallucination of LLMs. To address the challenges and limitations of LLM-based QA methods in complex QA, the roadmap of unifying LLMs with KGs for knowledge-intensive tasks is proposed [35]. Considering the popularity and mainstream adoptions of both LLMs and KGs and due to the wide applications of QA including query processing over databases [38], our tutorial is timely and relevant. This tutorial is intended for participants working in the broader area of LLMs, KGs, graph learning, information retrieval, and knowledge-augmented models from both academia and industry.

Why EDBT. The EDBT conference is an established and prestigious forum for the exchange of the latest research results in data management as well as for extending database technology. LLMs have emerged as a significant research topic within the data management and data science community, as evident by recent SIGMOD and VLDB keynotes, panels, tutorials, and workshops [6, 21, 24], Generative AI Day (KDD 24), LLM Day (WWW 24), etc. Our tutorial on unifying LLMs + KGs for QA emphasizes advanced data management techniques and integration strategies, making it highly relevant and beneficial to the interdisciplinary and broader data science research community.

2 TUTORIAL OUTLINE

This is a lecture-style tutorial, accompanied by discussions on domain-specific applications and demonstrations from industry. The outline of our tutorial is given below.

```
1. Introduction
    1.1 Motivation of QA
    1.2 Large Language Models for QA
    1.3 Knowledge Graphs for QA
    1.4 Overview of Unifying LLMs+KGs
2. Unifying LLMs with KGs for QA
    2.1 KGs as Background Knowledge
    2.2 KGs as Reasoning Guidelines
    2.3 KGs as Refiners and Validators
3. Advanced Topics on LLM+KG for QA
    3.1 Natural Language Questions to Structured Queries
    3.2 Explainable QA
    3.3 Optimization and Efficiency
4. Evaluations and Applications
    4.1 Performance Metrics
    4.2 Benchmark Datasets
    4.3 Industry Applications and Demonstrations
5. Future Directions
    5.1 Opportunities for Data Management
    5.2 Future Directions
```

The materials including covered papers, pointers to open-source codebase, datasets, and demonstrations are available on GitHub[1] for public access.

3 DESCRIPTION OF TOPICS

We categorize the methodology of unifying LLMs and KGs for QA tasks into different paradigms based on the role of KGs. Due to the lack of space, we only refer to the most relevant papers. However, this is not an exhaustive list of papers that are related and will be discussed during the tutorial.

3.1 KG as Background Knowledge

When KGs are used as background knowledge to enhance LLMs for QA, the questions are parsed to identify the relevant subgraphs from KGs, then they are integrated with LLMs based on knowledge fusion and retrieval-augmented generation (RAG).

Knowledge Integration and Fusion. Knowledge integration and fusion aims to enhance LLMs by integrating unknown knowledge into LLMs for knowledge-intensive tasks. In the phase of pre-training, the KGs and text are aligned (via local subgraph

[1]https://github.com/machuangtao/LLM-KG4QA

extraction and entity linking) and interacted to jointly train the language models for complex QA tasks [50]. To address knowledge forgetting during knowledge integration, InfuserKI [44] introduces the adaptive selection of the new knowledge that is integrated with LLMs. Fine-tuning LLMs with input text and knowledge graphs is another paradigm, as it can refine and improve their performance on domain-specified tasks. KG-Adapter [43] improves parameter-efficient fine-tuning of LLMs by introducing a knowledge adaptation layer to LLMs. GAIL [55] fine-tunes LLMs for lightweight KGQA models based on retrieved SPARQL-question pairs from KGs.

Retrieval Augmented Generation (RAG). RAG serves as a retrieval mechanism to retrieve relevant knowledge from the domain-specific knowledge organized in the form of text chunks, and augments the capability of LLMs by integrating the retrieved context with LLMs. However, the mainstream RAG methods retrieve the relevant knowledge from the embeddings of textual chunks, which ignores the structured information and interrelations of these textual chunks. To mitigate this limitation, Graph RAG [16, 26] is proposed. Instead of retrieving the knowledge from textual chunks, Graph RAG directly retrieves the relevant knowledge from graph data. Then it integrates the retrieved and pruned textual subgraphs with query by aggregating and aligning the graph embeddings with text vectors based on Graph Neural Networks (GNNs).

3.2 KGs as Reasoning Guidelines

KGs can serve as guidelines to LLMs for QA by providing structured factual knowledge. By integrating KGs, LLMs can access precise information and logical connections between concepts, thereby enhancing their ability to provide accurate and contextually relevant answers. Recent methods for integrating KG guidelines into LLM reasoning can be classified into three categories.

Offline KG Guidelines. In this paradigm, KG supplies potential subgraphs before the reasoning process of LLM. Then LLM selects the most relevant path for reasoning based on its existing knowledge. EtD [27] uses a lightweight GNN to extract fine-grained knowledge for creating knowledge-enhanced prompts, guiding a frozen LLM to determine answers. Recent studies have been exploring the application of novel formats of guidelines. GCR [32] transforms a KG into a KG-Trie for efficient reasoning path search and employs graph-constrained decoding with a specialized LLM to generate reasoning paths and answers.

Online KG Guidelines. This paradigm emphasizes that the guidance of the KG is directly involved in the reasoning process of LLMs. In each reasoning step, LLM needs to first retrieve the necessary knowledge from the KG and then makes a decision for the next step based on the retrieved knowledge. Oreo [17] uses contextualized random walks on KGs for single-step reasoning. LLM-ARK [19] treats reasoning as sequential decision-making optimized via Proximal Policy Optimization (PPO). ToG [39] enables LLMs to iteratively perform beam search on KGs to identify optimal reasoning paths and outcomes.

Agent-based KG Guidelines. KGs can also be integrated into the reasoning processes of LLMs as a component within an Agent. This integration allows the Agent to leverage structured knowledge for enhanced decision-making and problem-solving capabilities. KG-Agent [22] integrates LLM as a multifunctional toolbox with a KG-based executor and a knowledge memory system. It develops an iterative mechanism that autonomously selects tools and updates the memory to enhance reasoning over KGs. ODA [40] incorporates KG reasoning capabilities through a global observation approach, which improves reasoning abilities by employing a cyclical paradigm of observation, action, and reflection.

3.3 KGs as Refiners and Validators

KGs can enhance LLMs in QA tasks by serving as refiners and validators, providing structured knowledge to verify answers against factual knowledge. This integration helps filter and refine responses to improve precision and contextual relevance.

KG-Driven Filtering and Validation. KGs enhance the accuracy and reliability of LLM outputs by filtering and validating candidate answers through structured and verified information. For instance, ACT-Selection [37] filters and re-ranks answer candidates based on their types extracted from Wikidata. KGs contribute to improving factual accuracy, as demonstrated by KG-Rank [49], which integrates medical KGs with re-ranking techniques to increase the credibility of generated responses. Moreover, KGR [11] autonomously extracts and validates factual statements in model outputs, significantly boosting performance on factual QA benchmarks.

KG-Augmented Output Refinement. KGs are essential for enhancing the outputs of LLMs by integrating structured knowledge that enables LLMs to refine their responses for greater clarity and accuracy. EFSUM [25] optimizes an open-source LLM as a fact summarizer to generate relevant summaries from KGs, thereby improving performance in zero-shot QA. InteractiveKBQA [47] facilitates iterative interactions with the knowledge base, enabling LLMs to generate logical forms and refine outputs based on user feedback. Additionally, LPKG [45] improves the planning capabilities of LLMs by fine-tuning them with planning data derived from KGs, thus enabling more sophisticated reasoning in complex QA.

3.4 Advanced Topics

Recent advancements in unifying LLMs and KGs for QA have been applied to areas, e.g., explainable QA [9], visual QA [5], QA over multiple documents [46], and conversational QA [28]. However, these approaches face bottlenecks such as low efficiency and high computational costs due to large-scale graph reasoning and the processing of heterogeneous multi-modal data. To tackle these challenges, optimization techniques, e.g., index-based optimization [52], prompting-based optimization [46], and cost-based optimizations [4] have been introduced, significantly improving performance and scalability.

3.5 Evaluations and Applications

Metrics and Dataset. We summarize the evaluation metrics in unifying LLMs with KGs for QA: **(1)** the metrics measuring the retrieval of RAG, context relevance, precision, context recall [51]; **(2)** the metrics measuring the relevance of the generated answers, BERTScore and MRR (Mean Reciprocal Rank) [36], faithfulness, answer relevance, and context relevance [7]; **(3)** the metrics measuring the correctness of intermediate reasoning path for multi-hop QA, Hop-Acc [10]. The recent benchmark datasets are: **(1)** Complex QA – PATQA[33], MINTQA [14], MedQA [23]; **(2)** KBQA and KGQA – WebQSP [42], CAQA[15], CR-LT KGQA [13]; **(3)** LLM and KGs for QA – KGs+LLMs for QA [38], XplainLLM[2], LLM-KG-Bench [34].

Industrial Applications. We demonstrate domain-specified applications from industry in unifying LLMs+KGs for QA: **(1)** KAG (by Antgroup)[2] is a newly released domain-knowledge augmented generation framework that leverages KGs and vector retrieval to bi-directionally enhance LLMs for knowledge-intensive tasks such as QA for e-government and e-health; **(2)** Graph RAG (by NebulaGraph)[3] is an industrial demo of Graph RAG integrating NLP2Cypher-based KG query engine, vector RAG query engine, and Graph vector RAG query engine.

3.6 Opportunities for Data Management

The unification of LLMs and KGs provides exciting data management research opportunities across multiple dimensions.
NLQ to Structured Query. Using KGs and ontology/schema, LLMs can enable accurate conversion of natural language queries (NLQ) into structured graph queries (e.g., SPARQL and Cypher) by leveraging structured knowledge understanding [12, 31].
Efficient and Explainable RAG. KGs offer structured and reliable information, enabling efficient retrieval and accurate reasoning for LLMs [20]. They enhance explainability by linking generated answers to explicit KG relationships, reduce hallucinations, and support domain-specific or personalized use cases.
Knowledge Alignment and Dynamic Integration. Knowledge alignment between KGs and LLMs is a critical challenge since knowledge overlap and conflicts occur when integrating new knowledge from multimodal and multiple sources into LLMs [3, 41]. In addition to knowledge conflicts, incremental updates to KGs and dynamic integration with LLMs are essential for ensuring up-to-date knowledge integration.
Automated Prompt Engineering. Structured knowledge can be extracted from KGs, prompts can be generated using multi-view templates, and they can be optimized through bias detection and feedback loops. This workflow includes querying KGs, dynamically generating prompts [54], iteratively optimizing them, and evaluating their fairness and quality.
Roles of Vector and Graph Databases. Leveraging vector DBs for graph RAG creates new challenges and opportunities such as combining graph DBs with vector DBs [30], optimizing the index creation and similarity search over large-scale graph embeddings, multi-vector search, and hardware acceleration.

3.7 Challenges and Future Directions

We conclude by discussing open challenges and future roadmap.
Effectiveness and Efficiency of Subgraph Retrieval. The efficiency of relevant subgraphs extraction and retrieval is a challenging task since the KGs cannot be integrated and fused with LLMs directly. This is because knowledge graphs usually are large-scale graphs and the context length of LLM is limited.
Security and Privacy. With the unification of domain-specific KGs in QA, privacy and security concerns naturally arise. It is important to integrate privacy-preserving techniques and access control policies to ensure that the retrieved information is authorized and to maintain the confidentiality of sensitive information.
Explainable and Fairness-Aware QA. The explainable answers for QA are mainly based on the reasoning chains over the factual graph, while the low efficiency and high computing cost of iterative reasoning over the large graph remain challenging. The Graph RAG enhances the explainability of LLM responses by tracing relevant subgraphs within KGs, while also having the potential to rectify undesired biases.
Other Data Science Applications. The combination of LLMs and KGs leverages LLMs' natural language understanding and KGs' structured knowledge to enhance applications like personalized recommendations, customer service [48], accurate medical diagnostics, and financial decision-making, which enables more intelligent and knowledge-rich solutions across domains.

[2]https://github.com/OpenSPG/KAG
[3]https://github.com/wey-gu/demo-kg-build

4 RELATED TUTORIALS

The related tutorials are summarized below.

- **QA, LLMs, and KGs.** The relevant existing tutorials on QA [1], LLMs [53], and KGs [29] are mainly focused on open-domain question answering, KG reasoning, and LLMs for recommendations. Unlike these tutorials, our tutorial focuses on the state-of-the-art in unifying LLMs+KGs for knowledge-intensive QA.
- **LLMs+Graphs (KGs) and RAG.** Several tutorials on LLMs and graphs [18], LLMs and RAG [8] have been presented to introduce the paradigms of integrating LLMs with RAG.

Our tutorial differs from the above tutorials since we discuss the recent advances and directions in unifying LLMs+KGs for QA and emphasize the opportunities for data management.

5 BIOGRAPHY

Chuangtao Ma is a postdoctoral researcher at Aalborg University, Denmark. His research focuses on knowledge graphs, knowledge-augmented models, and their applications in data management. He is a member of the management committee of the COST action on the Global Network on Large-Scale, Cross-domain, and Multilingual Open Knowledge Graphs.
Yongrui Chen is a postdoctoral researcher at Southeast University, China. He specializes in incorporating structured and semi-structured knowledge into foundational LLMs, to improve their complex knowledge reasoning capability. He has presented numerous papers at prominent venues, including NeurIPS, TKDE, IJCAI, AAAI, ACL, ISWC, and NAACL.
Tianxing Wu is an associate professor at Southeast University, China. He is one of the main contributors to build Chinese large-scale encyclopedic knowledge graph: Zhishi.me and schema knowledge graph: Linked Open Schema. He has published over 60 papers in top-tier venues, e.g., ICDE, SIGIR, ACL, AAAI, IJCAI, ECAI, ISWC, TKDE, TKDD, JWS, and WWWJ.
Arijit Khan is an IEEE senior member, an ACM distinguished speaker, and an associate professor at Aalborg University, Denmark. He published over 90 papers in premier data management and mining venues including ACM SIGMOD, VLDB, TKDE, ICDE, ICLR, SDM, USENIX ATC, EDBT, WWW, WSDM, CIKM, and TKDD. Arijit co-presented tutorials on emerging graph queries, applications, big graph systems, and graph machine learning at VLDB, DSAA, CIKM, and ICDE.
Haofen Wang is a Professor at Tongji University, China. He is one of the initiators of OpenKG, the world's largest alliance for Chinese open knowledge graphs. He published over 100 high-level papers in the AI field, and developed the world's first interactive virtual idol–"Amber Xuyan". Additionally, the intelligent customer service robots he built have served over 1 billion users.

ACKNOWLEDGMENTS

This work was supported by the Novo Nordisk Foundation grant (NNF22OC0072415), the National Natural Science Foundation of China (U23B2057, 62176185, 62376058), and the GOBLIN COST Action (CA23147).

REFERENCES

[1] Danqi Chen and Wen tau Yih. 2020. Open-domain question answering. In *ACL*. 34–37.
[2] Zichen Chen, Jianda Chen, Ambuj Singh, and Misha Sra. 2024. XplainLLM: A knowledge-augmented dataset for reliable grounded explanations in LLMs. In *EMNLP*. 7578–7596.
[3] Philipp Christmann and Gerhard Weikum. 2024. RAG-based question answering over heterogeneous data and text. *IEEE Data Eng. Bull.* 48, 4 (2024), 71–86.

[4] Junnan Dong, Qinggang Zhang, Chuang Zhou, Hao Chen, Daochen Zha, and Xiao Huang. 2024. Cost-efficient knowledge-based question answering with large language models. In *NeurIPS*.

[5] Junnan Dong, Qinggang Zhang, Huachi Zhou, Daochen Zha, Pai Zheng, and Xiao Huang. 2024. Modality-aware integration with large language models for knowledge-based visual question answering. In *ACL*. 2417–2429.

[6] Xin Luna Dong. 2024. The journey to a knowledgeable assistant with retrieval-augmented generation (RAG). In *WSDM*. 4–4.

[7] Shahul Es, Jithin James, Luis Espinosa Anke, and Steven Schockaert. 2024. RAGAs: Automated evaluation of retrieval augmented generation. In *EACL*. 150–158.

[8] Wenqi Fan, Yujuan Ding, Liangbo Ning, Shijie Wang, Hengyun Li, Dawei Yin, Tat-Seng Chua, and Qing Li. 2024. A survey on RAG meeting LLMs: Towards retrieval-augmented large language models. In *SIGKDD*. 6491–6501.

[9] Jianzhou Feng, Qin Wang, Huaxiao Qiu, and Lirong Liu. 2025. Retrieval in decoder benefits generative models for explainable complex question answering. *Neural Netw.* 181 (2025), 106833.

[10] Hengrui Gu, Kaixiong Zhou, Xiaotian Han, Ninghao Liu, Ruobing Wang, and Xin Wang. 2024. PokeMQA: Programmable knowledge editing for multi-hop question answering. In *ACL*. 8069–8083.

[11] Xinyan Guan, Yanjiang Liu, Hongyu Lin, Yaojie Lu, Ben He, Xianpei Han, and Le Sun. 2024. Mitigating large language model hallucinations via autonomous knowledge graph-based retrofitting. In *AAAI*. 18126–18134.

[12] Aibo Guo, Xinyi Li, Guanchen Xiao, Zhen Tan, and Xiang Zhao. 2022. SpCQL: A semantic parsing dataset for converting natural language into Cypher. In *CIKM*. 3973–3977.

[13] Willis Guo, Armin Toroghi, and Scott Sanner. 2024. CR-LT-KGQA: A knowledge graph question answering dataset requiring commonsense reasoning and long-tail knowledge. *arXiv:2403.01395* (2024).

[14] Jie He, Nan Hu, Wanqiu Long, Jiaoyan Chen, and Jeff Z Pan. 2024. MINTQA: A multi-hop question answering benchmark for evaluating LLMs on new and tail knowledge. *arXiv:2412.17032* (2024).

[15] Nan Hu, Jiaoyan Chen, Yike Wu, Guilin Qi, Sheng Bi, Tongtong Wu, and Jeff Z Pan. 2024. Benchmarking large language models in complex question answering attribution using knowledge graphs. *arXiv:2401.14640* (2024).

[16] Yuntong Hu, Zhihan Lei, Zheng Zhang, Bo Pan, Chen Ling, and Liang Zhao. 2024. GRAG: Graph retrieval-augmented generation. *arXiv:2405.16506* (2024).

[17] Ziniu Hu, Yichong Xu, Wenhao Yu, Shuohang Wang, Ziyi Yang, Chenguang Zhu, Kai-Wei Chang, and Yizhou Sun. 2022. Empowering language models with knowledge graph reasoning for question answering. In *EMNLP*. 9562–9581.

[18] Chao Huang, Xubin Ren, Jiabin Tang, Dawei Yin, and Nitesh Chawla. 2024. Large language models for graphs: Progresses and directions. In *WWW*. 1284–1287.

[19] Yuxuan Huang. 2023. Evaluating and enhancing large language models for conversational reasoning on knowledge graphs. *arXiv:2312.11282* (2023).

[20] Yubo Huang and Guosun Zeng. 2024. RD-P: A Trustworthy retrieval-augmented prompter with knowledge graphs for LLMs. In *CIKM*. 942–952.

[21] Madelon Hulsebos, Matteo Interlandi, and Shreya Shankar. 2024. Eighth workshop on data management for end-to-end machine learning (DEEM). In *SIGMOD Companion*.

[22] Jinhao Jiang, Kun Zhou, Wayne Xin Zhao, Yang Song, Chen Zhu, Hengshu Zhu, and Ji-Rong Wen. 2024. Kg-Agent: An efficient autonomous agent framework for complex reasoning over knowledge graph. *arXiv:2402.11163* (2024).

[23] Di Jin, Eileen Pan, Nassim Oufattole, Wei-Hung Weng, Hanyi Fang, and Peter Szolovits. 2021. What disease does this patient have? a large-scale open domain question answering dataset from medical exams. *Appl. Sci.* 11, 14 (2021), 6421.

[24] Arijit Khan, Tianxing Wu, and Xi Chen. 2024. LLM+KG: Data management opportunities in unifying large language models + knowledge graphs. In *LLM+KG Workshop@VLDB*.

[25] Sungho Ko, Hyunjin Cho, Hyungjoo Chae, Jinyoung Yeo, and Dongha Lee. 2024. Evidence-focused fact summarization for knowledge-augmented zero-shot question answering. In *EMNLP*. 10636–10651.

[26] Ernests Lavrinovics, Russa Biswas, Johannes Bjerva, and Katja Hose. 2025. Knowledge graphs, Large language models, and Hallucinations: An NLP perspective. *J. Web Semant.* 85 (2025), 100844.

[27] Guangyi Liu, Yongqi Zhang, Yong Li, and Quanming Yao. 2024. Explore then Determine: A GNN-LLM synergy framework for reasoning over knowledge graph. *arXiv:2406.01145* (2024).

[28] Lihui Liu, Blaine Hill, Boxin Du, Fei Wang, and Hanghang Tong. 2024. Conversational question answering with language models generated reformulations over knowledge graph. In *ACL*. 839–850.

[29] Lihui Liu, Zihao Wang, Jiaxin Bai, Yangqiu Song, and Hanghang Tong. 2024. New frontiers of knowledge graph reasoning: Recent advances and future trends. In *WWW*. 1294–1297.

[30] Shige Liu, Zhifang Zeng, Li Chen, Adil Ainihaer, Arun Ramasami, Songting Chen, Yu Xu, Mingxi Wu, and Jianguo Wang. 2025. TigerVector: Supporting vector search in graph databases for advanced RAGs. *arXiv:2501.11216* (2025).

[31] Yang Liu, Xin Wang, Jiake Ge, Hui Wang, Dawei Xu, and Yongzhe Jia. 2024. Text to graph query using filter condition attributes. In *SGDA Workshop@VLDB*.

[32] Linhao Luo, Zicheng Zhao, Chen Gong, Gholamreza Haffari, and Shirui Pan. 2024. Graph-constrained reasoning: Faithful reasoning on knowledge graphs with large language models. *arXiv:2410.13080* (2024).

[33] Jannat Ara Meem, Muhammad Shihab Rashid, Yue Dong, and Vagelis Hristidis. 2024. PAT-Questions: A self-updating benchmark for present-anchored temporal question-answering. In *ACL*. 13129–13148.

[34] Lars-Peter Meyer, Johannes Frey, Kurt Junghanns, Felix Brei, Kirill Bulert, Sabine Gründer-Fahrer, and Michael Martin. 2023. Developing a scalable benchmark for assessing large language models in knowledge graph engineering. In *SEMANTICS*.

[35] Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. 2024. Unifying large language models and knowledge graphs: A roadmap. *IEEE Trans. Knowl. Data Eng.* 36, 7 (2024), 3580–3599.

[36] Boci Peng, Yun Zhu, Yongchao Liu, Xiaohe Bo, Haizhou Shi, Chuntao Hong, Yan Zhang, and Siliang Tang. 2024. Graph retrieval-augmented generation: A survey. *arXiv:2408.08921* (2024).

[37] Mikhail Salnikov, Maria Lysyuk, Pavel Braslavski, Anton Razzhigaev, Valentin A Malykh, and Alexander Panchenko. 2023. Answer candidate type selection: Text-to-text language model for closed book question answering meets knowledge graphs. In *KONVENS*. 155–164.

[38] Juan Sequeda, Dean Allemang, and Bryon Jacob. 2024. A benchmark to understand the role of knowledge graphs on large language model's accuracy for question answering on enterprise SQL databases. In *GRADES-NDA@SIGMOD/PODS*. 1–12.

[39] Jiashuo Sun, Chengjin Xu, Lumingyuan Tang, Saizhuo Wang, Chen Lin, Yeyun Gong, Lionel Ni, Heung-Yeung Shum, and Jian Guo. 2024. Think-on-Graph: Deep and responsible reasoning of large language model with knowledge graph. In *ICLR*.

[40] Lei Sun, Zhengwei Tao, Youdi Li, and Hiroshi Arakawa. 2024. ODA: Observation-driven agent for integrating LLMs and knowledge graphs. In *ACL*. 7417–7431.

[41] Nan Tang, Chenyu Yang, Zhengxuan Zhang, Yuyu Luo, Ju Fan, Lei Cao, Sam Madden, and Alon Halevy. 2024. Symphony: Towards trustworthy question answering and verification using RAG over multimodal data lakes. *IEEE Data Eng. Bull.* 48, 4 (2024), 135–146.

[42] Wen tau Yih, Matthew Richardson, Chris Meek, Ming-Wei Chang, and Jina Suh. 2016. The value of semantic parse labeling for knowledge base question answering. In *ACL*. 201–206.

[43] Shiyu Tian, Yangyang Luo, Tianze Xu, Caixia Yuan, Huixing Jiang, Chen Wei, and Xiaojie Wang. 2024. KG-Adapter: Enabling knowledge graph integration in large language models through parameter-efficient fine-tuning. In *ACL*. 3813–3828.

[44] Fali Wang, Runxue Bao, Suhang Wang, Wenchao Yu, Yanchi Liu, Wei Cheng, and Haifeng Chen. 2024. InfuserKI: Enhancing large language models with knowledge graphs via infuser-guided knowledge integration. In *EMNLP*. 3675–3688.

[45] Junjie Wang, Mingyang Chen, Binbin Hu, Dan Yang, Ziqi Liu, Yue Shen, Peng Wei, Zhiqiang Zhang, Jinjie Gu, Jun Zhou, Jeff Z. Pan, Wen Zhang, and Huajun Chen. 2024. Learning to plan for retrieval-augmented large language models from knowledge graphs. In *EMNLP*. 7813–7835.

[46] Yu Wang, Nedim Lipka, Ryan A. Rossi, Alexa Siu, Ruiyi Zhang, and Tyler Derr. 2024. Knowledge graph prompting for multi-document question answering. In *AAAI*, Vol. 38. 19206–19214.

[47] Guanming Xiong, Junwei Bao, and Wen Zhao. 2024. Interactive-KBQA: Multi-turn interactions for knowledge base question answering with large language models. In *ACL*. 10561–10582.

[48] Zhentao Xu, Mark Jerome Cruz, Matthew Guevara, Tie Wang, Manasi Deshpande, Xiaofeng Wang, and Zheng Li. 2024. Retrieval-augmented generation with knowledge graphs for customer service question answering. In *SIGIR*. 2905–2909.

[49] Rui Yang, Haoran Liu, Edison Marrese-Taylor, Qingcheng Zeng, Yuhe Ke, Wanxin Li, Lechao Cheng, Qingyu Chen, James Caverlee, Yutaka Matsuo, and Irene Li. 2024. KG-Rank: Enhancing large language models for medical QA with knowledge graphs and ranking techniques. In *BioNLP Workshop@ACL*. 155–166.

[50] Michihiro Yasunaga, Antoine Bosselut, Hongyu Ren, Xikun Zhang, Christopher D Manning, Percy S Liang, and Jure Leskovec. 2022. Deep bidirectional language-knowledge graph pretraining. In *NeurIPS*. 37309–37323.

[51] Hao Yu, Aoran Gan, Kai Zhang, Shiwei Tong, Qi Liu, and Zhaofeng Liu. 2024. Evaluation of retrieval-augmented generation: A survey. In *CCF Conference on Big Data*. 102–120.

[52] Haozhen Zhang, Tao Feng, and Jiaxuan You. 2024. Graph of records: Boosting retrieval augmented generation for long-context summarization with graphs. *arXiv:2410.11001* (2024).

[53] Jizhi Zhang, Keqin Bao, Yang Zhang, Wenjie Wang, Fuli Feng, and Xiangnan He. 2024. Large language models for recommendation: Progresses and future directions. In *WWW Companion* (2024). 1268–1271.

[54] Qinggang Zhang, Junnan Dong, Hao Chen, Daochen Zha, Zailiang Yu, and Xiao Huang. 2024. KnowGPT: Knowledge graph based prompting for large language models. In *NeurIPS*.

[55] Zhiqiang Zhang, Liqiang Wen, and Wen Zhao. 2024. A GAIL fine-tuned LLM enhanced framework for low-resource knowledge graph question answering. In *CIKM*. 3300–3309.