Knowledge Graphs for
Responsible AI Workshop

# The Role Evolution of KGs in Synthesizing with LLMs: From Background Knowledge to Joint Reasoning

Chuangtao Ma

Aalborg University
Denmark

# Outline

- Introduction
  - Comparison of LLMs and KGs
  - Why synthesize KGs with LLMs?
- Role Evolution of KGs in Synthesizing with LLMs
  - KGs as Background Knowledge
  - KGs as Reasoning Guidelines
  - Comparison of Approaches with Different Roles of KGs
- LLMs + KGs Application
  - Questions Answering
  - Schema Matching
- Conclusion

# Introduction

- Comparison of LLMs and KGs

| **Feature** | **Knowledge Graphs (KGs)** | **Large Language Models (LLMs)** |
|---|---|---|
| Data Structure | Structured, graph-based (triples) | Unstructured text-based, sequential tokens |
| Knowledge Type | Explicit, factual, domain-specific knowledge | Implicit, parametric, commonsense knowledge |
| Processing Style | Logical reasoning, graph query, path traversal | Intuitive, implicit, next token prediction |
| Primary Use Case | KGQA, recommendation, entity disambiguation … | QA, content generation, text summarization … |

# Introduction

- Why Do We need to synthesize LLMs with KGs?
  - Large Language Models (LLMs)
    - **Strengths**: Superior Capabilities in Natural Language Processing and Generation …
    - **Limitations**: Hallucination, Poor Reasoning, Lack of up-to-date Domain Knowledge, Black-box Model …
  - Knowledge Graphs (KGs)
    - **Strengths**: Highly curated Structured and Reliable Knowledge, Symbolic Reasoning and Inference …
    - **Limitations**: High cost of construction, Incompleteness, Domain-specified …
  - LLMs + KGs
    - **Neuro-symbolic System** : Hallucination Mitigation, Reasoning, Explainable and Responsible Results.
    - **Knowledge Fusion** : Combinations of Domain Specific factual Knowledge and Common Knowledge.

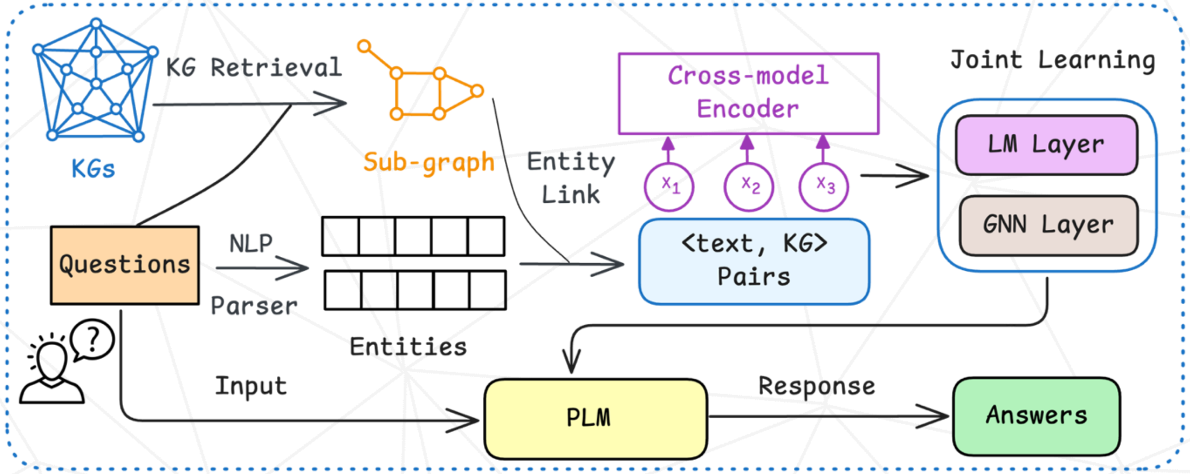
https://enterprise-knowledge.com/synergizing-knowledge-graphs-with-large-language-models-llms/

# Role Evolution of KGs in Synthesizing with LLMs

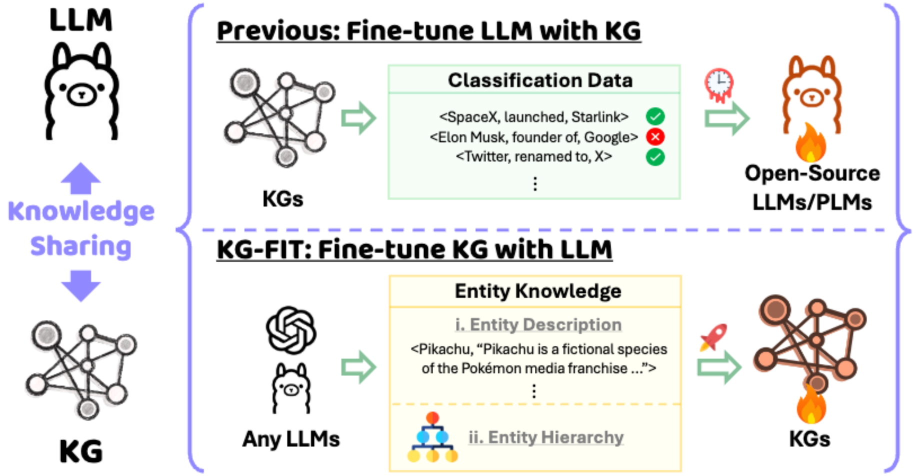
- KGs as Background Knowledge
  - **KGs and Text Alignment**
    - KGs (entities or relations) having the textual description (Text-KG pair)
    - KG and text data are stored separately (Common scenario for QA task)

Entity Linking

KG Retrieval




Joint learning

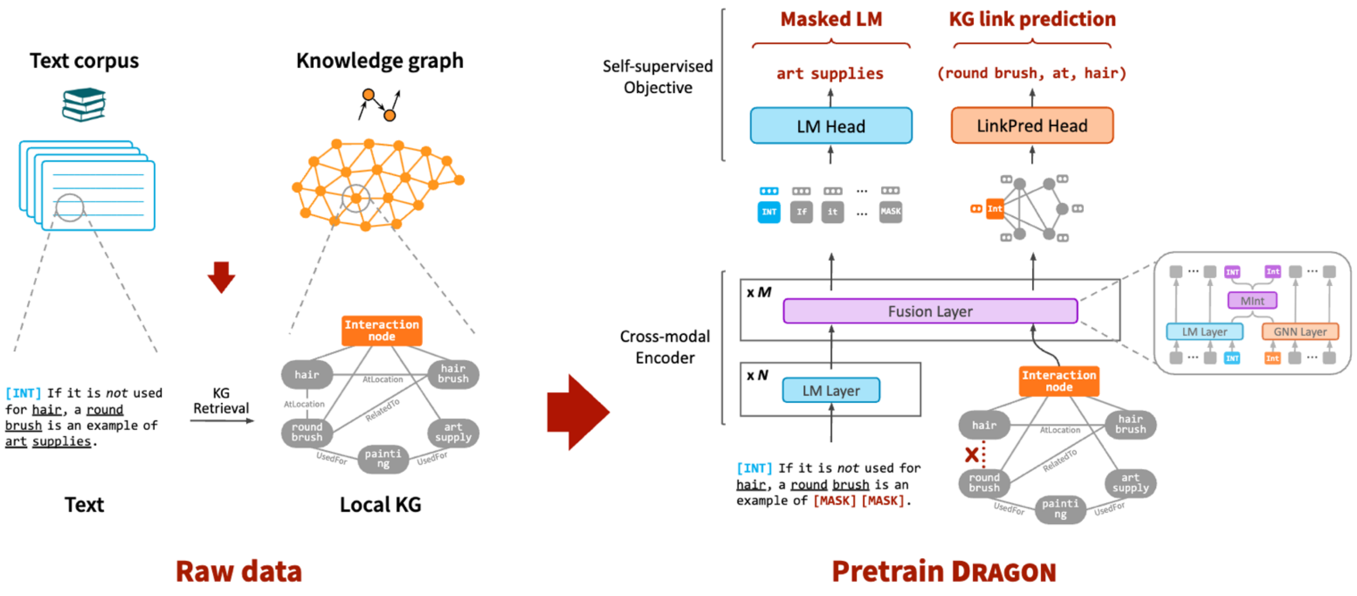



Fine-tuning

Jiang, Pengcheng, et al. "KG-FIT: Knowledge graph fine-tuning upon open-world knowledge." *Advances in Neural Information Processing Systems* 37 (2024): 136220-136258.

# Role Evolution of KGs in Synthesizing with LLMs

- Joint Learning: Bidirectional language and KG pretraining [NeurIPS2022]
  - Retrieving relevant subgraph from KG based on text to create **text-KG pair**.
  - Leveraging cross-modal encoder that fuses the input **text-KG pair** bidirectionally.
  - Unifying masked LM and KG link prediction for and joint learning reasoning.




Modality interaction (Mint) with interaction token and node to mix representation for joint learning

$$[\boldsymbol{h}_{int}^{(\ell)}; \boldsymbol{e}_{int}^{(\ell)}] = \mathrm{MInt}([\tilde{\boldsymbol{h}}_{int}^{(\ell)}; \tilde{\boldsymbol{e}}_{int}^{(\ell)}]),$$

Yasunaga, Michihiro, et al. "Deep bidirectional language-knowledge graph pretraining." *Advances in Neural Information Processing Systems* 35 (2022): 37309-37323.

# Role Evolution of KGs in Synthesizing with LLMs

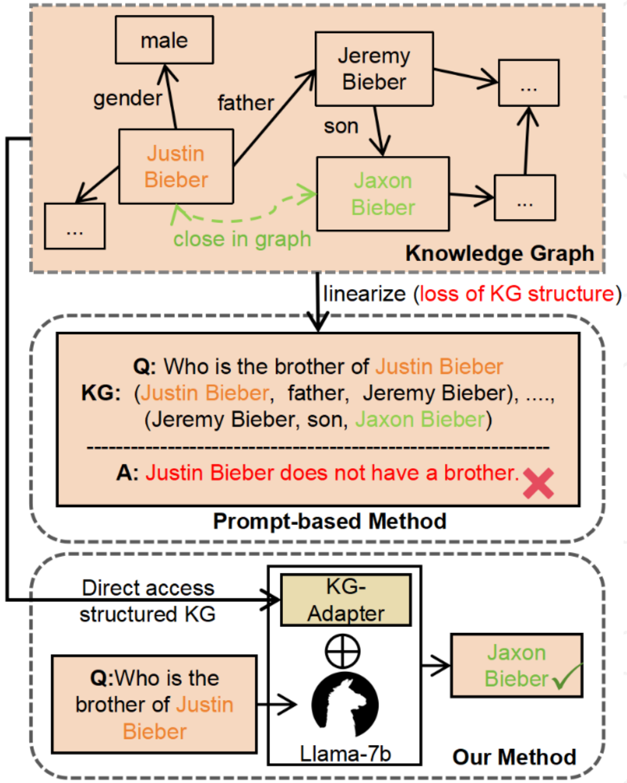
- Fine-Tuning LLMs with KGs

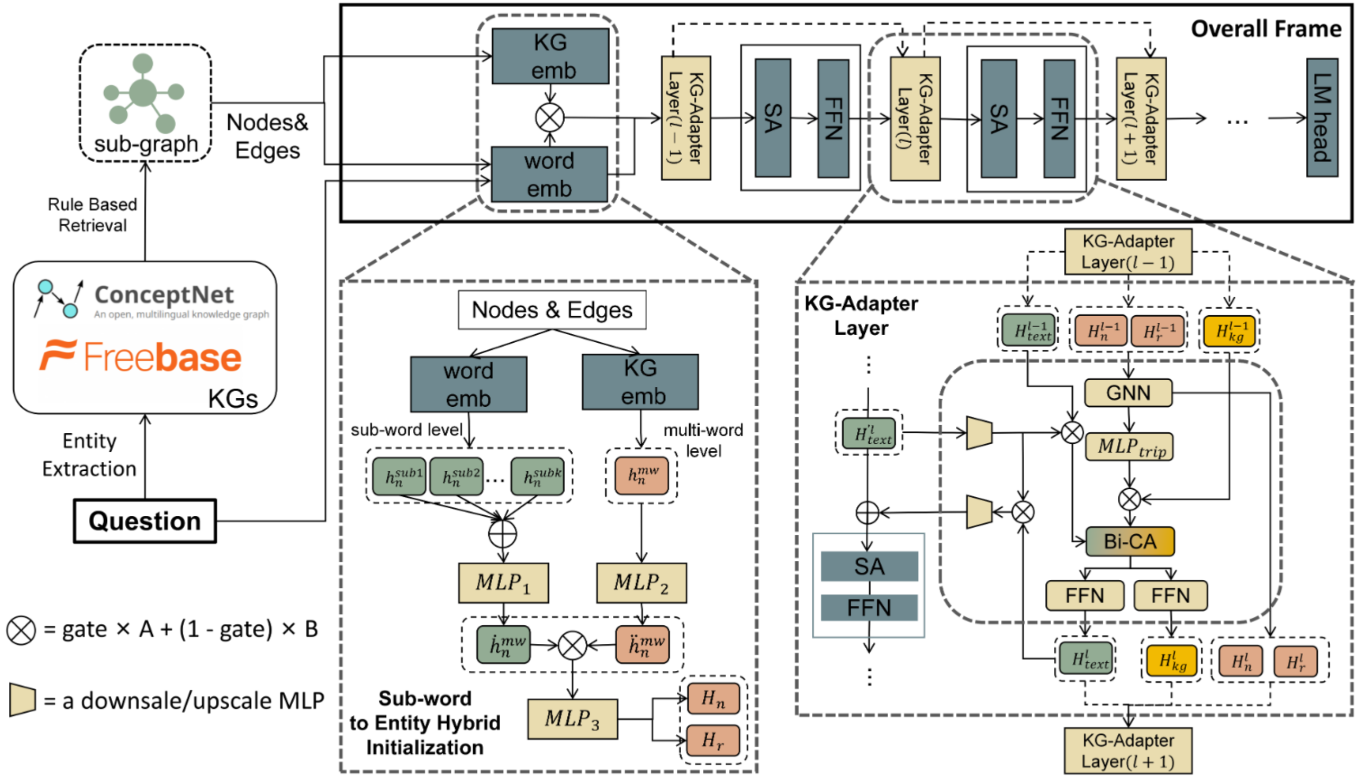
-KG-Adpter: Parameter-efficient fine-tuning (PEFT) for integrating KGs with LLM [ACL 2024]

Fuse representations of entity-level KG and sub-word KG

GNN-based KGs representations updating

Bi-directional cross- attention based reconstruction



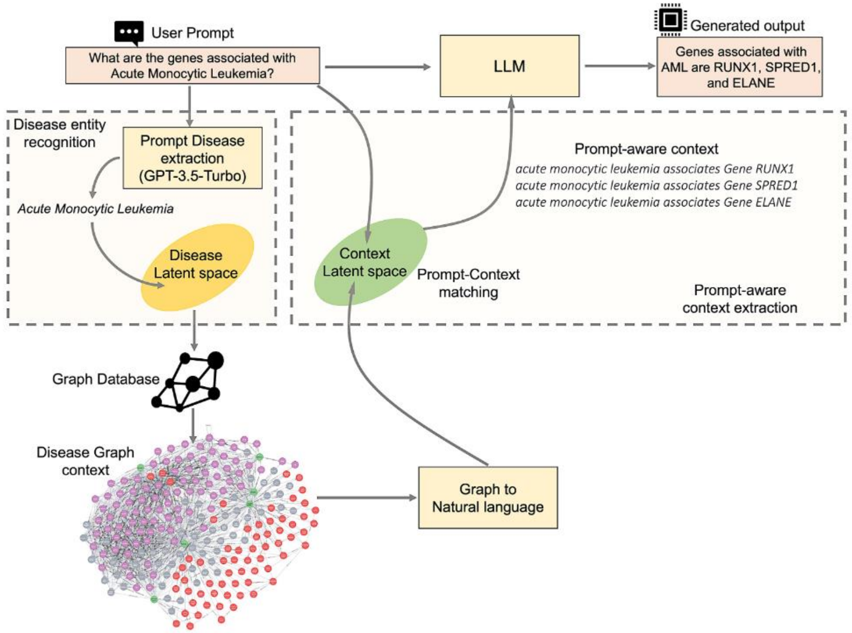




Tian, Shiyu, et al. "KG-Adapter: Enabling Knowledge Graph Integration in Large Language Models through Parameter-Efficient Fine-Tuning." *ACL Findings* (2024): 3813–3828.

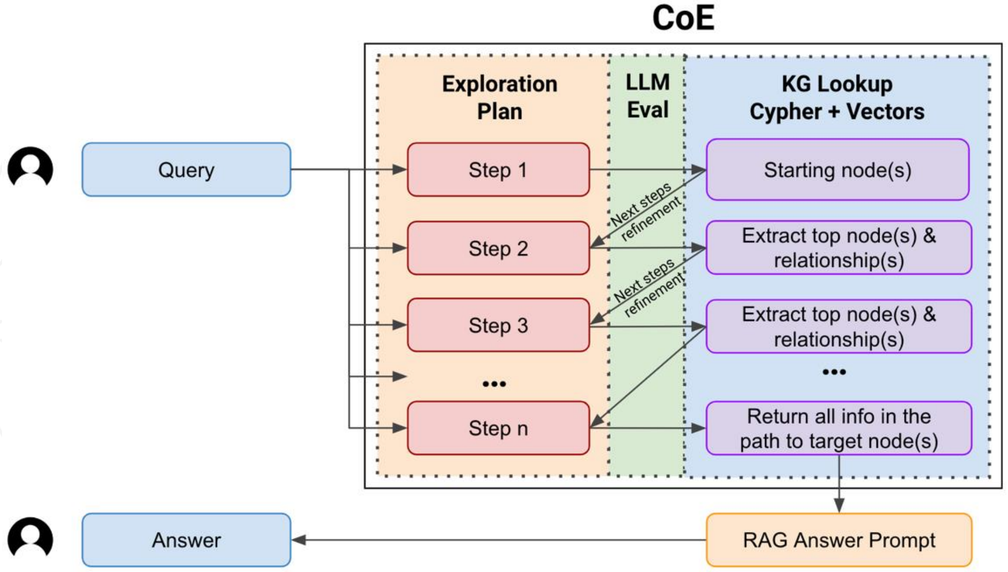
# Role Evolution of KGs in Synthesizing with LLMs

- KG based Retrieval Augmented Generation (KG-RAG)
  - KG-RAG for knowledge intensive tasks [Bioinformatics, 2024]
  - Chain of Explorations (CoE) for KG-RAG [arXiv2024]

Introduce CoT with LLM to select relevant nodes or relationships from KG




KG-RAG: Cypher-query based Retrieval

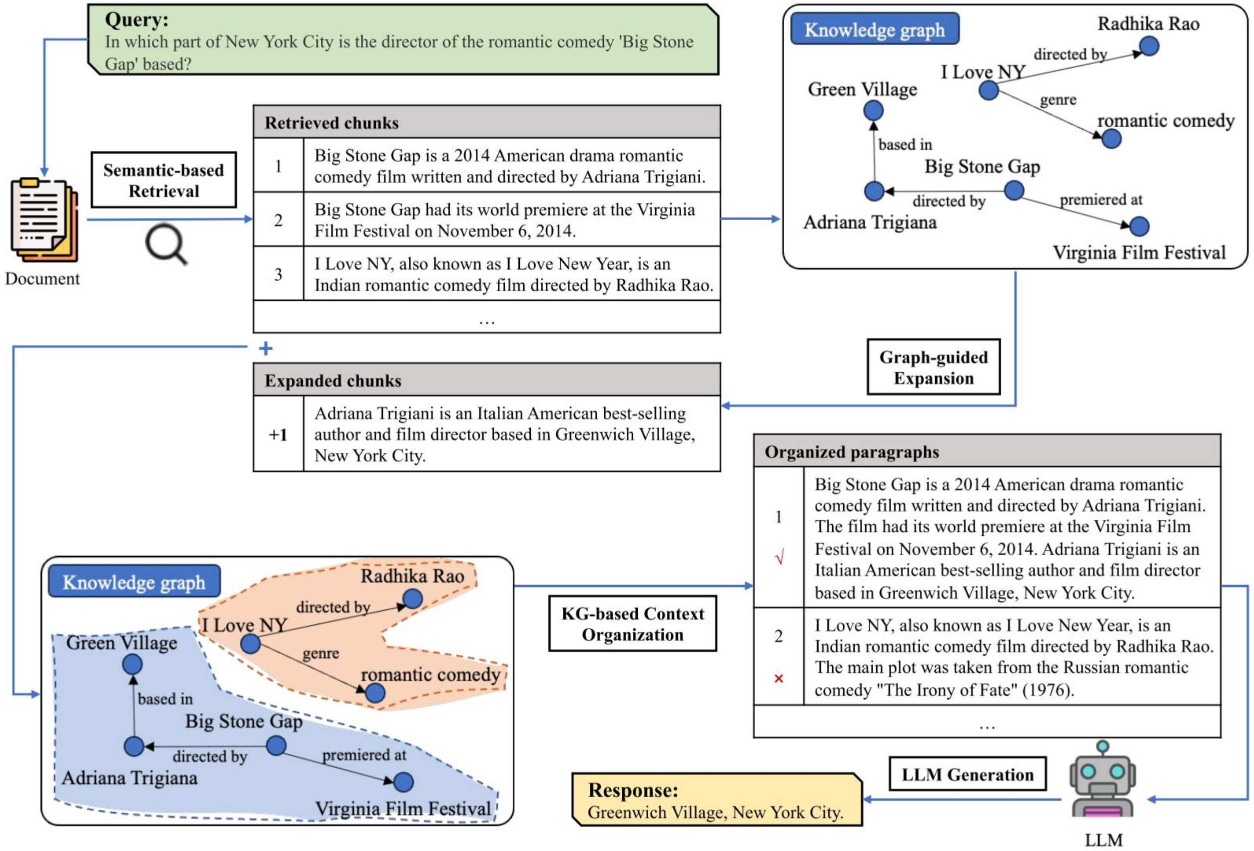



KGQA over a KG retrieval based on CoE


Soman, Karthik, et al. "Biomedical knowledge graph-optimized prompt generation for large language models." *Bioinformatics* 40.9 (2024): btae560.
Sanmartin, Diego. "KG-RAG: Bridging the gap between knowledge and creativity." *arXiv preprint arXiv:2405.12035* (2024).

# Role Evolution of KGs in Synthesizing with LLMs

- Retrieval Augmented Generation (RAG)
  - KG-guided RAG (KG²RAG) [arXiv 2025]




Text with available existing KG: establish linkage between text chunks and KG chunks

Text without KG: extract entities and relations from text chunks to form subgraph

$$\mathcal{S} = \{s(q,c) \,|\, c \in \mathcal{D}\},$$

a. Semantic-based chunks retrieval

$$\mathcal{G}_q^0 = \{(h,r,t,c) \,|\, c \in \mathcal{D}_q\} \subseteq \mathcal{G}.$$

b. Retrieve the relevant subgraph from KG

$$\mathcal{G}_q^m = \mathrm{traverse}(\mathcal{G}, \mathcal{G}_q^0, m),$$

c. Expand retrieved chunks with the m-hop BFS searched neighbor subgraphs on KG

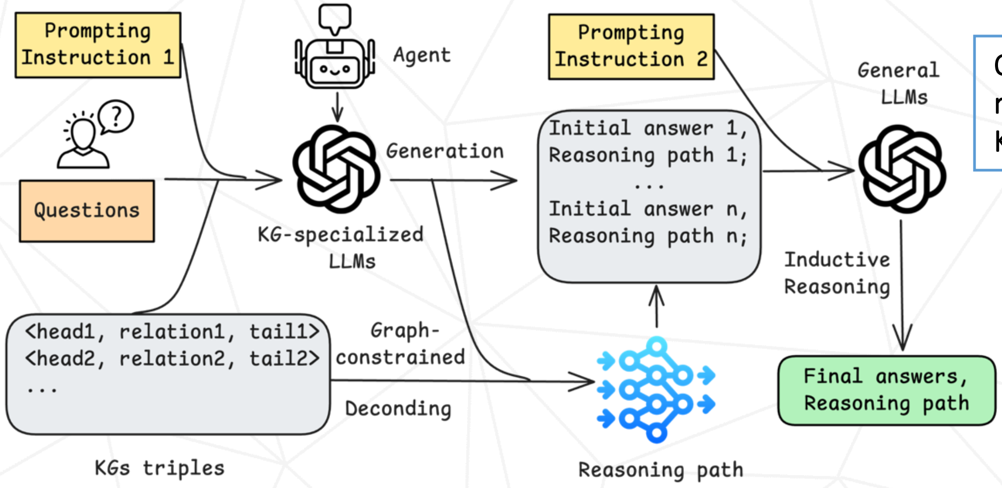
$$R(q, \mathcal{T}_i) = C(q, \mathrm{conc}(\mathcal{T}_i)),$$

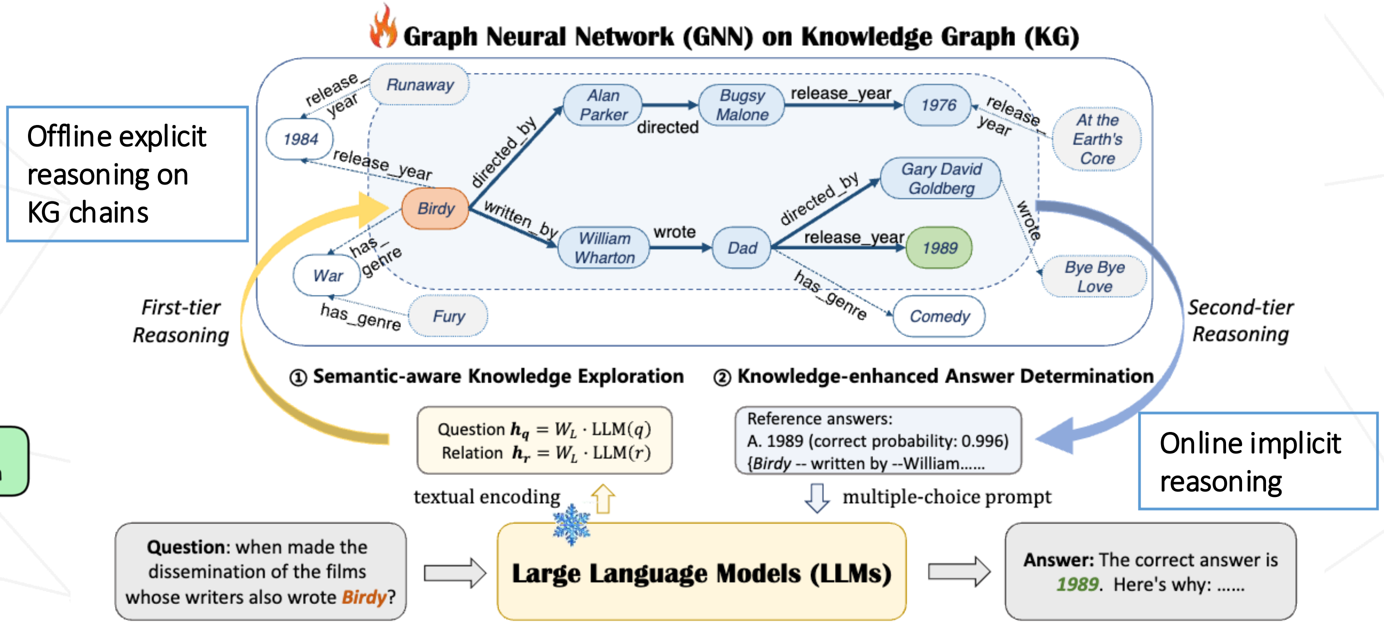
d. Rank the relevant expanded chunks and incorporates it with the retrieved chunks as context

Xiangrong, Zhu, et al. " Knowledge Graph-Guided Retrieval Augmented Generation." *arXiv preprint arXiv:2502.068641* (2025).

# Role Evolution of KGs in Synthesizing with LLMs

- KGs serves as guidelines to LLMs for Joint Reasoning
  - Offline Reasoning: KGs-based reasoning before LLMs reasoning
  - Online Reasoning: KGs-based reasoning directly involves in LLMs reasoning
  - Agent-based KG guidelines: Agent-based autonomous reasoning



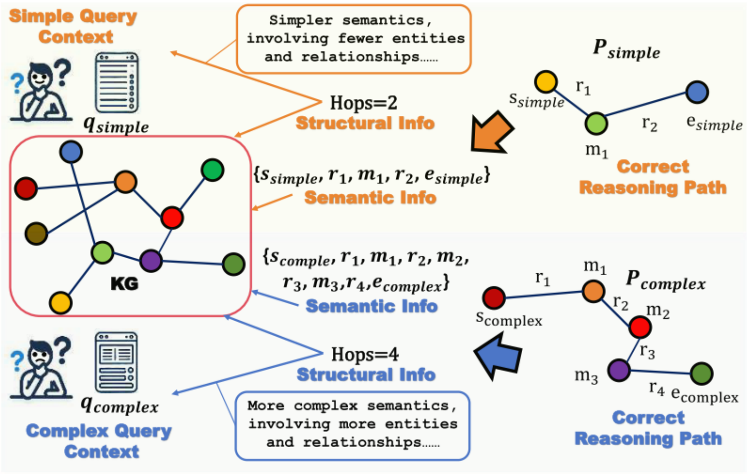



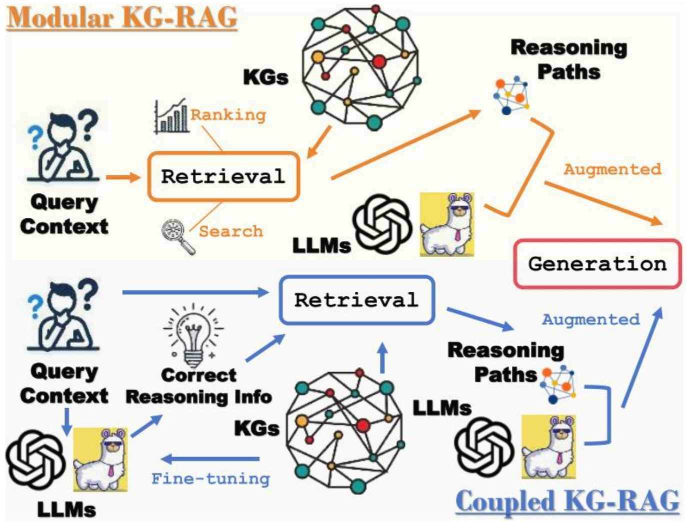

Liu, Guangyi, et al. Dual Reasoning: A GNN-LLM Collaborative Framework for Knowledge Graph Question Answering". CPAL, (2025).

# Role Evolution of KGs in Synthesizing with LLMs

- Offline Reasoning: [FRRAG] Modular KGRAG for KGQA [arXiv, 2025]
  - Estimate the hop of reasoning to classify the question
  - Employ BFS and ranking to retrieve subgraph for reasoning
  - Leverage LLM to extract reasoning paths
  - Generate the final answer based on retrieved reasoning paths

**Prompt:** You are an expert reasoner with a deep understanding of logical connections and relationships. Your task is to analyze the given reasoning paths and provide accurate reasoning path to the questions based on these paths. Based on the reasoning paths, please extract the correct reasoning path. If NO correct reasoning path, please just reply NO.
Reasoning Paths: {paths}
Question: {question}
Correct reasoning path:




**Prompt:** You are an expert reasoner with a deep understanding of logical connections and relationships. Your task is to analyze the given reasoning paths and provide clear and accurate answers to the questions based on these paths. Based on the reasoning paths, please answer the given question.
Reasoning Paths: {paths}
Question: {question}

Gao, Zengyi, et al. "FRAG: A Flexible Modular Framework for Retrieval-Augmented Generation based on Knowledge Graphs." *arXiv preprint arXiv:2501.09957* (2025).

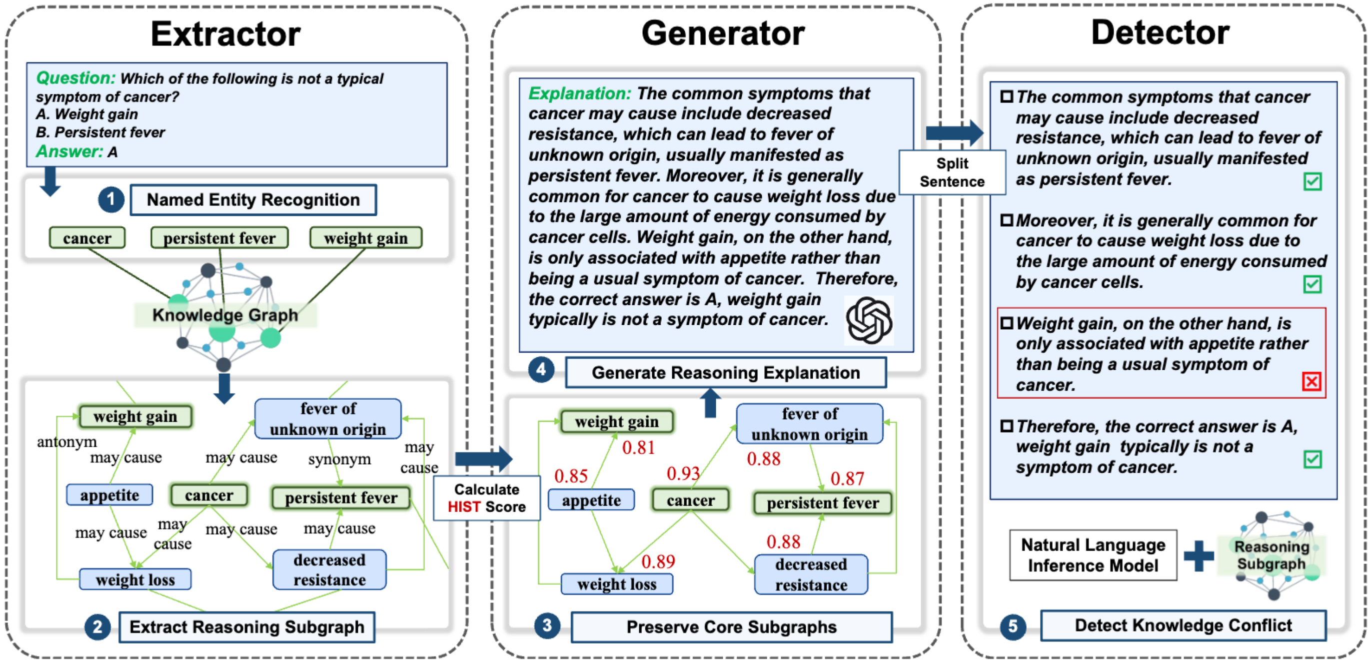
# Role Evolution of KGs in Synthesizing with LLMs

- Online Reasoning: [KG-SFT] KG augmented supervised fine-tuning LLM for KGQA [ICLR2025]
  - Search neighboring entities to obtain the reasoning subgraphs
  - Generate reasoning-based explanations via an external LLM
  - Detect knowledge conflict based on online reasoning (reasoning subgraph and natural language inference model)

(1) Perform NER on QA pairs for entity extraction

(2) Retrieve core subgraph from external KGs that is related to QA pairs vis HITS (Hyperlink-Induced Topic Search)

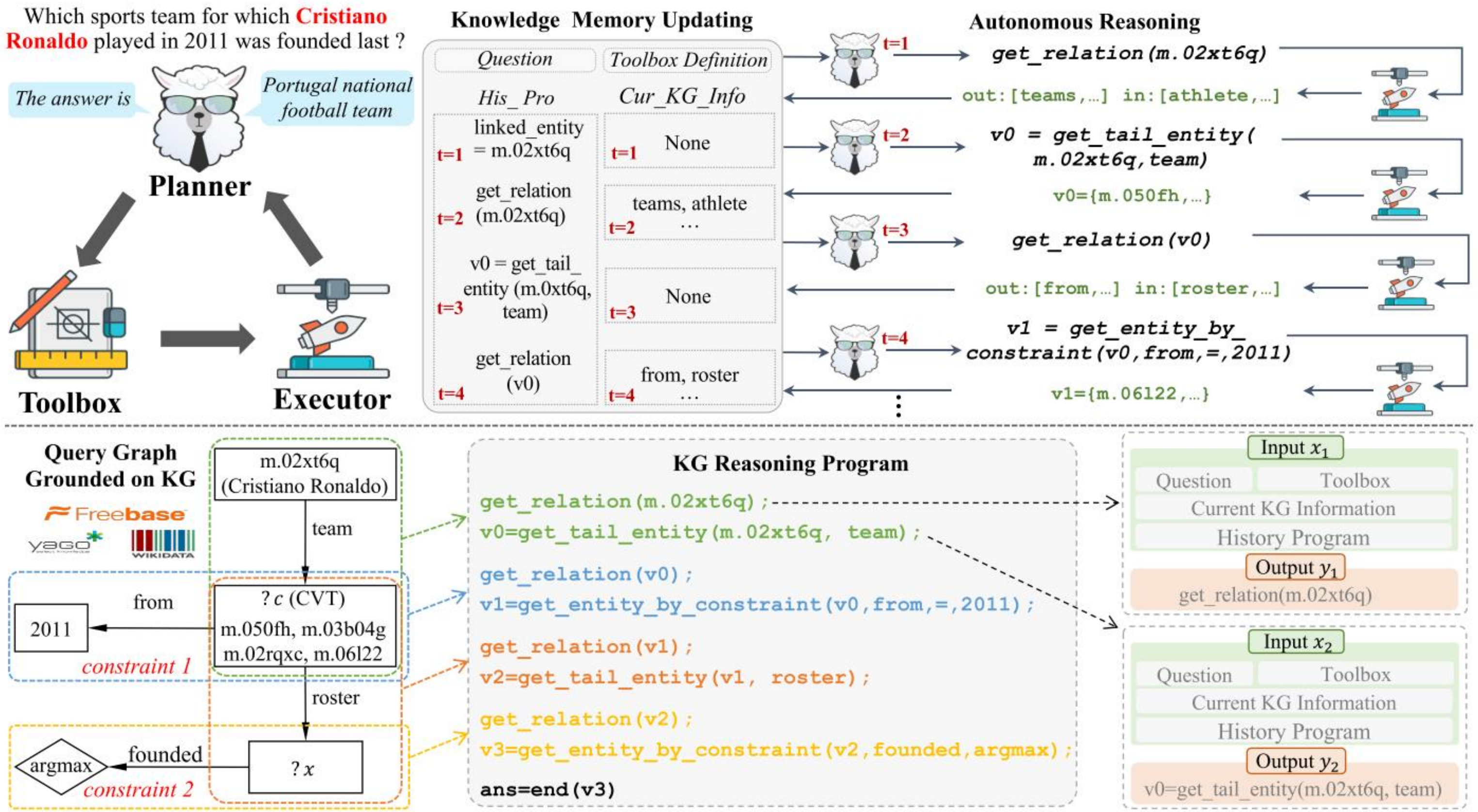
(3) Split the LLM generated reasoning explanations and fuse it with language inference model




Chen, Hanzhu, et al. "Knowledge Graph Finetuning Enhances Knowledge Manipulation in Large Language Models." *ICLR (2025)*.

# Role Evolution of KGs in Synthesizing with LLMs

- Agent based Reasoning: [KG-Agent] Agent-based autonomous reasoning for KGQA [arXiv2024]




Instruction-tuned LLM for planner

KG-based executor for knowledge updating

Agent-based autonomous reasoning over KG

Agent autonomously iterates the tool selection and knowledge updating for reasoning util reaching the answer entities

Example of instruction fine-tuning data synthesis and KG reasoning for the input-output pairs

Jiang, Jinhao, et al. "KG-Agent: An efficient autonomous agent framework for complex reasoning over knowledge graph." *arXiv preprint arXiv:2402.11163* (2024).

# Role Evolution of KGs in Synthesizing with LLMs

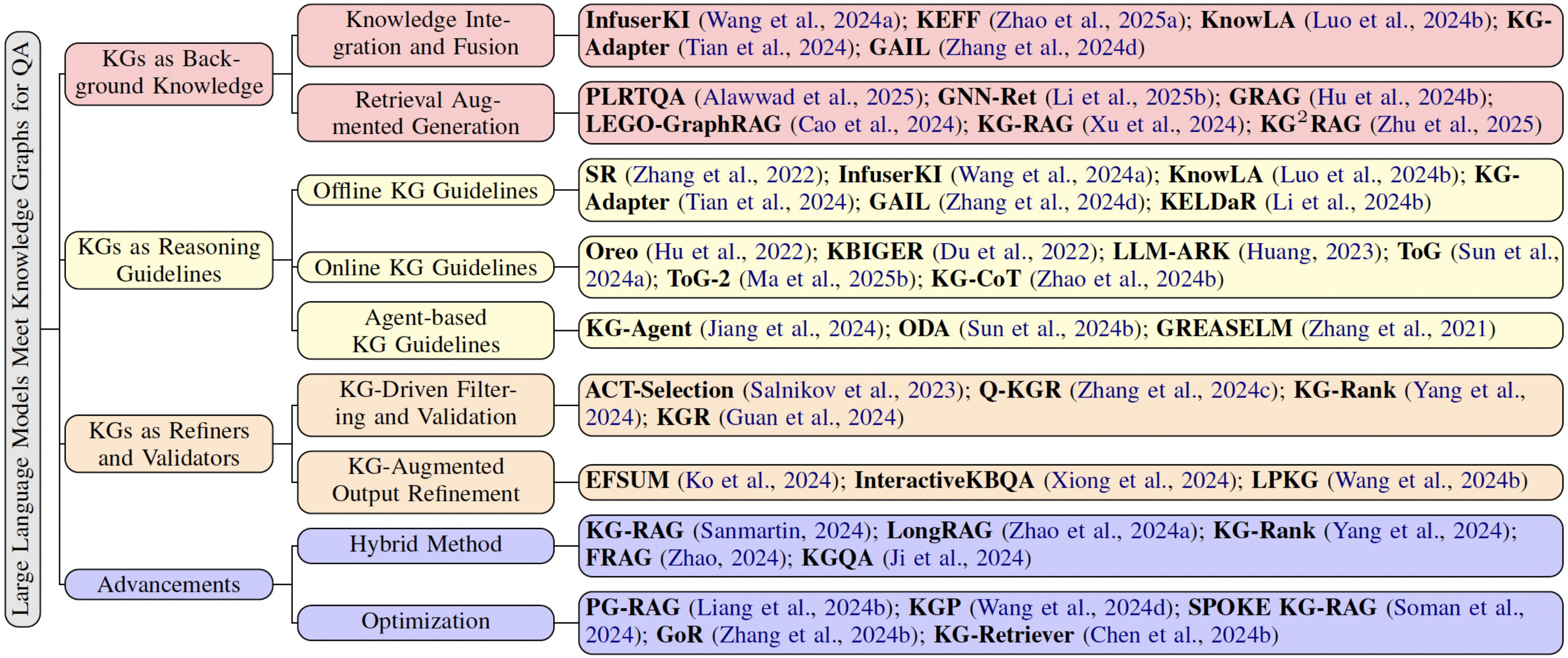
- Comparison of Approaches with Different Roles of KGs

| Approaches | Key Techniques | Strengths | Limitations | KG Requirements |
|---|---|---|---|---|
| KGs as background knowledge | Pre-training, fine-tuning, KG-based RAG | Hallucinations mitigation | Re-training is needed when updating KG | High domain coverage and up-to-date factual knowledge |
| KGs as reasoning guidelines | Rule-based reasoning, CoT-based path exploration | Explainable results with less hallucinations, Knowledge interaction | Costly rule mining and prompt token overhead, path explosion and reasoning latency | Rich logical and semantic Knowledge, Flexible interaction interface |
| Hybrid approaches | Agent-based knowledge reasoning, reinforcement learning | Iterative reasoning, knowledge updating | High reasoning complexity and computing cost | Dynamic knowledge adaptation |

Chuangtao, Ma, et al. " Large Language Models Meet Knowledge Graphs for Question Answering: Synthesis and Opportunities." *arXiv: 2505.20099* (2025).

# Role Evolution of KGs in Synthesizing with LLMs

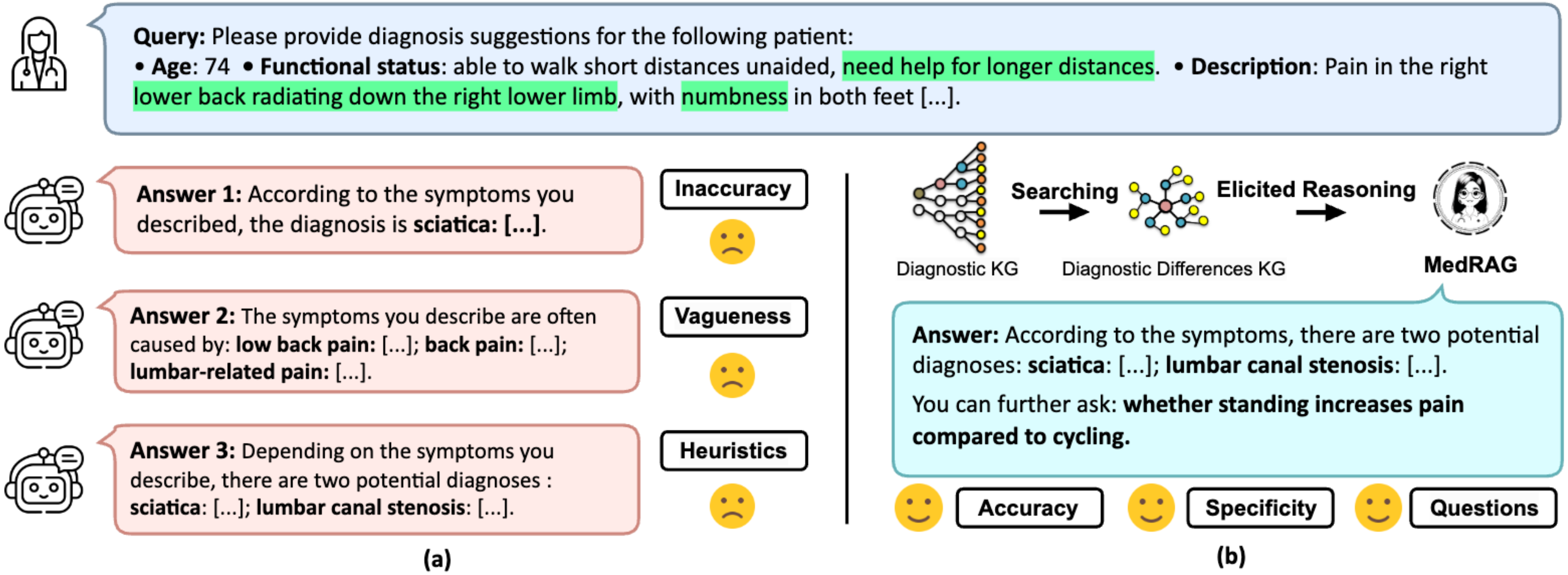
- **Our Recent Survey**
  - LLMs Meet Knowledge Graphs for Question Answering: Synthesis and Opportunities [arXiv2025] https://github.com/machuangtao/LLM-KG4QA




**Chuangtao, Ma**, et al. " Large Language Models Meet Knowledge Graphs for Question Answering: Synthesis and Opportunities." *arXiv: 2505.20099* (2025).

# LLMs + KGs Applications

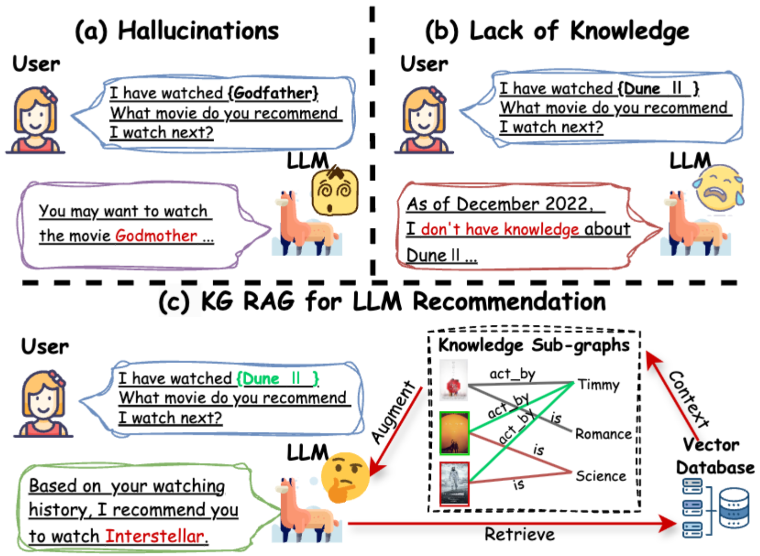
- Medical QA
  - MedRAG: KG-augmented LLMs for healthcare diseases diagnostic [WWW2025]
    - Leverage multi-level matching and upward traversal techniques to search the diagnostic differences KG.
    - Generate the final answers and follow-up questions for precise diagnosis based on the KG-elicited reasoning.

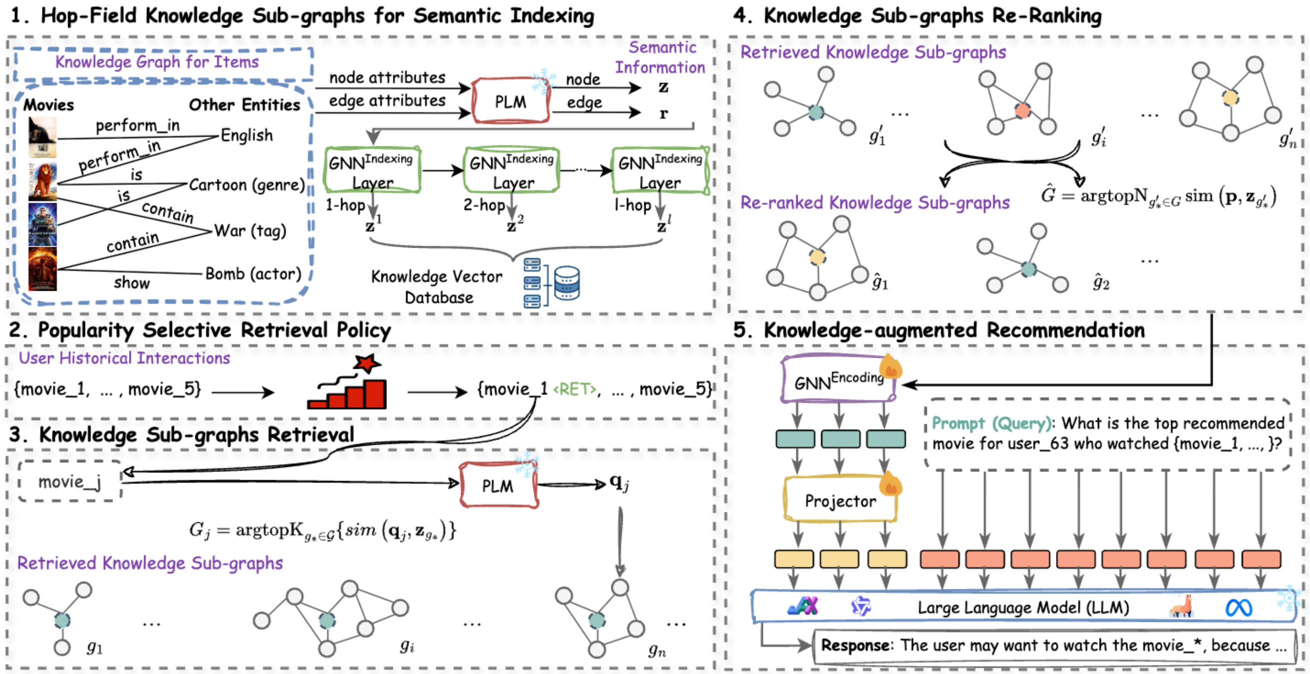



Zhao, Xuejiao, et al. (2025). MedRAG: Enhancing Retrieval-augmented Generation with Knowledge Graph-Elicited Reasoning for Healthcare Copilot. *WWW2025.*

# LLMs + KGs Applications

- Recommendation
  - K-RAGRec: KG RAG for LLM-based Recommendation [ACL2025]
    - Introduce a popularity selective and similarity-based ranking to retrieve the relevant subgraphs from item's KGs.
    - Leverage GNN and projector to align the retrieved sub-graphs into semantic space of LLM for recommendation.



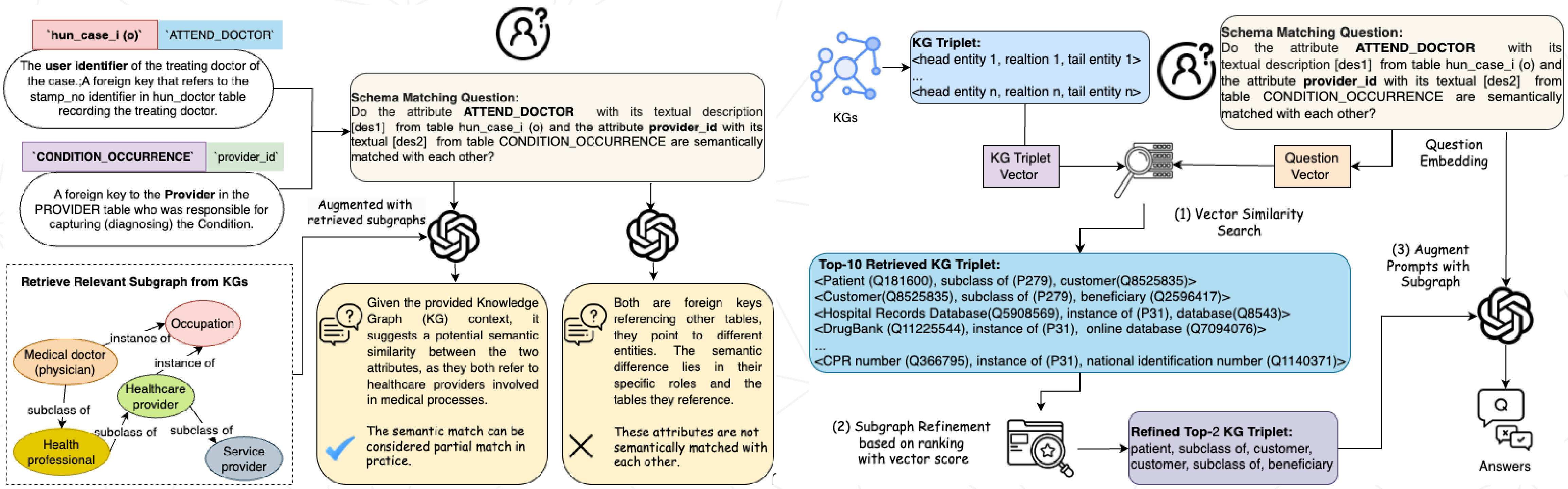




Wang, Shijie, et al. (2025). Knowledge Graph Retrieval-Augmented Generation for LLM-based Recommendation. ACL*2025.*

# LLMs + KGs Applications

- Schema Matching and Data Integration
  - KGRAG4SM: KG based RAG for Schema Matching [arXiv 2025]

    - A hybrid retrievals with vector-based, graph traversal-based, and ranking-based graph refinement to retrieve relevant subgraphs from external large KGs to augment LLMs for schema matching.




**Chuangtao Ma**, Sriom Chakrabarti, **Arijit Khan**, Bálint Molnár. Knowledge Graph-based Retrieval-Augmented Generation for Schema Matching. *arXiv:* 2501.08686 (2025).

# Conclusion

- KG-RAG
  - *Vector-based graph retrieval:* vector-based search is computing-consuming task for large KGs.
  - *Query-based graph retrieval*: Text2GQL is a challengeable task as the schema is agnostic for LLMs.
- KG-guided Reasoning
  - *Faithful of reasoning :* generate the reasoning paths from KGs is primary based on LLMs while the faithful of generated reasoning paths from KGs need to be addressed.
  - *Complex reasoning*: reasoning over large-scale KGs is a time-consuming and computing-consuming task, efficient graph reasoning techniques need to be further investigated.
- LM and KG Alignment
  - *Effective knowledge fusion:* integrating LLMs with KGs with the prompt-based fusion is not the optimal one as the topological information of KGs will be lost during the conversion.
  - *Dynamic Knowledge Integration*: keep the up-to-date knowledge in KGs and support incremental knowledge updating of KGs.
  - *Knowledge conflicts mitigation:* the knowledge conflicts between the internal knowledge of LLMs and the retrieved external knowledge need to be mitigated.

# Thanks

https://machuangtao.github.io