

Knowledge Graph and Large Language Model Co-learning via Structure-oriented Retrieval Augmented Generation

Carl Yang*, Ran Xu, Linhao Luo, Shirui Pan

Abstract

Recent years have witnessed major technical breakthroughs in AI– facilitated by tremendous data and high-performance computers, large language models (LLMs) have brought disruptive progress to information technology from accessing data to performing analysis. While demonstrating unprecedented capabilities, LLMs have been found unreliable in tasks requiring factual knowledge and rigorous reasoning. Despite recent works discussing the hallucination problem of LLMs, systematic studies on empowering LLMs with the ability to plan, reason, and ground with explicit knowledge are still lacking. On the other hand, real-world data are enormous and complex, coming from different sources and bearing various modalities. Data professionals have spent tremendous efforts collecting and curating countless datasets with different schemas and standards. Transforming the separate datasets into unified knowledge graphs (KGs) can facilitate their integrative analysis and utilization, but these processes would often require strong domain expertise and significant human labor. In this paper, we discuss recent progress and promise in the co-learning of KGs and LLMs, through LLM-aided KG construction, KG-guided LLM enhancement, and knowledge-aware multi-agent federation, particularly emphasizing a structure-oriented retrieval augmented generation (SRAG) paradigm, towards fully utilizing the value of complex data, unleashing the power of generative models, and expediting next-generation trustworthy AI.

1 Introduction

Large language models (LLMs) have reshaped AI research and implementations, with unprecedented capabilities widely shown in various language-related tasks [1–7], bringing humans ever close to general AI [8–10]. Recent research on multi-agent systems has further magnified LLMs’ advantages of *broad knowledge*, *language comprehension*, and *generalizability* through conversational collaborations, showing strong promise for further human-model collaboration for critical applications [11–15]. Recent studies have also revealed the limitations of LLMs regarding their *lack of planning* [16–20], *fuzzy inference* [21–25], and *hallucination* [26–33]. Specifically, in many real-world application scenarios, the lack of accurate planning can be caused by the lack of access to high-quality domain knowledge, especially the rapidly evolving new knowledge; the fuzzy inference nature can lead to difficulties in conducting reliable comprehension and stable predictions for complex questions; and hallucination creating factual errors and misinformation can cause fatal and life-threatening problems in critical applications [34–39].

Knowledge graphs (KGs) have been widely studied across academia and industry, due to their advantages in *storing accurate knowledge*, *facilitating explicit inference*, and *allowing easy editing of the knowledge* [40–43]. However, the creation of KGs relies much on the *standardization of data*, which requires significant schematic designs and human efforts. In many application domains, researchers and professionals have spent decades collecting, processing, and curating various types of data towards the

*Corresponding author (j.carlyang@emory.edu); Department of Computer Science, Emory University, Atlanta, GA 30322, USA

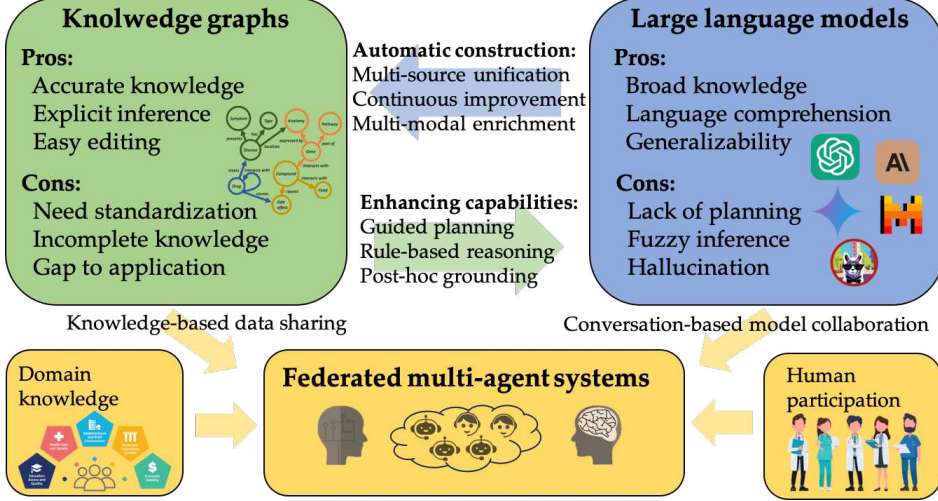


Figure 1: Overview of the proposed knowledge graph and large language model co-learning framework.

construction of KGs [44–53], which are widely used to support downstream application such as search and ranking [54–56], user modeling and recommendation [57–59], basic science research [60–65], healthcare [66–68] and education [69, 70]. Nevertheless, KGs still suffer from *incomplete knowledge coverage*, due to the stringent requirements on data standardization and limited data volumes compared to unstructured data in the wild, while specific algorithms are often needed to fill *their gaps to applications*. Moreover, real-world data are noisy and complex, where entities and relations come from various sources such as institutions using different data schemas and naming conventions [71–73], and the data can also include multiple modalities such as tables, texts, images, and time series [74–85]. While such multi-source and multi-modality data hold great promise in integrative and comprehensive analysis, unifying and extracting high-quality knowledge from them is non-trivial.

Recently, significant research attention has been drawn to the synergies between KGs and LLMs [86–92], due to their naturally complementary advantages (Figure 1). The construction and modeling of KGs have often relied on advances in natural language processing (NLP), with recent efforts intensively exploring language models towards the embedding [93–98], completion [99–111] and construction [112–140] of KGs. Studies in the recent years have also bloomed to explore the utilization of KGs for enhancing LLMs through providing new sources of knowledge during pre-training [141–179] or inference [180–207], and enabling knowledge-based interpretation and evaluation [21, 22, 208–217]. Finally, pioneering studies have also been conducted to explore LLM-based multi-agent systems, mostly through prompt-based role-plays to simulate human collaborations [218–226].

In this paper, we re-emphasize the promise of KG and LLM co-learning, especially through a structure-oriented retrieval augmented generation (SRAG) paradigm, where LLMs extract structured knowledge from unstructured data, which can be further retrieved to enhance the capabilities and reliabilities of LLMs during applications. Specifically, we give examples and discuss several natural and promising use cases where LLMs can be utilized to automate the construction of high-quality KGs. Furthermore, we summarize and highlight several limitations of current LLMs that can be potentially mitigated through the utilization of KGs. Finally, we envision a federated multi-agent system where models and data are disentangled while humans and knowledge are deeply engaged. Future directions are further discussed in the end.

2 LLM-aided KG Construction

In this section, we study and establish the advantages of utilizing LLMs for the construction of KGs, by demonstrating their effectiveness in improving the *accuracy*, *consistency*, *coverage*, and *freshness* of knowledge. Popular KGs such as Freebase [227], Yago [228] and Wikidata [229] contain hundreds of millions of real-world entities like people, places, and things, along with their multi-typed relations. However, since the KGs are collected and curated by different platforms and institutions, they do not use a unified coding system or thesaurus. The varying terminologies due to different conventions or abbreviations can lead to high degrees of duplication and inconsistency when multiple KGs are directly put together. Moreover, the sheer amounts of data in existing KGs are enormous, but the knowledge is still never comprehensive enough to serve various needs of real-world applications, especially those requiring rapidly updated knowledge. Recently, pioneering studies including ours have demonstrated strong promise of utilizing LLMs to automate the construction, integration, and enrichment of KGs [112–122]. In the following, we give several examples of promising attempts of these kinds and discuss more natural use cases of LLMs and multi-modal foundation models (MMFMs) toward constructing high-quality KGs as promising future directions.

2.1 Integrating existing KGs

KG integration, also known as knowledge fusion or knowledge alignment, represents a fundamental challenge in the broader landscape of knowledge engineering, which involves integrating multiple KGs that originate from varied sources and formats [230, 231]. While individual KGs often excel in specific domains or use cases, their true potential can be unlocked through effective integration, enabling more comprehensive and robust knowledge representation [47, 232]. As the number and diversity of KGs continue to grow, the need for effective integration methods becomes increasingly critical.

However, the integration of existing KGs faces several key challenges: (1) *semantic heterogeneity across sources*: Different KGs often use varying terminologies, definitions, and contextual frameworks to represent similar concepts [233]; (2) *varying granularity levels in knowledge representation*: KGs may differ in the detail and depth with which they describe entities and relationships, impacting the consistency and usability of integrated data. Although several neural approaches have been proposed for entity alignment on KGs [234–236], these methods generally depend heavily on labeled data for training. However, obtaining sufficient labeled data often involves substantial manual effort and can be rather costly.

LLMs have emerged as a promising solution to these challenges with unique advantages: First, their strong natural language understanding capabilities enable them to capture semantic relationships among concepts that may be missed by traditional string-matching or embedding-based approaches [109]. Second, LLMs can draw on their extensive knowledge acquired during pre-training to aid in disambiguating entities and mapping relationships across different KGs [237]. Third, LLMs possess robust few-shot learning abilities, making them particularly valuable for specialized domain applications where labeled data are limited [8, 238].

Recent work has demonstrated the effectiveness of LLM-based approaches in KG integration. For example, Lu et al. [239] developed HiPrompt (framework shown in Figure 2), which aligns entities between biomedical KGs and standardized hierarchical entity taxonomies. This task poses significant challenges due to the scarcity of available pairs and the inconsistent naming conventions between KGs and entity taxonomies. Their two-stage approach combines traditional information retrieval techniques (BM25) with LLM-based re-ranking using hierarchy-oriented prompts, achieving superior performance in few-shot biomedical knowledge-graph integration. Building on this direction, Xie et al. [240] developed PromptLink, a framework that leverages both domain-specific language models and GPT-4 for cross-

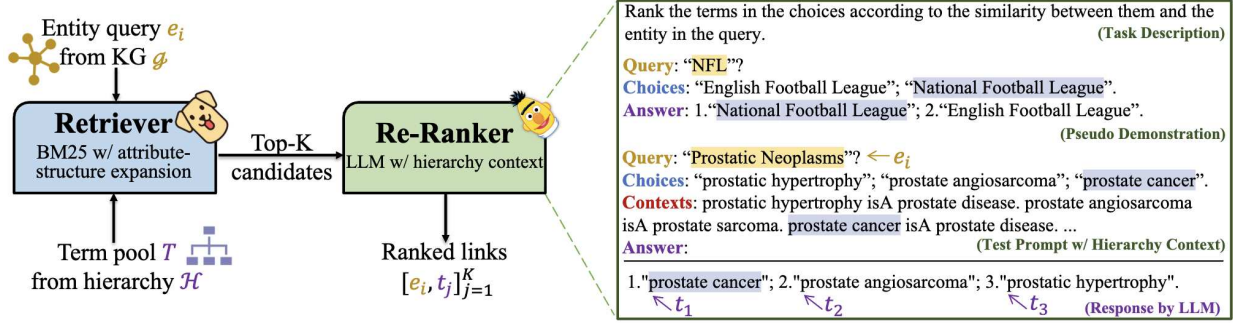


Figure 2: The overall framework of Hiprompt.

source biomedical concept linking. The framework employs a two-stage prompting mechanism by first eliciting the biomedical prior knowledge from the LLM for the concept linking task and then enforcing the LLM to reflect on its own predictions to further enhance their reliability. PromptLink’s success in zero-shot scenarios illustrates the potential of LLMs to generalize across diverse data sources without extensive training data.

In contrast to the two LLM approaches that primarily use LLMs for ranking candidate sets, AutoAlign [241] employs off-the-shelf LLMs to construct a predicate-proximity graph that captures relationships between entity types rather than individual entities. It then aligns the entity embeddings of two KGs into a common vector space by calculating similarity based on entity attributes.

These advances suggest several promising future directions for LLM-aided KG integration. For example, LLMs could potentially facilitate the continuous integration of new knowledge into existing KGs dynamically and automatically by identifying and resolving conflicts between new and existing information, while maintaining consistency across the integrated knowledge base. Such real-time data integration is especially beneficial for dynamic applications like live event monitoring and real-world decision-making. Furthermore, integrating KGs with LLMs remains challenging due to the risk of generating false or untrustworthy information [242]. To mitigate this issue, there is a growing need for improved human-in-the-loop systems. Specifically, enhanced interfaces [243, 244] can enable experts to more effectively verify and interpret LLM-generated recommendations, ensuring greater reliability and transparency.

2.2 Constructing and Completing KGs

KGs have high-standard requirements on the quality of knowledge, regarding accuracy, consistency, coverage and freshness. No matter constructed through manual curation, NLP tools, or their combinations, KGs can unavoidably include erroneous knowledge. Moreover, when multiple KGs are integrated, conflicting knowledge can emerge. Finally, new knowledge is constantly generated from new experiments and research, making existing knowledge inaccurate and incomplete. LLMs have emerged as a promising solution, leveraging the vast and adaptable knowledge acquired during pre-training to overcome these limitations [208, 245–247]. The key advantage of LLMs for KG construction and completion is their ability to generate novel, semantically coherent information with minimal reliance on additional labeled data.

Recent studies have highlighted the effectiveness of LLM-based approaches for KG construction and completion. Zhu et al. [112] utilize in-context learning capabilities of LLMs to complete tuples with missing entities or relations to generate new knowledge triplets for augmenting existing KGs. Wei et al. [248] and Wang et al. [249] tackle KG completion as a candidate identification and ranking task, proposing a “retrieve-rank” pipeline where LLMs are used to rerank top-retrieved entities, thus creating

additional knowledge triplets.

An alternative approach to prompting LLMs involves using code-based prompts, rather than natural language, to incorporate new entities into existing KGs [250, 251]. Code LLMs, extensively trained on structured data such as programming code, are inherently well-suited to the structured nature of KGs. Using a code-based interface enables more effective handling of graph-like structures, logical relationships, and precise reasoning [252–254]. Specifically, Bi et al. [250] first encoded the schema of KGs by modeling code definitions, to capture the structural information inherent in the data. They then employed chain-of-thought prompting to produce accurate knowledge triples. This methodology demonstrates improved performance over traditional natural language-based prompts. Similarly, Zeng et al. [251] proposed CodeTaxo, which represents entities within a base 'Entity' class, mirroring hierarchical relationships in programming constructs. This approach enables LLMs to efficiently create taxonomic structures by leveraging syntactic capabilities commonly used in code tasks, thus enhancing the organization and completeness of KGs.

The above methods primarily focus on prompting LLMs for KG construction and completion. While these methods show promise, they fall short in fully adapting LLMs to target tasks and can suffer from hallucination issues. To address these limitations, several studies aim to improve the quality of LLM-generated content for KG tasks. Zhang et al. [116] presented a three-phase framework for constructing KGs with LLMs to enhance contextual understanding and schema alignment. It starts with open information extraction to identify relation triplets from unlabeled textual corpora, followed by schema definition where LLMs generate contextually relevant descriptions for schema components. Finally, the extracted triplets are aligned with the schema. To further enhance the quality of the extracted triplets and minimize the risk of misinformation, a schema retriever is used to generate a list of candidate entities and relations to guide the triplet extraction steps. This approach works for both predefined schemas and situations where the schema needs to be inferred from the context. Additionally, various studies explored fine-tuning LLMs specifically for KG completion. Zhang et al. [255] proposed KOPA that first applies structural pre-training to create embeddings of entities and relations in KGs, and project embeddings into the textual space as virtual knowledge tokens. These tokens act as prefixes in LLM input prompts, enabling structure-aware reasoning that leverages both the generative power of LLMs and the retrieval of structured KG information to improve the accuracy and completeness of KG completion tasks. Jiang et al. [256] introduced KG-FIT for using open-world knowledge from LLMs to enhance KG embeddings. It initially constructs a semantically coherent, hierarchical structure of entity clusters, guided by LLM-powered entity representations. It then fine-tunes these embeddings by integrating the hierarchical structure with textual embeddings. This hybrid approach allows KG-FIT to capture both the semantic depth of LLMs and the structural information intrinsic to KGs, resulting in more comprehensive KG representations.

2.3 Enriching KGs with multi-modality data

Traditionally, specialized models and algorithms have been developed to process and analyze various modalities of data such as tables, texts, images, and time series. These methods can hardly perform integrative analysis across data modalities and generalize across different data platforms. Recently, LLM-based multi-modality foundation models (MMFMs) have shown strong promise in analyzing multi-modality data through the unified interface of languages [257–261]. Integrating multi-modal data into KGs creates a more comprehensive representation of entities and their relationships, enhancing performance in open-world applications like image classification and visual question answering [262]. Many studies focus on using MMFMs for specific tasks, such as entity extraction [263–265], relation extraction [266–268], or event extraction [269, 270]. However, these works often isolate extraction tasks without unifying entities and relations into a structured KG. A pioneering approach in this direction is

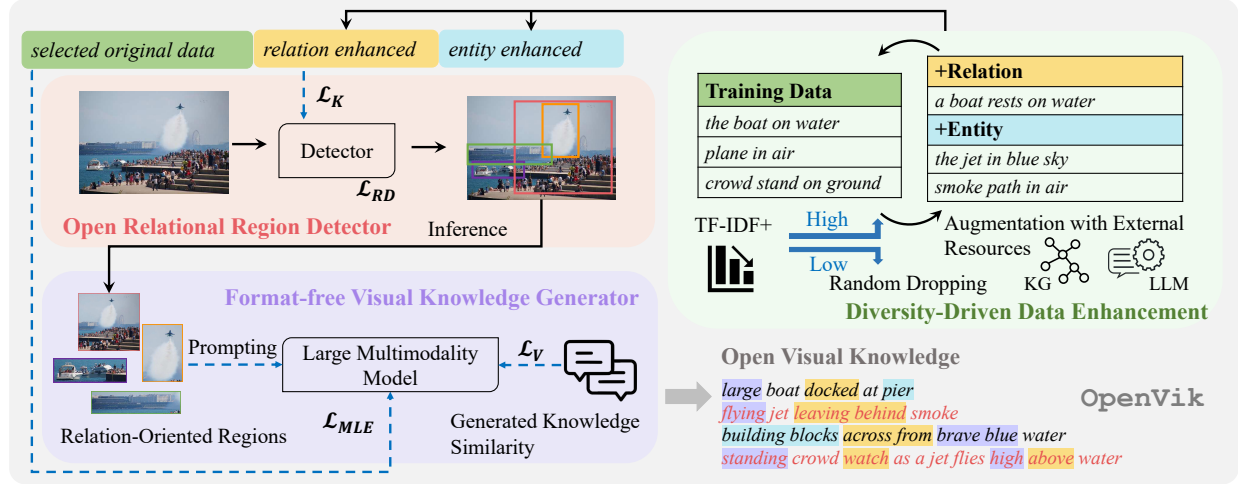


Figure 3: The overall framework of OpenVik consists of two main components: (1) an open relational region detector, highlighted in the orange and purple panels, which includes a region regression loss (\mathcal{L}_{RD}) and a regional description loss (\mathcal{L}_K), and (2) a format-free visual knowledge generator, incorporating knowledge generation loss (\mathcal{L}_{MLE}) and diversity regularization (\mathcal{L}_V). These modules work collaboratively to extract open visual knowledge, incorporating novel entities and diverse relations in a format-free manner.

OpenVik [271] (framework shown in Figure 3). It first trains an open relational region detector to locate image regions containing relational information. It then employs a visual knowledge generator to create format-free knowledge descriptions by prompting a large visual language model. Through the use of MMFM, OpenVik advances KG completion by integrating rich visual context, expanding knowledge coverage, and enhancing the accuracy of representation within the resulting KG. Application-wise, Yang et al. [272] proposed an automated approach to constructing product KGs from raw images in e-commerce. This method first employs vision-language models (VLMs) to extract detailed image information and then uses an LLM to reason and infer additional KG properties not visually present, hierarchically expanding, and linking nodes to develop comprehensive, scalable KGs without human input.

3 KG-guided LLM Enhancement

LLMs have shown impressive communication and question-answering capabilities, demonstrating strong promise in various applications [238, 261, 273–288]. However, to reliably model domain-specific data and generate factual and accurate answers, LLMs still face the challenges of lacking domain knowledge, fuzzy inferences, and hallucination [16–33]. Retrieval augmented generation (RAG) [289], which aims at retrieving query-relevant evidence and generating evidence-based answers, has strong promise in evidence-critical domains. However, effective and efficient RAG for complex queries is still challenging which requires LLMs to be able to (1) generate logical plans for retrieving multiple pieces of relevant evidence from complex data, (2) conduct valid reasoning and inference to compose the pieces of evidence towards generating coherent answers, and (3) reliably guarantee the detection and removal of errors. In the following, we discuss how these challenges can be addressed with well-designed planning, reasoning, and reflection frameworks with the help of KGs.

3.1 Planning with domain knowledge

While LLMs excel in many NLP tasks [8, 290], they still face challenges in acquiring domain knowledge. To address this issue, many attempts seek assistance from KGs, which are often constructed to represent knowledge in specific domains, such as medicine [50], law [291], and finance [292]. The integration of KGs and LLMs has shown promising results in various applications, such as question answering [181], recommendation [186], and dialogue systems [293]. Despite the success, there are still challenges in effectively obtaining useful information from KGs and incorporating them into LLMs.

Existing methods typically depend on a retriever to obtain relevant triples. For example, Baek et al. [294] proposed a direct retrieval method to retrieve relevant triples from KGs. However, the retriever may not always retrieve the most relevant triples, leading to suboptimal performance. Additionally, KGs contain a wealth of domain-specific knowledge, making it challenging for LLMs with limited domain expertise to comprehend and utilize this information. To further unleash LLMs’ capabilities of leveraging domain knowledge, the *plan-and-solve* paradigm [295] has been proposed, in which LLMs are prompted to first generate a plan. Based on the plan, LLMs can retrieve the relevant domain knowledge and conduct reasoning to generate answers [296]. However, existing methods are incapable of handling the complex structured knowledge in KGs to enable effective planning and reasoning. To address this issue, we propose a *planning-retrieval-reasoning* framework named **RoG** that enables LLMs to plan and reason on KGs [182]. The overall framework is illustrated in Figure 4.

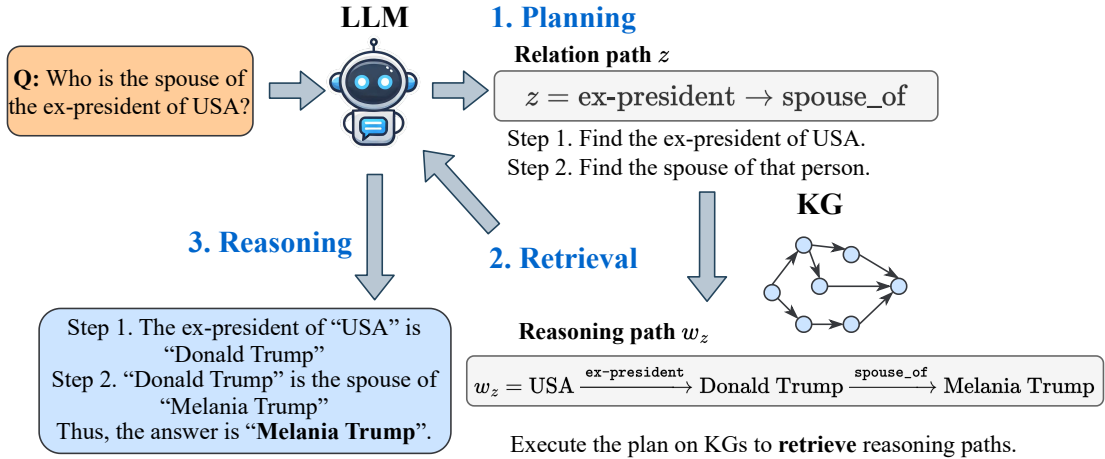


Figure 4: The overall framework of planning and reasoning on KGs (**RoG**).

RoG first generates several relation paths that are grounded by KGs as plans. Relation paths, which capture semantic relations between entities, have been utilized in many reasoning tasks on KGs [297, 298]. Based on relation paths, we can always retrieve the latest knowledge from KGs with a simple constrained breadth-first search. Therefore, relation paths can serve as faithful plans to guide the retrieval and reasoning on domain-specific KGs. Additionally, by treating relation paths as plans, we can make sure the plans are grounded by KGs, which enables LLMs to retrieve relevant knowledge and conduct faithful reasoning. To this end, we formulate our **RoG** as an optimization problem that aims to maximize the probability of reasoning the answer from a KG \mathcal{G} w.r.t the question q by generating relation paths z as the plan:

$$P_{\theta}(a|q, \mathcal{G}) = \sum_{z \in \mathcal{Z}} P_{\theta}(a|q, z, \mathcal{G}) P_{\theta}(z|q), \quad (1)$$

where θ denotes the parameters of LLMs and a denotes the final answer. To enable accurate planning

with domain knowledge, we design two instruction tuning tasks: 1) *planning optimization*, which distills the knowledge from KGs into LLMs to generate faithful relation paths as plans; 2) *retrieval-reasoning optimization*, which enables LLMs to reason based on the retrieved reasoning paths. The final objective function of **RoG** is the combination of the planning optimization and retrieval-reasoning optimization, which can be formulated as

$$\mathcal{L} = \underbrace{\log P_{\theta}(a|q, \mathcal{Z}_K^*, \mathcal{G})}_{\text{Retrieval-reasoning}} + \underbrace{\frac{1}{|\mathcal{Z}^*|} \sum_{z \in \mathcal{Z}^*} \log P_{\theta}(z|q)}_{\text{Planning}}, \quad (2)$$

where we use the shortest paths $\mathcal{Z}^* \subseteq \mathcal{Z}$ between q and a in KGs as supervision signals. We maximize the probability of LLMs generating faithful relation paths through distilling the knowledge from KGs. In this way, with the proposed **RoG**, LLMs can effectively retrieve domain knowledge from KGs with planning, which significantly enhances the reasoning capability of LLMs.

3.2 Reasoning with structured knowledge

KGs capture abundant factual knowledge in a structured format, which provides a faithful knowledge source for improving the reasoning abilities of LLMs [87]. Nevertheless, because of the unstructured nature of LLMs, directly applying them to reason on structured KGs is challenging. Early works focus on fine-tuning LLMs together with structured knowledge from KGs to enrich the knowledge of LLMs for better reasoning [152, 159]. For example, KEPLER [89] directly employs both KG embedding training objective and Masked token pre-training objective into a shared transformer-based encoder. Through fine-tuning, LLMs can better understand the structured knowledge in KGs for reasoning. However, the fine-tuning process is computationally expensive and incapable of efficiently adapting to the evolving real-world knowledge.

Recently, researchers have combined the strengths of retrieval-based methods with the prompting technique to enable LLMs to reason on KGs [181, 289]. CoK [25] and KD-CoT [299] retrieve facts from an external KG to guide the CoT performed by LLMs. To capture graph structure, GNN-RAG [148] adopts a lightweight graph neural network to effectively retrieve knowledge from KGs, which are formatted as a sentence path to elicit the reasoning process of LLMs. Mindmap [196] builds a prompt-based method that endows LLMs with the capability of comprehending KG and reasoning with it. Despite the success of these methods, they still face challenges in designing principled prompts to represent KGs and conduct reasoning. Moreover, LLMs still have limited capabilities in understanding the graph structure and reasoning with the text-based graph prompts [300].

Different from existing efforts that require a computationally expensive fine-tuning phase or design ad-hoc prompts for LLMs, we recently introduced a KG-constrained reasoning (**GCR**) paradigm [301]. **GCR** connects unstructured reasoning in LLMs with structured knowledge in KGs, seeking to achieve efficient and effective reasoning on structured knowledge. The overall framework is illustrated in Figure 5.

Graph-constrained reasoning, inspired by the concept that LLMs reason through decoding [302], incorporates the KG structure into the LLM decoding process. This enables LLMs to directly reason on graphs by generating reliable reasoning paths grounded in KGs that lead to correct answers. Specifically, given a question, we first adopt a retrieval module to find a relevant KG that is helpful for reasoning. Then, we convert the KG into a structured index, KG-Trie, to facilitate efficient reasoning on KG using LLMs. Trie is also known as the prefix tree [303] that compresses a set of strings, which can be used to restrict LLM output tokens to those starting with valid prefixes. KG-Trie encodes the reasoning paths in KGs as formatted strings to constrain the decoding process of LLMs. Then, we propose graph-constrained decoding that employs a lightweight KG-specialized LLM to generate multiple KG-grounded reasoning paths and answers. With the constraints from KG-Trie, we ensure faithful reasoning while leveraging the strong reasoning capabilities of LLMs to efficiently explore paths on

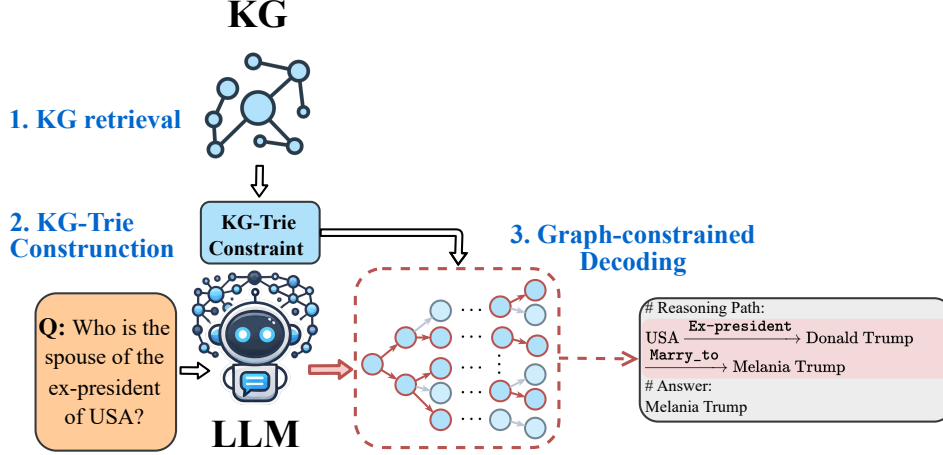


Figure 5: The overall framework of KG-constrained reasoning (GCR).

KGs in constant time. In this way, **GCR** bridges the gap between structured knowledge in KGs and unstructured reasoning in LLMs, allowing for efficient reasoning on KGs via LLM decoding.

3.3 Reflecting with atomic knowledge

LLMs have shown impressive capabilities in encapsulating massive knowledge and conducting reasoning. However, they still face challenges in generating factually correct and faithful responses, especially in the presence of hallucinations [304]. KGs store atomic knowledge in a structured format, which can be used to verify the correctness of generated responses and detect hallucinations [305]. To incorporate the factual knowledge from KGs into LLM hallucination detection, Guan et al. [200] proposed a retrieval-based method called KG-based retrofitting (KGR). KGR retrieves relevant facts from KGs during the LLM reasoning process, which are used to mitigate factual hallucination by retrofitting the initial responses. KGR enables an autonomous knowledge verifying and refining procedure with the factual knowledge retrieved from KGs, which significantly improves the reliability of LLMs.

The hallucination of LLMs is usually attributed to the lack of factual knowledge of LLMs. To systematically evaluate the factual knowledge inside LLMs, as shown in Figure 6a, we propose a novel framework to automatically assess the factual knowledge in LLMs by using KGs [214]. Unlike conventional methods that rely on human-annotated question-answering datasets, we systematically generate valid and diverse questions from KGs with different difficulties while also ensuring knowledge coverage. Specifically, we retire the atomic knowledge from KGs as sets of triples. Then, we utilize different question generation methods, e.g., template-based and LLM-based methods, to convert the triples into question-answer pairs. The generated pairs are used to evaluate the factual knowledge of LLMs by comparing the generated answers with the ground-truth answers. The evaluation results can be used to reflect the factual knowledge of LLMs. In this way, we can systematically evaluate the factual knowledge of LLMs and provide insights into the hallucination behavior of LLMs, which can be used to improve the reliability of LLMs in various applications.

Apart from the factual knowledge, the structure of KGs can be also utilized to justify the reasoning process of LLMs. Minh-Vuong et al [306] designed a framework that delves deeper into the CoT reasoning capabilities of LLMs in multi-hop question answering by utilizing KGs, as shown in Figure 6b. The framework contains two evaluation modules: discriminative evaluation and generative evaluation. The discriminative evaluation aims to analyze whether the LLMs possess enough knowledge to conduct

faithful reasoning. It feeds both valid and invalid reasoning paths retrieved from KGs into LLMs and asks them to predict the validity of these paths. The generative evaluation, on the other hand, aims to evaluate the faithfulness of the reasoning process of LLMs by grounding it on KGs. Given a reasoning process generated by LLMs, the generative evaluation module retrieves the facts from KGs, which are compared with the ground-truth reasoning paths. The evaluation results can be used to reflect the reasoning capabilities of LLMs and provide insights into the faithfulness of LLM reasoning. Based on the findings, although LLMs have shown impressive reasoning capabilities, they still face challenges in conducting faithful reasoning, especially in multi-hop question answering.

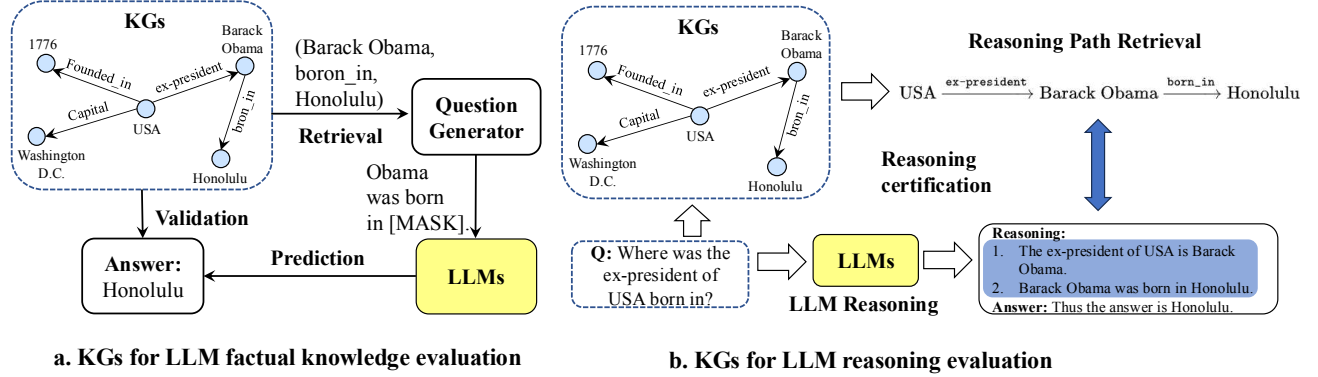


Figure 6: The illustration of LLM reflection with KGs. (a) The evaluation of the factual knowledge inside LLMs. (b) The evaluation of the reasoning process of LLMs with KGs.

4 Knowledge-aware Multi-agent Federation

While KGs and LLMs can mutually enhance each other, data are properties and many real-world datasets are privately collected and owned by different institutions, which cannot be simply put together to train more powerful models. Moreover, many real-world applications require domain-specific knowledge that may not have been captured by general-purpose KGs and LLMs yet, and such knowledge can also be private properties. Finally, while the development of KGs and LLMs is highly automated and data-driven, the values and needs of different human stakeholders may not have been properly reflected in the data and models. Federated learning (FL) provides a robust and principled framework for privacy-protected multi-site collaboration, but proper implementation of FL in the new era of generative AI remains unclear; the further incorporation of domain knowledge and human participation is also highly under-explored. In the following, we will envision an innovative Federated Multi-Agent System (FedMAS) for multi-site privacy-protected, knowledge-infused, and human-engaged KG-LLM co-learning scenarios.

Nowadays, while common practices in AI applications still largely resort to in-house development of models based on public and local data, the successes of generative AI, where complex models are trained with large-scale data, have demonstrated a strong need to collaboratively utilize local data towards obtaining powerful models that can generalize across institutions, finding and utilizing deep data patterns underlying common and rare use cases. Towards protecting local data privacy during collaborative model training, FL provides a promising solution [307, 307, 308, 308–310]. However, existing FL frameworks, by merely preventing the direct sharing of training data, are not effective in the scenarios of KG-LLM co-learning, because (1) as the construction of comprehensive KGs necessitates the

incorporation of knowledge discovered from local data, private information may get reversely inferred from the collaboratively constructed KG; (2) as powerful generative models like LLMs can easily memorize training data, collaboratively trained LLMs may expose private information facing deliberately composed jailbreaking prompts [311–313].

In our pioneering studies on FL for graphs [314–321], we developed several novel algorithms for different graph separation scenarios. In FedDEP [317], we developed a prototype-based embedding sharing algorithm with local graph differential privacy (DP) guarantees, and demonstrated its utility in FL for global graph embedding models across private local subgraphs; in FedR [319], we showed that sharing relation embeddings across local KGs can help FL for global KG embeddings with less privacy leakage. These studies have laid the foundations for our envisioned framework here, which will build on the private embedding sharing algorithms to construct *multi-view KGs* that can facilitate multi-site knowledge sharing with minimum risks of exposing local private data.

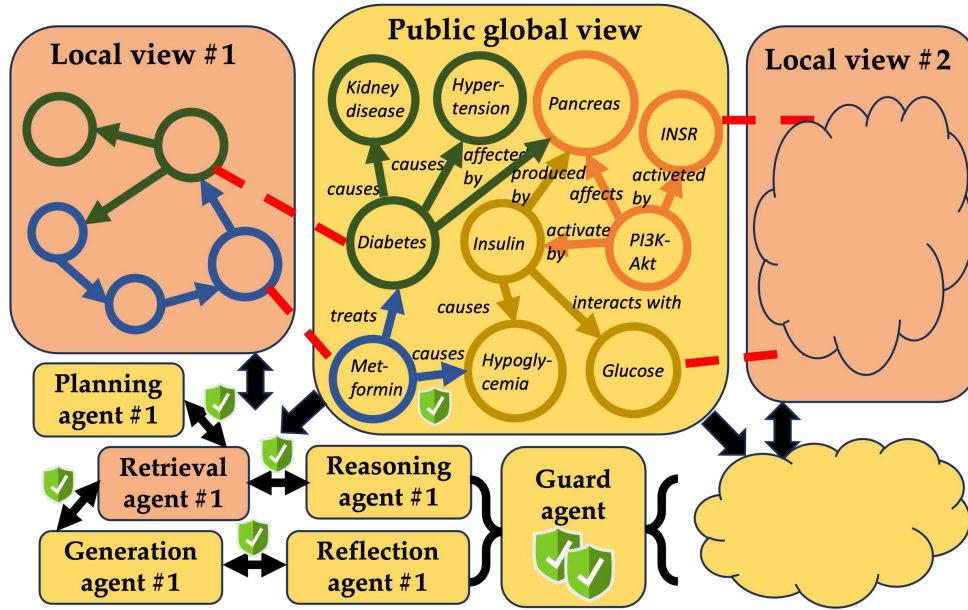


Figure 7: The envisioned framework of a federated multi-agent system (FedMAS).

As illustrated in Figure 7, our envisioned FedMAS will include a multi-view KG and various LLM agents. The federation on KG will be implemented by collectively constructing a multi-view KG by all participating sites, where knowledge from public resources is integrated into the global view, and knowledge from private resources is kept in each site’s local view only visible to itself. For each site, entities in its local view are linked with the global view, and only the embeddings related to these linked entities are shared with the server and other sites, via privacy-guaranteed embedding sharing algorithms such as those developed in FedDEP [317] and FedR [319]. In this way, each local site can compute embeddings on its local view and the public view as if they can see all other sites’ local views, allowing them to effectively adjust their local knowledge and further enhance their local LLMs, all without actually seeing the other sites’ local views (knowledge). The server will periodically adjust knowledge in the global view also based on the shared privacy-guaranteed embeddings, and apply additional privacy checks to make sure no sensitive knowledge gets propagated into the global view.

To rigorously protect the more sensitive patient data during the federation on LLM, it is possible to train multiple LLM agents in each site with different functions and let them collaborate through conversations instead of traditional model sharing, so the system can strictly control the level of private

data each agent can access and monitor or moderate their collaborations. Specifically, since the retrieval agents need to access all local patient data, one plan is to implement programmable guardrails [322] on its conversations with all other agents within the same site and forbid them from directly communicating with agents from other sites. Since all other agents can only access patient data indirectly through the retrieval agents, it is possible to adapt techniques such as our recently developed GuardAgent [323] based on the clear goals and typical outputs of each agent to monitor all cross-site conversations, detecting and removing any suspicious private information.

When applying FedMAS to specific application domains such as finance, law, education, and healthcare, it is promising to leverage the knowledge-based data sharing mechanism to incorporate existing domain knowledge towards further alleviating knowledge gaps and mitigating potential biases. Built on recent promising results from LLM-based data annotations as discussed in Section 2, FedMAS can utilize specialized LLM agents to perform comprehensive extraction of structured knowledge from existing guidelines and tutorials and automatically integrate them with existing general and domain-specific KGs. For example, in healthcare, one type of important domain-specific entity is social determinants of health (SDOH) [324–326]. The system can start with a set of known SDOH such as defined by WHO [327, 328], and further extend the set and discover their impacts and relationships with various risk factors by investigating relevant healthcare literature. The KGs enhanced through these steps are supposed to facilitate the alleviation of various health disparities when utilized by subsequent LLM agents in the FedMAS.

While FedMAS utilizes AI advances to automate multi-site data integration and modeling, comprehensive and trustworthy AI systems need to also incorporate the values and needs of various stakeholders, who can have different and even contradictory perspectives. LLMs, especially in our multi-agent conversational environment, provide unique convenience for effective and efficient human participation, where different stakeholders can verify, influence, and complement the decision processes and outputs of different LLM agents, all based on natural languages as the interface. Specifically, we envision a novel multi-stage intervention mechanism to efficiently enable the participation of different stakeholders in the LLM-based multi-agent conversational environment. The potential stages could include (1) LLM uncertainty quantification, where LLMs highlight their own uncertain outputs; (2) Rubrics-based rating, where humans create rubrics to automatically rate the LLM outputs; (3) Focused human interactions, where humans directly interact with LLMs, focusing on the problematic scenarios identified in the previous stages. The overall multi-stage mechanism is supposed to allow FedMAS to adapt to human values through iteratively integrating the language-based feedback via interactions with various stakeholders.

5 Conclusions and Future Directions

In this paper, we discuss the trending efforts of co-learning KGs and LLMs. Through the lens of SRAG, we showcase promising attempts to utilize LLMs to automate the construction, integration, and enrichment of KGs, and discuss how KGs can help with planning paths, guide reasoning with structure, and ground knowledge with reflection, enhancing the reliability of LLMs for downstream tasks. We also envision a novel system of multiple agents collaborating in a conversational federated learning environment based on the knowledge-infused, human-engaged LLMs. While the co-learning of KGs and LLMs holds great potential, we envision several promising directions especially from the SRAG perspective.

Effective evaluation of LLM-generated knowledge. To achieve effective knowledge enrichment for KGs with LLMs, it is critical to evaluate and guarantee the quality of added and/or modified knowledge. However, new knowledge is hard to evaluate in nature due to the lack of ground truth. Exhaustive

human evaluation is costly, but LLMs can be utilized to lubricate the collaboration between humans and machines toward efficient new knowledge evaluation. For example, humans can create guidelines and rubrics for LLMs to screen and rate the new knowledge from different perspectives. LLMs can also evaluate the quality of knowledge with confidence or uncertainty quantifications. Humans can then focus on the LLM-flagged suspicious or uncertain new knowledge to conduct close manual evaluation.

Unified versus specialized KGs. Due to the diversity and breadth of knowledge, it might be difficult to integrate all knowledge into a single unified KG, which might potentially harm the knowledge integrity. As a potential alternative, it may become practical to construct specialized KGs depending on the knowledge needs of different applications. It then remains an open problem regarding how to measure the relevance of knowledge with respect to specific applications and decide what to include/exclude from the specialized KGs.

More powerful KGs. Current KGs mostly include general, binary, and pair-wise relations. However, when KGs are used in certain applications, the knowledge may not equally hold for every context. For example, one drug may treat a disease for only certain groups of patients. In such scenarios, specific mechanisms are needed to model the various contexts for knowledge. Moreover, relations are not always binary and pair-wise (between pairs of entities). They can be true with a probability and involve more than two entities. Such scenarios are ubiquitous in reality, so probabilistic KGs and n-ary KGs should receive wider adoption and study.

Trade-off between effectiveness and efficiency of retrieval. Most existing retrieval-based methods focus on developing an effective retrieval mechanism to accurately retrieve relevant knowledge from KGs [180, 329]. However, they often overlook the efficiency of the retrieval process. In practice, the retrieval process can be computationally expensive, especially when the KG is large. Meanwhile, real-world application often requires prompt responses, which further exacerbates the efficiency issue. Therefore, it is essential to strike a balance between the effectiveness and efficiency of the retrieval process [330].

Resolving knowledge conflicts (internal LLM knowledge versus external knowledge). LLMs contain a vast amount of knowledge obtained via pre-training. However, the knowledge might be inaccurate or outdated, which could conflict with the knowledge retrieved from KGs [331]. To resolve the conflict, SPARE [332] utilizes the internal activations of LLMs to identify the conflict. AstuteRAG [333] uses a novel RAG approach to adaptively elicit LLM internal knowledge and iteratively consolidate internal and external knowledge. Despite the attempts, how to effectively identify and resolve the conflict between the internal knowledge of LLMs and the external knowledge retrieved from KGs remains an open problem.

Retrieval from multi-modal data. KGs store knowledge in diverse modalities such as text, image, and video [334]. Existing KG retrieval methods mainly focus on retrieving textual knowledge. However, the retrieval from multi-modal data is still under-explored. Knowledge from different modalities can complement each other, which could potentially enhance the retrieval performance. Therefore, it is essential to develop retrieval methods that can effectively retrieve knowledge from multi-modal data [335].

Robustness/safety of SRAG for LLMs. The safety and robustness of LLMs are receiving increasing attention due to their critical role in developing trustworthy AI systems. Previous research has primarily focused on attacking the LLMs themselves [336]. However, integrating LLMs with KG retrieval systems

expands the attack surface. Attackers could manipulate KGs and the retrieval systems to mislead LLMs, potentially leading to severe consequences [337]. Therefore, enhancing the robustness and safety of the combined KG and LLM systems is an important research direction.

Acknowledgements

This research was partially supported by the National Science Foundation under Award Number 2319449 and Award Number 2312502, as well as the National Institute Of Diabetes And Digestive And Kidney Diseases of the National Institutes of Health under Award Number K25DK135913. Any opinions, findings, and conclusions or recommendations expressed herein are those of the authors and do not necessarily represent the views, either expressed or implied, of the National Science Foundation, National Institutes of Health, or the U.S. government. The authors wish to thank the editors and reviewers for their valuable efforts and suggestions.

References

- [1] OpenAI, “Gpt-4 technical report,” arXiv preprint arXiv:2303.08774, 2023.
- [2] C. Leiter, R. Zhang, Y. Chen, J. Belouadi, D. Larionov, V. Fresen, and S. Eger, “Chatgpt: A meta-analysis after 2.5 months,” Machine Learning with Applications, vol. 16, p. 100541, 2024.
- [3] G. T. Google, “Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context,” arXiv preprint arXiv:2403.05530, 2024.
- [4] NVIDIA, “Nemotron-4 340b technical report,” arXiv preprint arXiv:2406.11704, 2024.
- [5] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. Le Scao, T. Lavril, T. Wang, T. Lacroix, and W. El Sayed, “Mistral 7b,” arXiv preprint arXiv:2310.06825, 2023.
- [6] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, and G. Lample, “Llama: Open and efficient foundation language models,” arXiv preprint arXiv:2302.13971, 2023.
- [7] G. T. Google, “Gemma: Open models based on gemini research and technology,” arXiv preprint arXiv:2403.08295, 2024.
- [8] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language models are few-shot learners,” in Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2020.
- [9] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro, and Y. Zhang, “Sparks of artificial general intelligence: Early experiments with gpt-4,” arXiv preprint arXiv:2303.12712, 2023.
- [10] Y. Ge, W. Hua, K. Mei, J. Tan, S. Xu, Z. Li, and Y. Zhang, “Openagi: When llm meets domain experts,” in Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2024.

- [11] S. C. Bankes, “Agent-based modeling: A revolution?” Proceedings of the National Academy of Sciences, vol. 99, no. suppl_3, pp. 7199–7200, 2002.
- [12] Y. Bai, S. Kadavath, S. Kundu, A. Askill, J. Kernion, A. Jones, A. Chen, A. Goldie, A. Mirhoseini, C. McKinnon, C. Chen, C. Olsson, C. Olah, D. Hernandez, D. Drain, D. Ganguli, D. Li, E. Tran-Johnson, E. Perez, J. Kerr, J. Mueller, J. Ladish, J. Landau, K. Ndousse, K. Lukosuite, L. Lovitt, M. Sellitto, N. Elhage, N. Schiefer, N. Mercado, N. DasSarma, R. Lasenby, R. Larson, S. Ringer, S. Johnston, S. Kravec, S. E. Showk, S. Fort, T. Lanham, T. Telleen-Lawton, T. Conerly, T. Henighan, T. Hume, S. R. Bowman, Z. Hatfield-Dodds, B. Mann, D. Amodei, N. Joseph, S. McCandlish, T. Brown, and J. Kaplan, “Constitutional ai: Harmlessness from ai feedback,” arXiv preprint arXiv:2212.08073, 2022.
- [13] E. Bonabeau, “Agent-based modeling: Methods and techniques for simulating human systems,” Proceedings of the National Academy of Sciences, vol. 99, no. suppl_3, pp. 7280–7287, 2002.
- [14] G. Li, H. Hammoud, H. Itani, D. Khizbullin, and B. Ghanem, “Camel: Communicative agents for “mind” exploration of large language model society,” in Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2023.
- [15] Q. Wu, G. Bansal, J. Zhang, Y. Wu, B. Li, E. Zhu, L. Jiang, X. Zhang, S. Zhang, J. Liu, A. H. Awadallah, R. W. White, D. Burger, and C. Wang, “Autogen: Enabling next-gen llm applications via multi-agent conversation,” in First Conference on Language Modeling, 2024.
- [16] X. Hu, J. Chen, X. Li, Y. Guo, L. Wen, P. S. Yu, and Z. Guo, “Do large language models know about facts?” arXiv preprint arXiv:2310.05177, 2023.
- [17] S. M. Mousavi, S. Alghisi, and G. Riccardi, “Is your llm outdated? benchmarking llms & alignment algorithms for time-sensitive knowledge,” arXiv preprint arXiv:2404.08700, 2024.
- [18] Y. A. Yadkori, I. Kuzborskij, A. György, and C. Szepesvári, “To believe or not to believe your llm,” arXiv preprint arXiv:2406.02543, 2024.
- [19] A. Asai, Z. Wu, Y. Wang, A. Sil, and H. Hajishirzi, “Self-RAG: Learning to retrieve, generate, and critique through self-reflection,” in The Twelfth International Conference on Learning Representations (ICLR), 2024.
- [20] Y. Yu, W. Ping, Z. Liu, B. Wang, J. You, C. Zhang, M. Shoenybi, and B. Catanzaro, “RankRAG: Unifying context ranking with retrieval-augmented generation in LLMs,” in Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS), 2024.
- [21] H. Liu, R. Ning, Z. Teng, J. Liu, Q. Zhou, and Y. Zhang, “Evaluating the logical reasoning ability of chatgpt and gpt-4,” arXiv preprint arXiv:2304.03439, 2023.
- [22] K. Zhu, J. Chen, J. Wang, N. Z. Gong, D. Yang, and X. Xie, “Dyval: Dynamic evaluation of large language models for reasoning tasks,” in Proceedings of the International Conference on Learning Representations (ICLR), 2023.
- [23] H. H. Zhuo, X. Chen, and R. Pan, “On the roles of llms in planning: Embedding llms into planning graphs,” arXiv preprint arXiv:2403.00783, 2024.
- [24] C. Yuan, Q. Xie, J. Huang, and S. Ananiadou, “Back to the future: Towards explainable temporal reasoning with large language models,” in Proceedings of the International World Wide Web Conference (WWW), 2024.

- [25] J. Wang, Q. Sun, X. Li, and M. Gao, “Boosting language models reasoning with chain-of-knowledge prompting,” arXiv preprint arXiv:2306.06427, 2023.
- [26] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, and P. Fung, “Survey of hallucination in natural language generation,” ACM Computing Surveys, vol. 55, no. 12, pp. 1–38, 2023.
- [27] Z. Bai, P. Wang, T. Xiao, T. He, Z. Han, Z. Zhang, and M. Z. Shou, “Hallucination of multimodal large language models: A survey,” arXiv preprint arXiv:2404.18930, 2024.
- [28] S. Tonmoy, S. Zaman, V. Jain, A. Rani, V. Rawte, A. Chadha, and A. Das, “A comprehensive survey of hallucination mitigation techniques in large language models,” arXiv preprint arXiv:2401.01313, 2024.
- [29] J. Maynez, S. Narayan, B. Bohnet, and R. McDonald, “On faithfulness and factuality in abstractive summarization,” in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2020.
- [30] Y. Xiao and W. Y. Wang, “On hallucination and predictive uncertainty in conditional language generation,” in Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2021.
- [31] S. Farquhar, J. Kossen, L. Kuhn, and Y. Gal, “Detecting hallucinations in large language models using semantic entropy,” Nature, vol. 630, no. 8017, pp. 625–630, 2024.
- [32] Z. Ji, T. Yu, Y. Xu, N. Lee, E. Ishii, and P. Fung, “Towards mitigating llm hallucination via self reflection,” in Findings of the Association for Computational Linguistics: EMNLP (EMNLP-Findings), 2023.
- [33] C. Chen, K. Liu, Z. Chen, Y. Gu, Y. Wu, M. Tao, Z. Fu, and J. Ye, “Inside: Llms’ internal states retain the power of hallucination detection,” in Proceedings of the International Conference on Learning Representations (ICLR), 2024.
- [34] M. Wornow, Y. Xu, R. Thapa, B. Patel, E. Steinberg, S. Fleming, M. A. Pfeffer, J. Fries, and N. H. Shah, “The shaky foundations of large language models and foundation models for electronic health records,” NPJ Digital Medicine, vol. 6, no. 1, p. 135, 2023.
- [35] Y. Shen, L. Heacock, J. Elias, K. D. Hentel, B. Reig, G. Shih, and L. Moy, “Chatgpt and other large language models are double-edged swords,” Radiology, vol. 307, no. 2, p. e230163, 2023.
- [36] A. Pal, L. K. Umapathi, and M. Sankarasubbu, “Med-halt: Medical domain hallucination test for large language models,” in Proceedings of the Conference on Computational Natural Language Learning (CoNLL), 2023.
- [37] R. Xu, H. Liu, S. Nag, Z. Dai, Y. Xie, X. Tang, C. Luo, Y. Li, J. C. Ho, C. Yang, and Q. He, “Simrag: Self-improving retrieval-augmented generation for adapting large language models to specialized domains,” arXiv preprint arXiv:2410.17952, 2024.
- [38] D. P. Panagoulas, M. Virvou, and G. A. Tsihrintzis, “Evaluating llm-generated multimodal diagnosis from medical images and symptom analysis,” arXiv preprint arXiv:2402.01730, 2024.
- [39] Z. Gu, C. Yin, F. Liu, and P. Zhang, “Medvh: Towards systematic evaluation of hallucination for large vision language models in the medical context,” arXiv preprint arXiv:2407.02730, 2024.

- [40] A. Hogan, E. Blomqvist, M. Cochez, C. D’amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, and S. Neumaier, “Knowledge graphs,” ACM Computing Surveys, vol. 54, no. 4, pp. 1–37, 2021.
- [41] S. Ji, S. Pan, E. Cambria, P. Marttinen, and S. Y. Philip, “A survey on knowledge graphs: Representation, acquisition, and applications,” IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 2, pp. 494–514, 2021.
- [42] Q. Wang, Z. Mao, B. Wang, and L. Guo, “Knowledge graph embedding: A survey of approaches and applications,” IEEE Transactions on Knowledge and Data Engineering (TKDE), vol. 29, no. 12, pp. 2724–2743, 2017.
- [43] X. Zou, “A survey on application of knowledge graph,” in Journal of Physics: Conference Series, vol. 1487, no. 1. IOP Publishing, 2020, p. 012016.
- [44] H. Cui, J. Lu, S. Wang, R. Xu, W. Ma, S. Yu, Y. Yu, X. Kan, T. Fu, C. Ling, J. Ho, F. Wang, and C. Yang, “A survey on knowledge graphs for healthcare: Resources, application progress, and promise,” in ICML 3rd Workshop on Interpretable Machine Learning in Healthcare (IMLH), 2023.
- [45] S. Chang, Y. Hou, S. Rajendran, J. R. M. A. Maasch, Z. Abedi, H. Zhang, Z. Bai, A. Cuturrufo, W. Guo, F. F. Chaudhry, G. Ghahramani, J. Tang, F. Cheng, Y. Li, R. Zhang, J. Bian, and F. Wang, “Biomedical discovery through the integrative biomedical knowledge hub (ibkh),” iScience, vol. 26, no. 4, p. 106460, 2023.
- [46] R. Cornet and N. de Keizer, “Forty years of snomed: a literature review,” BMC Medical Informatics and Decision Making, vol. 8, pp. 1–6, 2008.
- [47] A. Santos, A. R. Colação, A. B. Nielsen, L. Niu, P. E. Geyer, F. Coscia, N. J. W. Albrechtsen, F. Mundt, L. J. Jensen, and M. Mann, “Clinical knowledge graph integrates proteomics data into clinical decision-making,” bioRxiv, pp. 2020–05, 2020.
- [48] J. E. Harrison, S. Weber, R. Jakob, and C. G. Chute, “Icd-11: an international classification of diseases for the twenty-first century,” BMC Medical Informatics and Decision Making, vol. 21, pp. 1–10, 2021.
- [49] C. E. Lipscomb, “Medical subject headings (mesh),” Bulletin of the Medical Library Association, vol. 88, no. 3, p. 265, 2000.
- [50] O. Bodenreider, “The unified medical language system (umls): integrating biomedical terminology,” Nucleic Acids Research, vol. 32, no. suppl_1, pp. D267–D270, 2004.
- [51] J. Xu, S. Kim, M. Song, M. Jeong, D. Kim, J. Kang, J. F. Rousseau, X. Li, W. Xu, V. I. Torvik, Y. Bu, C. Chen, I. A. Ebeid, D. Li, and Y. Ding, “Building a pubmed knowledge graph,” Scientific Data, vol. 7, no. 1, p. 205, 2020.
- [52] L. Li, P. Wang, J. Yan, Y. Wang, S. Li, J. Jiang, Z. Sun, B. Tang, T.-H. Chang, S. Wang, and Y. Liu, “Real-world data medical knowledge graph: construction and applications,” Artificial Intelligence in Medicine, vol. 103, p. 101817, 2020.
- [53] Q. Lv, G. Chen, H. He, Z. Yang, L. Zhao, K. Zhang, and C. Y.-C. Chen, “Tcmbank-the largest tcm database provides deep learning-based chinese-western medicine exclusion prediction,” Signal Transduction and Targeted Therapy, vol. 8, no. 1, p. 127, 2023.

- [54] C. Xiong, R. Power, and J. Callan, “Explicit semantic ranking for academic search via knowledge graph embedding,” in Proceedings of the International World Wide Web Conference (WWW), 2017.
- [55] Z. Liu, C. Xiong, M. Sun, and Z. Liu, “Entity-duet neural ranking: Understanding the role of knowledge graph semantics in neural information retrieval,” in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2018.
- [56] T. Vu, T. D. Nguyen, D. Q. Nguyen, D. Phung et al., “A capsule network-based embedding model for knowledge graph completion and search personalization,” in Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019.
- [57] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, “Kgat: Knowledge graph attention network for recommendation,” in Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD), 2019.
- [58] X. Wang, T. Huang, D. Wang, Y. Yuan, Z. Liu, X. He, and T.-S. Chua, “Learning intents behind interactions with knowledge graph for recommendation,” in Proceedings of the International World Wide Web Conference (WWW), 2021.
- [59] Y. Yang, C. Huang, L. Xia, and C. Li, “Knowledge graph contrastive learning for recommendation,” in Proceedings of the ACM International Conference on Research and Development in Information Retrieval (SIGIR), 2022.
- [60] F. Feng, F. Tang, Y. Gao, D. Zhu, T. Li, S. Yang, Y. Yao, Y. Huang, and J. Liu, “Genomickb: a knowledge graph for the human genome,” Nucleic Acids Research, vol. 51, no. D1, pp. D950–D956, 2023.
- [61] X. Shao, C. Li, H. Yang, X. Lu, J. Liao, J. Qian, K. Wang, J. Cheng, P. Yang, H. Chen, X. Xu, and X. Fan, “Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with spatalk,” Nature Communications, vol. 13, no. 1, p. 4429, 2022.
- [62] X. Quan, W. Cai, C. Xi, C. Wang, and L. Yan, “Aimedgraph: a comprehensive multi-relational knowledge graph for precision medicine,” Database, vol. 2023, p. baad006, 2023.
- [63] S. Zhang, X. Lin, and X. Zhang, “Discovering dti and ddi by knowledge graph with mhrw and improved neural network,” in Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021.
- [64] B. Li and H. Chen, “Prediction of compound synthesis accessibility based on reaction knowledge graph,” Molecules, vol. 27, no. 3, p. 1039, 2022.
- [65] J. Jeong, N. Lee, Y. Shin, and D. Shin, “Intelligent generation of optimal synthetic pathways based on knowledge graph inference and retrosynthetic predictions using reaction big data,” Journal of the Taiwan Institute of Chemical Engineers, vol. 130, p. 103982, 2022.
- [66] Y. Xu, X. Chu, K. Yang, Z. Wang, P. Zou, H. Ding, J. Zhao, Y. Wang, and B. Xie, “Seqcare: Sequential training with external medical knowledge graph for diagnosis prediction in healthcare data,” in Proceedings of the ACM Web Conference 2023, 2023, pp. 2819–2830.

- [67] P. Jiang, C. Xiao, A. R. Cross, and J. Sun, “Graphcare: Enhancing healthcare predictions with personalized knowledge graphs,” in The Twelfth International Conference on Learning Representations, 2024.
- [68] R. Xu, W. Shi, Y. Yu, Y. Zhuang, B. Jin, M. D. Wang, J. C. Ho, and C. Yang, “Ram-ehr: Retrieval augmentation meets clinical predictions on electronic health records,” in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2024.
- [69] P. Chen, Y. Lu, V. W. Zheng, X. Chen, and B. Yang, “Knowedu: A system to construct knowledge graph for education,” IEEE Access, vol. 6, pp. 31 553–31 563, 2018.
- [70] M. Rizun, “Knowledge graph application in education: a literature review,” Acta Universitatis Lodzensis. Folia Oeconomica, vol. 3, no. 342, pp. 7–19, 2019.
- [71] H. Tang, “Intelligent processing and classification of multisource health big data from the perspective of physical and medical integration,” Scientific Programming, vol. 2022, no. 1, p. 5799354, 2022.
- [72] L. Yuan, Y. Wang, P. M. Thompson, V. A. Narayan, and J. Ye, “Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data,” NeuroImage, vol. 61, no. 3, pp. 622–632, 2012.
- [73] J. Zhang, “Multi-source remote sensing data fusion: status and trends,” International Journal of Image and Data Fusion, vol. 1, no. 1, pp. 5–24, 2010.
- [74] Y. Huang, C. Du, Z. Xue, X. Chen, H. Zhao, and L. Huang, “What makes multi-modal learning better than single (provably),” in Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2021.
- [75] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, “Multimodal machine learning: A survey and taxonomy,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423–443, 2018.
- [76] J. N. Acosta, G. J. Falcone, P. Rajpurkar, and E. J. Topol, “Multimodal biomedical ai,” Nature Medicine, vol. 28, no. 9, pp. 1773–1784, 2022.
- [77] Q. Cai, H. Wang, Z. Li, and X. Liu, “A survey on multimodal data-driven smart healthcare systems: approaches and applications,” IEEE Access, vol. 7, pp. 133 583–133 599, 2019.
- [78] C. Zhang, X. Chu, L. Ma, Y. Zhu, Y. Wang, J. Wang, and J. Zhao, “M3care: Learning with missing modalities in multimodal healthcare data,” in Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD), 2022.
- [79] L. R. Soenksen, Y. Ma, C. Zeng, L. Boussieux, K. Villalobos Carballo, L. Na, H. M. Wiberg, M. L. Li, I. Fuentes, and D. Bertsimas, “Integrated multimodal artificial intelligence framework for healthcare applications,” NPJ Digital Medicine, vol. 5, no. 1, p. 149, 2022.
- [80] T. Shaik, X. Tao, L. Li, H. Xie, and J. D. Velásquez, “A survey of multimodal information fusion for smart healthcare: Mapping the journey from data to wisdom,” Information Fusion, p. 102040, 2023.
- [81] F. Krones, U. Marikkar, G. Parsons, A. Szmul, and A. Mahdi, “Review of multimodal machine learning approaches in healthcare,” arXiv preprint arXiv:2402.02460, 2024.

- [82] F. Cremonesi, V. Planat, V. Kalokyri, H. Kondylakis, T. Sanavia, V. M. M. Resinas, B. Singh, and S. Uribe, “The need for multimodal health data modeling: A practical approach for a federated-learning healthcare platform,” Journal of Biomedical Informatics, vol. 141, p. 104338, 2023.
- [83] D. Iakovidis and C. Smailis, “A semantic model for multimodal data mining in healthcare information systems,” in Quality of Life through Quality of Information. IOS Press, 2012, pp. 574–578.
- [84] A. Kline, H. Wang, Y. Li, S. Dennis, M. Hutch, Z. Xu, F. Wang, F. Cheng, and Y. Luo, “Multimodal machine learning in precision health: A scoping review,” NPJ Digital Medicine, vol. 5, no. 1, p. 171, 2022.
- [85] S. R. Stahlschmidt, B. Ulfenborg, and J. Synnergren, “Multimodal deep learning for biomedical data fusion: a review,” Briefings in Bioinformatics, vol. 23, no. 2, p. bbab569, 2022.
- [86] J. Pan, S. Razniewski, J.-C. Kalo, S. Singhanian, J. Chen, S. Dietze, H. Jabeen, J. Omeliyanenko, W. Zhang, M. Lissandrini, R. Biswas, G. de Melo, A. Bonifati, E. Vakaj, M. Dragoni, and D. Graux, “Large language models and knowledge graphs: Opportunities and challenges,” Transactions on Graph Data and Knowledge, 2023.
- [87] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu, “Unifying large language models and knowledge graphs: A roadmap,” IEEE Transactions on Knowledge and Data Engineering, vol. 1, no. 1, pp. 2:1–2:38, 2024.
- [88] S. Pan, Y. Zheng, and Y. Liu, “Integrating graphs with large language models: Methods and prospects,” IEEE Intelligent Systems, vol. 39, no. 1, pp. 64–68, 2024.
- [89] X. Wang, T. Gao, Z. Zhu, Z. Zhang, Z. Liu, J. Li, and J. Tang, “Kepler: A unified model for knowledge embedding and pre-trained language representation,” Transactions of the Association for Computational Linguistics, vol. 9, pp. 176–194, 2021.
- [90] D. Yu, C. Zhu, Y. Yang, and M. Zeng, “Jaket: Joint pre-training of knowledge graph and language understanding,” in Proceedings of the AAAI International Conference on Artificial Intelligence (AAAI), 2022.
- [91] M. Yasunaga, A. Bosselut, H. Ren, X. Zhang, C. D. Manning, P. S. Liang, and J. Leskovec, “Deep bidirectional language-knowledge graph pretraining,” in Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2022.
- [92] Z. Jiang, L. Zhong, M. Sun, J. Xu, R. Sun, H. Cai, S. Luo, and Z. Zhang, “Efficient knowledge infusion via kg-llm alignment,” arXiv preprint arXiv:2406.03746, 2024.
- [93] X. Wang, Q. He, J. Liang, and Y. Xiao, “Language models as knowledge embeddings,” in Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 2022.
- [94] Y. Yu, C. Xiong, S. Sun, C. Zhang, and A. Overwijk, “Coco-dr: Combating distribution shifts in zero-shot dense retrieval with contrastive and distributionally robust learning,” in Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.
- [95] X. Xie, Z. Li, X. Wang, Z. Xi, and N. Zhang, “Lambdakg: A library for pre-trained language model-based knowledge graph embeddings,” in Proceedings of the International Joint Conference on Natural Language Processing, 2023.

- [96] N. Choudhary and C. K. Reddy, “Complex logical reasoning over knowledge graphs using large language models,” arXiv preprint arXiv:2305.01157, 2023.
- [97] Z. Zhang, X. Liu, Y. Zhang, Q. Su, X. Sun, and B. He, “Pretrain-kge: learning knowledge representation from pretrained language models,” in Findings of the Association for Computational Linguistics: EMNLP (EMNLP-Findings), 2020.
- [98] P. Ke, H. Ji, Y. Ran, X. Cui, L. Wang, L. Song, X. Zhu, and M. Huang, “Jointgt: Graph-text joint representation learning for text generation from knowledge graphs,” in Findings of the Association for Computational Linguistics: ACL (ACL-Findings), 2021.
- [99] L. Yao, C. Mao, and Y. Luo, “Kg-bert: Bert for knowledge graph completion,” arXiv preprint arXiv:1909.03193, 2019.
- [100] B. Kim, T. Hong, Y. Ko, and J. Seo, “Multi-task learning for knowledge graph completion with pre-trained language models,” in Proceedings of the International Conference on Computational Linguistics (COLING), 2020.
- [101] X. Lv, Y. Lin, Y. Cao, L. Hou, J. Li, Z. Liu, P. Li, and J. Zhou, “Do pre-trained models benefit knowledge graph completion? a reliable evaluation and a reasonable approach,” in Findings of the Association for Computational Linguistics: ACL (ACL-Findings), 2022.
- [102] J. Shen, C. Wang, L. Gong, and D. Song, “Joint language semantic and structure embedding for knowledge graph completion,” in Proceedings of the International Conference on Computational Linguistics (COLING), 2022.
- [103] B. Choi and Y. Ko, “Knowledge graph extension with a pre-trained language model via unified learning method,” Knowledge-Based Systems, vol. 262, p. 110245, 2023.
- [104] B. Wang, T. Shen, G. Long, T. Zhou, Y. Wang, and Y. Chang, “Structure-augmented text representation learning for efficient knowledge graph completion,” in Proceedings of the International World Wide Web Conference (WWW), 2021.
- [105] L. Wang, W. Zhao, Z. Wei, and J. Liu, “Simkgc: Simple contrastive knowledge graph completion with pre-trained language models,” in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2022.
- [106] D. Li, B. Zhu, S. Yang, K. Xu, M. Yi, Y. He, and H. Wang, “Multi-task pre-training language model for semantic network completion,” ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 22, no. 11, pp. 1–20, 2023.
- [107] A. Saxena, A. Kochsiek, and R. Gemulla, “Sequence-to-sequence knowledge graph completion and question answering,” in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2022.
- [108] C. Chen, Y. Wang, B. Li, and K.-Y. Lam, “Knowledge is flat: A seq2seq generative framework for various knowledge graph completion,” in Proceedings of the International Conference on Computational Linguistics (COLING), 2022.
- [109] C. Chen, Y. Wang, A. Sun, B. Li, and K.-Y. Lam, “Dipping plms sauce: Bridging structure and text for effective knowledge graph completion via conditional soft prompting,” in Findings of the Association for Computational Linguistics: ACL (ACL-Findings), 2023.

- [110] J. Lovelace and C. Rosé, “A framework for adapting pre-trained language models to knowledge graph completion,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.
- [111] X. Xie, N. Zhang, Z. Li, S. Deng, H. Chen, F. Xiong, M. Chen, and H. Chen, “From discrimination to generation: Knowledge graph completion with generative transformer,” in Companion Proceedings of the International World Wide Web Conference (WWW), 2022.
- [112] Y. Zhu, X. Wang, J. Chen, S. Qiao, Y. Ou, Y. Yao, S. Deng, H. Chen, and N. Zhang, “Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities,” arXiv preprint arXiv:2305.13168, 2023.
- [113] H. Ye, N. Zhang, H. Chen, and H. Chen, “Generative knowledge graph construction: A review,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022, pp. 1–17.
- [114] C. Qin, A. Zhang, Z. Zhang, J. Chen, M. Yasunaga, and D. Yang, “Is chatgpt a general-purpose natural language processing task solver?” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.
- [115] V. K. Kommineni, B. König-Ries, and S. Samuel, “From human experts to machines: An llm supported approach to ontology and knowledge graph construction,” arXiv preprint arXiv:2403.08345, 2024.
- [116] B. Zhang and H. Soh, “Extract, define, canonicalize: An LLM-based framework for knowledge graph construction,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024, pp. 9820–9836.
- [117] J. Vizcarra, S. Haruta, and M. Kurokawa, “Representing the interaction between users and products via llm-assisted knowledge graph construction,” in Proceedings of the IEEE International Conference on Semantic Computing (ICSC), 2024.
- [118] S. Yu, T. Huang, M. Liu, and Z. Wang, “Bear: Revolutionizing service domain knowledge graph construction with llm,” in Proceedings of the International Conference on Service-Oriented Computing, 2023.
- [119] Y. Hu, F. Zou, J. Han, X. Sun, and Y. Wang, “Llm-tikg: Threat intelligence knowledge graph construction utilizing large language model,” Computers & Security, p. 103999, 2024.
- [120] L.-P. Meyer, C. Stadler, J. Frey, N. Radtke, K. Junghanns, R. Meissner, G. Dziwis, K. Bulert, and M. Martin, “Llm-assisted knowledge graph engineering: Experiments with chatgpt,” in Proceedings of the Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow, 2023.
- [121] M. Hofer, J. Frey, and E. Rahm, “Towards self-configuring knowledge graph construction pipelines using llms-a case study with rml,” in ESWC Workshop on Knowledge Graph Construction, 2024.
- [122] R. Yang, B. Yang, A. Feng, S. Ouyang, M. Blum, T. She, Y. Jiang, F. Lecue, J. Lu, and I. Li, “Graphusion: A rag framework for knowledge graph construction with a global perspective,” arXiv preprint arXiv:2410.17600, 2024.

- [123] I. Melnyk, P. Dognin, and P. Das, “Grapher: Multi-stage knowledge graph construction using pretrained language models,” in NeurIPS Workshop on Deep Generative Models and Downstream Applications, 2021.
- [124] A. Kumar, A. Pandey, R. Gadia, and M. Mishra, “Building knowledge graph using pre-trained language model for learning entity-aware relationships,” in Proceedings of the IEEE International Conference on Computing, Power and Communication Technologies, 2020.
- [125] X. Chen, N. Zhang, X. Xie, S. Deng, Y. Yao, C. Tan, F. Huang, L. Si, and H. Chen, “Know-prompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction,” in Proceedings of the International World Wide Web Conference (WWW), 2022.
- [126] Z. Wan, F. Cheng, Z. Mao, Q. Liu, H. Song, J. Li, and S. Kurohashi, “Gpt-re: In-context learning for relation extraction using large language models,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.
- [127] Y. Wang, H. Xin, and L. Chen, “Kglink: A column type annotation method that combines knowledge graph and pre-trained language model,” arXiv preprint arXiv:2406.00318, 2024.
- [128] H. Li, G. Appleby, and A. Suh, “A preliminary roadmap for llms as assistants in exploring, analyzing, and visualizing knowledge graphs,” arXiv preprint arXiv:2404.01425, 2024.
- [129] H. Yan, T. Gui, J. Dai, Q. Guo, Z. Zhang, and X. Qiu, “A unified generative framework for various ner subtasks,” in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2021.
- [130] N. Ding, Y. Chen, X. Han, G. Xu, X. Wang, P. Xie, H. Zheng, Z. Liu, J. Li, and H.-G. Kim, “Prompt-learning for fine-grained entity typing,” in Findings of the Association for Computational Linguistics: EMNLP (EMNLP-Findings), 2022.
- [131] S. Efeoglu and A. Paschke, “Retrieval-augmented generation-based relation extraction,” arXiv preprint arXiv:2404.13397, 2024.
- [132] B. Li, W. Yin, and M. Chen, “Ultra-fine entity typing with indirect supervision from natural language inference,” Transactions of the Association for Computational Linguistics, vol. 10, pp. 607–622, 2022.
- [133] T. Ayoola, S. Tyagi, J. Fisher, C. Christodoulopoulos, and A. Pierleoni, “Refined: An efficient zero-shot-capable approach to end-to-end entity linking,” in Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2022.
- [134] A. Cattan, A. Eirew, G. Stanovsky, M. Joshi, and I. Dagan, “Cross-document coreference resolution over predicted mentions,” in Findings of the Association for Computational Linguistics: ACL (ACL-Findings), 2021.
- [135] R. Thirukovalluru, N. Monath, K. Shridhar, M. Zaheer, M. Sachan, and A. McCallum, “Scaling within document coreference to long texts,” in Findings of the Association for Computational Linguistics: ACL (ACL-Findings), 2021.
- [136] C. Alt, M. Hübner, and L. Hennig, “Improving relation extraction by pre-trained language representations,” in Proceedings of the Conference on Automated Knowledge Base Construction (AKBC), 2019.

- [137] Q. Guo, Y. Sun, G. Liu, Z. Wang, Z. Ji, Y. Shen, and X. Wang, “Constructing chinese historical literature knowledge graph based on bert,” in Proceedings of the International Conference on Web Information Systems and Applications (WISA), 2021.
- [138] A. Bosselut, H. Rashkin, M. Sap, C. Malaviya, A. Celikyilmaz, and Y. Choi, “Comet: Commonsense transformers for knowledge graph construction,” in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2019.
- [139] S. Hao, B. Tan, K. Tang, H. Zhang, E. P. Xing, and Z. Hu, “Bertnet: Harvesting knowledge graphs from pretrained language models,” arXiv preprint arXiv:2206.14268, 2022.
- [140] P. West, C. Bhagavatula, J. Hessel, J. Hwang, L. Jiang, R. Le Bras, X. Lu, S. Welleck, and Y. Choi, “Symbolic knowledge distillation: from general language models to commonsense models,” in Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2022.
- [141] L. Hu, Z. Liu, Z. Zhao, L. Hou, L. Nie, and J. Li, “A survey of knowledge enhanced pre-trained language models,” IEEE Transactions on Knowledge and Data Engineering (TKDE), vol. 36, pp. 1413–1430, 2023.
- [142] X. Wei, S. Wang, D. Zhang, P. Bhatia, and A. Arnold, “Knowledge enhanced pretrained language models: A comprehensive survey,” arXiv preprint arXiv:2110.08455, 2021.
- [143] D. Yin, L. Dong, H. Cheng, X. Liu, K.-W. Chang, F. Wei, and J. Gao, “A survey of knowledge-intensive nlp with pre-trained language models,” arXiv preprint arXiv:2202.08772, 2022.
- [144] S. Wang, Z. Wei, J. Xu, T. Li, and Z. Fan, “Unifying structure reasoning and language model pre-training for complex reasoning,” arXiv preprint arXiv:2301.08913, 2023.
- [145] Y. Sun, Q. Shi, L. Qi, and Y. Zhang, “Jointlk: Joint reasoning with language models and knowledge graphs for commonsense question answering,” in Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2022.
- [146] Y. Liu, Y. Wan, L. He, H. Peng, and S. Y. Philip, “Kg-bart: Knowledge graph-augmented bart for generative commonsense reasoning,” in Proceedings of the AAAI International Conference on Artificial Intelligence (AAAI), 2021.
- [147] P. Qi, H. Lee, O. T. Sido, and C. D. Manning, “Answering open-domain questions of varying reasoning steps from text,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021.
- [148] C. Mavromatis and G. Karypis, “Gnn-rag: Graph neural retrieval for large language model reasoning,” arXiv preprint arXiv:2405.20139, 2024.
- [149] X. Zhang, A. Bosselut, M. Yasunaga, H. Ren, P. Liang, C. D. Manning, and J. Leskovec, “Greaselm: Graph reasoning enhanced language models,” in Proceedings of the International Conference on Learning Representations (ICLR), 2022.
- [150] Y. Feng, X. Chen, B. Y. Lin, P. Wang, J. Yan, and X. Ren, “Scalable multi-hop relational reasoning for knowledge-aware question answering,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.

- [151] M. Yasunaga, H. Ren, A. Bosselut, P. Liang, and J. Leskovec, “Qa-gnn: Reasoning with language models and knowledge graphs for question answering,” in Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2021.
- [152] Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu, “Ernie: Enhanced language representation with informative entities,” in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2019.
- [153] Y. Sun, S. Wang, S. Feng, S. Ding, C. Pang, J. Shang, J. Liu, X. Chen, Y. Zhao, Y. Lu, W. Liu, Z. Wu, W. Gong, J. Liang, Z. Shang, P. Sun, W. Liu, X. Ouyang, D. Yu, H. Tian, H. Wu, and H. Wang, “Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation,” arXiv preprint arXiv:2107.02137, 2021.
- [154] W. Liu, P. Zhou, Z. Zhao, Z. Wang, Q. Ju, H. Deng, and P. Wang, “K-bert: Enabling language representation with knowledge graph,” in Proceedings of the AAAI International Conference on Artificial Intelligence (AAAI), 2020.
- [155] D. Dai, L. Dong, Y. Hao, Z. Sui, B. Chang, and F. Wei, “Knowledge neurons in pretrained transformers,” in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2022.
- [156] T. Shen, Y. Mao, P. He, G. Long, A. Trischler, and W. Chen, “Exploiting structured knowledge in text via graph-guided representation learning,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.
- [157] D. Zhang, Z. Yuan, Y. Liu, F. Zhuang, H. Chen, and H. Xiong, “E-bert: A phrase and product knowledge enhanced language model for e-commerce,” arXiv preprint arXiv:2009.02835, 2020.
- [158] H. Tian, C. Gao, X. Xiao, H. Liu, B. He, H. Wu, H. Wang, and F. Wu, “Skep: Sentiment knowledge enhanced pre-training for sentiment analysis,” in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2020.
- [159] C. Rosset, C. Xiong, M. Phan, X. Song, P. Bennett, and S. Tiwary, “Knowledge-aware language model pretraining,” arXiv preprint arXiv:2007.00655, 2020.
- [160] S. Li, X. Li, L. Shang, C.-J. Sun, B. Liu, Z. Ji, X. Jiang, and Q. Liu, “Pre-training language models with deterministic factual knowledge,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.
- [161] W. Xiong, J. Du, W. Y. Wang, and V. Stoyanov, “Pretrained encyclopedia: Weakly supervised knowledge-pretrained language model,” in Proceedings of the International Conference on Learning Representations (ICLR), 2020.
- [162] B. He, D. Zhou, J. Xiao, X. Jiang, Q. Liu, N. J. Yuan, and T. Xu, “Bert-mk: Integrating graph contextualized knowledge into pre-trained language models,” in Findings of the Association for Computational Linguistics: EMNLP (EMNLP-Findings), 2020.
- [163] Y. Su, X. Han, Z. Zhang, Y. Lin, P. Li, Z. Liu, J. Zhou, and M. Sun, “Cokebert: Contextual knowledge selection and embedding towards enhanced pre-trained language models,” AI Open, vol. 2, pp. 127–134, 2021.

- [164] H. Zhu, H. Peng, Z. Lyu, L. Hou, J. Li, and J. Xiao, “Pre-training language model incorporating domain-specific heterogeneous knowledge into a unified representation,” Expert Systems with Applications, vol. 215, p. 119369, 2023.
- [165] C. Feng, X. Zhang, and Z. Fei, “Knowledge solver: Teaching llms to search for domain knowledge from knowledge graphs,” arXiv preprint arXiv:2309.03118, 2023.
- [166] N. Huang, Y. R. Deshpande, Y. Liu, H. Alberts, K. Cho, C. Vania, and I. Calixto, “Endowing language models with multimodal knowledge graph representations,” arXiv preprint arXiv:2206.13163, 2022.
- [167] O. Agarwal, H. Ge, S. Shakeri, and R. Al-Rfou, “Knowledge graph based synthetic corpus generation for knowledge-enhanced language model pre-training,” in Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2021.
- [168] Y. Xu, M. Namazifar, D. Hazarika, A. Padmakumar, Y. Liu, and D. Hakkani-Tur, “Kilm: Knowledge injection into encoder-decoder language models,” in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2023.
- [169] B. Oguz, X. Chen, V. Karpukhin, S. Peshterliev, D. Okhonko, M. Schlichtkrull, S. Gupta, Y. Mehdad, and S. Yih, “Unik-qa: Unified representations of structured and unstructured knowledge for open-domain question answering,” in Findings of the Association for Computational Linguistics: NAACL (NAACL-Findings), 2022.
- [170] Y. Tan, Z. Zhou, H. Lv, W. Liu, and C. Yang, “Walklm: A uniform language model fine-tuning framework for attributed graph embedding,” in Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2024.
- [171] R. Xu, W. Shi, Y. Yu, Y. Zhuang, Y. Zhu, M. D. Wang, J. C. Ho, C. Zhang, and C. Yang, “Bmretriever: Tuning large language models as better biomedical text retrievers,” in Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024.
- [172] T. Sun, Y. Shao, X. Qiu, Q. Guo, Y. Hu, X.-J. Huang, and Z. Zhang, “Colake: Contextualized language and knowledge embedding,” in Proceedings of the International Conference on Computational Linguistics (COLING), 2020.
- [173] T. Zhang, C. Wang, N. Hu, M. Qiu, C. Tang, X. He, and J. Huang, “Dkplm: decomposable knowledge-enhanced pre-trained language model for natural language understanding,” in Proceedings of the AAAI International Conference on Artificial Intelligence (AAAI), 2022.
- [174] H. Ye, N. Zhang, S. Deng, X. Chen, H. Chen, F. Xiong, X. Chen, and H. Chen, “Ontology-enhanced prompt-tuning for few-shot learning,” in Proceedings of the International World Wide Web Conference (WWW), 2022.
- [175] H. Luo, Z. Tang, S. Peng, Y. Guo, W. Zhang, C. Ma, G. Dong, M. Song, W. Lin, Y. Zhu, and L. A. Tuan, “Chatkbqa: A generate-then-retrieve framework for knowledge base question answering with fine-tuned large language models,” arXiv preprint arXiv:2310.08975, 2023.
- [176] A. Martino, M. Iannelli, and C. Truong, “Knowledge injection to counter large language model (llm) hallucination,” in Proceedings of the European Semantic Web Conference (ESWC), 2023.

- [177] W. Chen, Y. Su, X. Yan, and W. Y. Wang, “Kgpt: Knowledge-grounded pre-training for data-to-text generation,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.
- [178] C. N. d. Santos, Z. Dong, D. Cer, J. Nham, S. Shakeri, J. Ni, and Y.-H. Sung, “Knowledge prompts: Injecting world knowledge into language models through soft prompts,” arXiv preprint arXiv:2210.04726, 2022.
- [179] F. Moiseev, Z. Dong, E. Alfonseca, and M. Jaggi, “Skill: Structured knowledge infusion for large language models,” in Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2022.
- [180] S. Li, Y. Gao, H. Jiang, Q. Yin, Z. Li, X. Yan, C. Zhang, and B. Yin, “Graph reasoning for question answering with triplet retrieval,” in Findings of the Association for Computational Linguistics: ACL (ACL-Findings), 2023.
- [181] J. Jiang, K. Zhou, X. Zhao, and J.-R. Wen, “Unikgqa: Unified retrieval and reasoning for solving multi-hop question answering over knowledge graph,” in Proceedings of the International Conference on Learning Representations (ICLR), 2023.
- [182] L. Luo, Y.-F. Li, R. Haf, and S. Pan, “Reasoning on graphs: Faithful and interpretable large language model reasoning,” in Proceedings of the International Conference on Learning Representations (ICLR), 2024.
- [183] J. Sun, C. Xu, L. Tang, S. Wang, C. Lin, Y. Gong, L. Ni, H.-Y. Shum, and J. Guo, “Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph,” in Proceedings of the International Conference on Learning Representations (ICLR), 2024.
- [184] Y. Wang, B. Jiang, Y. Luo, D. He, P. Cheng, and L. Gao, “Reasoning on efficient knowledge paths: Knowledge graph guides large language model for domain question answering,” arXiv preprint arXiv:2404.10384, 2024.
- [185] L. Luo, J. Ju, B. Xiong, Y.-F. Li, G. Haffari, and S. Pan, “Chatrule: Mining logical rules with large language models for knowledge graph reasoning,” arXiv preprint arXiv:2309.01538, 2023.
- [186] Y. Wang, Z. Chu, X. Ouyang, S. Wang, H. Hao, Y. Shen, J. Gu, S. Xue, J. Y. Zhang, Q. Cui, L. Li, J. Zhou, and S. Li, “Enhancing recommender systems with large language model reasoning graphs,” arXiv preprint arXiv:2308.10835, 2023.
- [187] M. Besta, N. Blach, A. Kubicek, R. Gerstenberger, M. Podstawski, L. Gianinazzi, J. Gajda, T. Lehmann, H. Niewiadomski, P. Nyczyk, P. Nyczyk, and T. Hoefler, “Graph of thoughts: Solving elaborate problems with large language models,” in Proceedings of the AAAI International Conference on Artificial Intelligence (AAAI), 2024.
- [188] X. Wang, P. Kapanipathi, R. Musa, M. Yu, K. Talamadupula, I. Abdelaziz, M. Chang, A. Fokoue, B. Makni, N. Mattei, and M. Witbrock, “Improving natural language inference using external knowledge in the science questions domain,” in Proceedings of the AAAI International Conference on Artificial Intelligence (AAAI), 2019.
- [189] Y. Wang, N. Lipka, R. A. Rossi, A. Siu, R. Zhang, and T. Derr, “Knowledge graph prompting for multi-document question answering,” in Proceedings of the AAAI International Conference on Artificial Intelligence (AAAI), 2024.

- [190] D. Allemang and J. Sequeda, “Increasing the llm accuracy for question answering: Ontologies to the rescue!” arXiv preprint arXiv:2405.11706, 2024.
- [191] Z. Ji, Z. Liu, N. Lee, T. Yu, B. Wilie, M. Zeng, and P. Fung, “Rho: Reducing hallucination in open-domain dialogues with knowledge grounding,” in Findings of the Association for Computational Linguistics: ACL (ACL-Findings), 2023.
- [192] S. Feng, V. Balachandran, Y. Bai, and Y. Tsvetkov, “Factkb: Generalizable factuality evaluation using language models enhanced with factual knowledge,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.
- [193] D. McDonald, R. Papadopoulos, and L. Benningfield, “Reducing llm hallucination using knowledge distillation: A case study with mistral large and mmlu benchmark,” Preprint available at techrxiv.171665607, 2024.
- [194] Y. Lee, J. J. Y. Chung, T. S. Kim, J. Y. Song, and J. Kim, “Promptiverse: Scalable generation of scaffolding prompts through human-ai hybrid knowledge graph annotation,” in Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI), 2022.
- [195] R. Brate, M.-H. Dang, F. Hoppe, Y. He, A. Meroño-Peñuela, and V. Sadashivaiah, “Improving language model predictions via prompts enriched with knowledge graphs?” in ISWC Wrokshop on Deep Learning for Knowledge Graphs, 2022.
- [196] Y. Wen, Z. Wang, and J. Sun, “Mindmap: Knowledge graph prompting sparks graph of thoughts in large language models,” arXiv preprint arXiv:2308.09729, 2023.
- [197] D. Wilmot and F. Keller, “Memory and knowledge augmented language models for inferring salience in long-form stories,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021.
- [198] R. Logan, N. F. Liu, M. E. Peters, M. Gardner, and S. Singh, “Barack’s wife hillary: Using knowledge graphs for fact-aware language modeling,” in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2019.
- [199] Y. Wu, Y. Zhao, B. Hu, P. Minervini, P. Stenetorp, and S. Riedel, “An efficient memory-augmented transformer for knowledge-intensive nlp tasks,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.
- [200] X. Guan, Y. Liu, H. Lin, Y. Lu, B. He, X. Han, and L. Sun, “Mitigating large language model hallucinations via autonomous knowledge graph-based retrofitting,” in Proceedings of the AAAI International Conference on Artificial Intelligence (AAAI), 2024.
- [201] J. Dong, B. Fatemi, B. Perozzi, L. F. Yang, and A. Tsitsulin, “Don’t forget to connect! improving rag with graph-based reranking,” arXiv preprint arXiv:2405.18414, 2024.
- [202] P. Sarthi, S. Abdullah, A. Tuli, S. Khanna, A. Goldie, and C. D. Manning, “Raptor: Recursive abstractive processing for tree-organized retrieval,” in Proceedings of the International Conference on Learning Representations (ICLR), 2024.
- [203] X. He, Y. Tian, Y. Sun, N. V. Chawla, T. Laurent, Y. LeCun, X. Bresson, and B. Hooi, “G-retriever: Retrieval-augmented generation for textual graph understanding and question answering,” arXiv preprint arXiv:2402.07630, 2024.

- [204] J. Han, N. Collier, W. Buntine, and E. Shareghi, “Pive: Prompting with iterative verification improving graph-based generative capability of llms,” arXiv preprint arXiv:2305.12392, 2023.
- [205] D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody, S. Truitt, and J. Larson, “From local to global: A graph rag approach to query-focused summarization,” arXiv preprint arXiv:2404.16130, 2024.
- [206] B. J. Gutiérrez, Y. Shu, Y. Gu, M. Yasunaga, and Y. Su, “Hipporag: Neurobiologically inspired long-term memory for large language models,” arXiv preprint arXiv:2405.14831, 2024.
- [207] X. Liang, S. Niu, S. Zhang, S. Song, H. Wang, J. Yang, F. Xiong, B. Tang, and C. Xi, “Empowering large language models to set up a knowledge retrieval indexer via self-learning,” arXiv preprint arXiv:2405.16933, 2024.
- [208] F. Petroni, T. Rocktäschel, S. Riedel, P. Lewis, A. Bakhtin, Y. Wu, and A. Miller, “Language models as knowledge bases?” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019.
- [209] Z. Jiang, F. F. Xu, J. Araki, and G. Neubig, “How can we know what language models know?” Transactions of the Association for Computational Linguistics, vol. 8, pp. 423–438, 2020.
- [210] L. Adolphs, S. Dhuliawala, and T. Hofmann, “How to query language models?” arXiv preprint arXiv:2108.01928, 2021.
- [211] T. Shin, Y. Razeghi, R. L. Logan IV, E. Wallace, and S. Singh, “Autoprompt: Eliciting knowledge from language models with automatically generated prompts,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.
- [212] A. Mallen, A. Asai, V. Zhong, R. Das, H. Hajishirzi, and D. Khashabi, “When not to trust language models: Investigating effectiveness and limitations of parametric and non-parametric memories,” arXiv preprint arXiv:2212.10511, 2022.
- [213] W. W. Cohen, W. Chen, M. De Jong, N. Gupta, A. Presta, P. Verga, and J. Wieting, “Qa is the new kr: Question-answer pairs as knowledge bases,” in Proceedings of the AAAI International Conference on Artificial Intelligence (AAAI), 2023.
- [214] L. Luo, T. Vu, D. Phung, and R. Haf, “Systematic assessment of factual knowledge in large language models,” in Findings of the Association for Computational Linguistics: EMNLP (EMNLP-Findings), 2023.
- [215] A. Orogat, I. Liu, and A. El-Roby, “Cbench: towards better evaluation of question answering over knowledge graphs,” in Proceedings of the International Conference on Very Large Data Bases (VLDB), 2021.
- [216] Y. Bai, S. Feng, V. Balachandran, Z. Tan, S. Lou, T. He, and Y. Tsvetkov, “Kgquiz: Evaluating the generalization of encoded knowledge in large language models,” in Proceedings of the International World Wide Web Conference (WWW), 2024.
- [217] X. Ho, A.-K. D. Nguyen, S. Sugawara, and A. Aizawa, “Constructing a multi-hop qa dataset for comprehensive evaluation of reasoning steps,” in Proceedings of the International Conference on Computational Linguistics (COLING), 2020.

- [218] M. Tracy, M. Cerdá, and K. M. Keyes, “Agent-based modeling in public health: current applications and future directions,” Annual Review of Public Health, vol. 39, no. 1, pp. 77–94, 2018.
- [219] X. Tang, A. Zou, Z. Zhang, Z. Li, Y. Zhao, X. Zhang, A. Cohan, and M. Gerstein, “Medagents: Large language models as collaborators for zero-shot medical reasoning,” in ICLR Workshop on LLM Agents, 2024.
- [220] D. Kaur, S. Uslu, M. Durresi, and A. Durresi, “Llm-based agents utilized in a trustworthy artificial conscience model for controlling ai in medical applications,” in International Conference on Advanced Information Networking and Applications, 2024.
- [221] J. Li, S. Wang, M. Zhang, W. Li, Y. Lai, X. Kang, W. Ma, and Y. Liu, “Agent hospital: A simulacrum of hospital with evolvable medical agents,” arXiv preprint arXiv:2405.02957, 2024.
- [222] Y. Kim, C. Park, H. Jeong, Y. S. Chan, X. Xu, D. McDuff, C. Breazeal, and H. W. Park, “Adaptive collaboration strategy for llms in medical decision making,” arXiv preprint arXiv:2404.15155, 2024.
- [223] S. A. Gebreab, K. Salah, R. Jayaraman, M. H. ur Rehman, and S. Ellaham, “Llm-based framework for administrative task automation in healthcare,” in Proceedings of the International Symposium on Digital Forensics and Security, 2024.
- [224] L. Yue and T. Fu, “Ct-agent: Clinical trial multi-agent with large language model-based reasoning,” arXiv preprint arXiv:2404.14777, 2024.
- [225] B. Pan, J. Lu, K. Wang, L. Zheng, Z. Wen, Y. Feng, M. Zhu, and W. Chen, “Agentcoord: Visually exploring coordination strategy for llm-based multi-agent collaboration,” arXiv preprint arXiv:2404.11943, 2024.
- [226] Y. Xiao, J. Liu, Y. Zheng, X. Xie, J. Hao, M. Li, R. Wang, F. Ni, Y. Li, J. Luo, S. Jiao, and J. Peng, “Cellagent: An llm-driven multi-agent framework for automated single-cell data analysis,” bioRxiv, pp. 2024–05, 2024.
- [227] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, “Freebase: a collaboratively created graph database for structuring human knowledge,” in Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD), 2008.
- [228] F. M. Suchanek, G. Kasneci, and G. Weikum, “Yago: A large ontology from wikipedia and wordnet,” Journal of Web Semantics, vol. 6, no. 3, pp. 203–217, 2008.
- [229] D. Vrandečić and M. Krötzsch, “Wikidata: a free collaborative knowledgebase,” Communications of the ACM, vol. 57, no. 10, pp. 78–85, 2014.
- [230] Y. Yan, Y. Hou, Y. Xiao, R. Zhang, and Q. Wang, “Knownet: Guided health information seeking from llms via knowledge graph integration,” IEEE Transactions on Visualization and Computer Graphics, 2024.
- [231] J. Lu and C. Yang, “Open-world taxonomy and knowledge graph co-learning,” in Proceedings of the Conference on Automated Knowledge Base Construction (AKBC), 2022.
- [232] D. S. Himmelstein, A. Lizee, C. Hessler, L. Brueggeman, S. L. Chen, D. Hadley, A. Green, P. Khankhanian, and S. E. Baranzini, “Systematic integration of biomedical knowledge prioritizes drugs for repurposing,” Elife, vol. 6, p. e26726, 2017.

- [233] X. Liu, H. Hong, X. Wang, Z. Chen, E. Kharlamov, Y. Dong, and J. Tang, “Selfkg: Self-supervised entity alignment in knowledge graphs,” in Proceedings of the ACM Web Conference 2022, 2022, pp. 860–870.
- [234] Z. Wang, Q. Lv, X. Lan, and Y. Zhang, “Cross-lingual knowledge graph alignment via graph convolutional networks,” in Proceedings of the 2018 conference on empirical methods in natural language processing, 2018, pp. 349–357.
- [235] Q. Zhu, H. Wei, B. Sisman, D. Zheng, C. Faloutsos, X. L. Dong, and J. Han, “Collective multi-type entity alignment between knowledge graphs,” in Proceedings of The Web Conference 2020, 2020, pp. 2241–2252.
- [236] Y. Yan, L. Liu, Y. Ban, B. Jing, and H. Tong, “Dynamic knowledge graph alignment,” in Proceedings of the AAAI conference on artificial intelligence, vol. 35, no. 5, 2021, pp. 4564–4572.
- [237] P. Sancheti, K. Karlapalem, and K. Vemuri, “Llm driven web profile extraction for identical names,” in Companion Proceedings of the ACM on Web Conference 2024, 2024, pp. 1616–1625.
- [238] M. Agrawal, S. Hegselmann, H. Lang, Y. Kim, and D. Sontag, “Large language models are few-shot clinical information extractors,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.
- [239] J. Lu, J. Shen, B. Xiong, W. Ma, S. Staab, and C. Yang, “Hiprompt: Few-shot biomedical knowledge fusion via hierarchy-oriented prompting,” in Proceedings of the ACM International Conference on Research and Development in Information Retrieval (SIGIR), 2023.
- [240] Y. Xie, J. Lu, J. Ho, F. Nahab, X. Hu, and C. Yang, “Promptlink: Leveraging large language models for cross-source biomedical concept linking,” in Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2024, pp. 2589–2593.
- [241] R. Zhang, Y. Su, B. D. Trisedya, X. Zhao, M. Yang, H. Cheng, and J. Qi, “Autoalign: fully automatic and effective knowledge graph alignment enabled by large language models,” IEEE Transactions on Knowledge and Data Engineering, 2023.
- [242] L. Yang, H. Chen, Z. Li, X. Ding, and X. Wu, “Give us the facts: Enhancing large language models with knowledge graphs for fact-aware language modeling,” IEEE Transactions on Knowledge and Data Engineering, 2024.
- [243] N. Hassan, G. Zhang, F. Arslan, J. Caraballo, D. Jimenez, S. Gawsane, S. Hasan, M. Joseph, A. Kulkarni, A. K. Nayak et al., “Claimbuster: The first-ever end-to-end fact-checking system,” Proceedings of the VLDB Endowment, vol. 10, no. 12, pp. 1945–1948, 2017.
- [244] P. Nakov, D. Corney, M. Hasanain, F. Alam, T. Elsayed, A. Barrón-Cedeño, P. Papotti, S. Shaar, G. Da San Martino et al., “Automated fact-checking for assisting human fact-checkers,” in Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, 2021, pp. 4551–4558.
- [245] J. Yu, X. Wang, S. Tu, S. Cao, D. Zhang-Li, X. Lv, H. Peng, Z. Yao, X. Zhang, H. Li, C. Li, Z. Zhang, Y. Bai, Y. Liu, A. Xin, K. Yun, L. GONG, N. Lin, J. Chen, Z. Wu, Y. Qi, W. Li, Y. Guan, K. Zeng, J. Qi, H. Jin, J. Liu, Y. Gu, Y. Yao, N. Ding, L. Hou, Z. Liu, X. Bin, J. Tang, and J. Li, “KoLA: Carefully benchmarking world knowledge of large language models,” in The Twelfth International Conference on Learning Representations, 2024.

- [246] R. Xu, H. Cui, Y. Yu, X. Kan, W. Shi, Y. Zhuang, W. Jin, J. Ho, and C. Yang, “Knowledge-infused prompting: Assessing and advancing clinical text data generation with large language models,” in Findings of the Association for Computational Linguistics: ACL 2024, 2024.
- [247] B. AlKhamissi, M. Li, A. Celikyilmaz, M. Diab, and M. Ghazvininejad, “A review on language models as knowledge bases,” arXiv preprint arXiv:2204.06031, 2022.
- [248] Y. Wei, Q. Huang, Y. Zhang, and J. Kwok, “KICGPT: Large language model with knowledge in context for knowledge graph completion,” in Findings of the Association for Computational Linguistics: EMNLP 2023. Singapore: Association for Computational Linguistics, 2023, pp. 8667–8683.
- [249] Y. Wang, M. Hu, Z. Huang, D. Li, D. Yang, and X. Lu, “KC-GenRe: A knowledge-constrained generative re-ranking method based on large language models for knowledge graph completion,” in Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, 2024, pp. 9668–9680.
- [250] Z. Bi, J. Chen, Y. Jiang, F. Xiong, W. Guo, H. Chen, and N. Zhang, “Codekgc: Code language model for generative knowledge graph construction,” ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 23, no. 3, pp. 1–16, 2024.
- [251] Q. Zeng, Y. Bai, Z. Tan, Z. Wu, S. Feng, and M. Jiang, “Codetaxo: Enhancing taxonomy expansion with limited examples via code language prompts,” arXiv preprint arXiv:2408.09070, 2024.
- [252] A. Madaan, S. Zhou, U. Alon, Y. Yang, and G. Neubig, “Language models of code are few-shot commonsense learners,” in Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022, pp. 1384–1403.
- [253] W. Chen, X. Ma, X. Wang, and W. W. Cohen, “Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks,” Transactions on Machine Learning Research, 2023.
- [254] W. Shi, R. Xu, Y. Zhuang, Y. Yu, J. Zhang, H. Wu, Y. Zhu, J. C. Ho, C. Yang, and M. D. Wang, “Ehrgent: Code empowers large language models for few-shot complex tabular reasoning on electronic health records,” in ICLR 2024 Workshop on Large Language Model (LLM) Agents, 2024.
- [255] Y. Zhang, Z. Chen, L. Guo, Y. Xu, W. Zhang, and H. Chen, “Making large language models perform better in knowledge graph completion,” in Proceedings of the 32nd ACM International Conference on Multimedia. Association for Computing Machinery, 2024, p. 233–242.
- [256] P. Jiang, L. Cao, C. Xiao, P. Bhatia, J. Sun, and J. Han, “Kg-fit: Knowledge graph fine-tuning upon open-world knowledge,” arXiv preprint arXiv:2405.16412, 2024.
- [257] Z. Yang, L. Li, K. Lin, J. Wang, C.-C. Lin, Z. Liu, and L. Wang, “The dawn of lmms: Preliminary explorations with gpt-4v (ision),” arXiv preprint arXiv:2309.17421, 2023.
- [258] N. Fei, Z. Lu, Y. Gao, G. Yang, Y. Huo, J. Wen, H. Lu, R. Song, X. Gao, T. Xiang, H. Sun, and J.-R. Wen, “Towards artificial general intelligence via a multimodal foundation model,” Nature Communications, vol. 13, no. 1, p. 3094, 2022.

- [259] C. Li, Z. Gan, Z. Yang, J. Yang, L. Li, L. Wang, and J. Gao, “Multimodal foundation models: From specialists to general-purpose assistants,” Foundations and Trends in Computer Graphics and Vision, vol. 16, no. 1-2, pp. 1–214, 2024.
- [260] C. Li, C. Wong, S. Zhang, N. Usuyama, H. Liu, J. Yang, T. Naumann, H. Poon, and J. Gao, “Llava-med: Training a large language-and-vision assistant for biomedicine in one day,” in Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2024.
- [261] K. Zhang, J. Yu, E. Adhikarla, R. Zhou, Z. Yan, Y. Liu, Z. Liu, L. He, B. Davison, X. Li, H. Ren, S. Fu, J. Zou, W. Liu, J. Huang, C. Chen, Y. Zhou, T. Liu, X. Chen, Y. Chen, Q. Li, H. Liu, and L. Sun, “Biomedgpt: a unified and generalist biomedical generative pre-trained transformer for vision, language, and multimodal tasks,” arXiv preprint arXiv:2305.17100, 2023.
- [262] Z. Chen, Y. Zhang, Y. Fang, Y. Geng, L. Guo, X. Chen, Q. Li, W. Zhang, J. Chen, Y. Zhu *et al.*, “Knowledge graphs meet multi-modal learning: A comprehensive survey,” arXiv preprint arXiv:2402.05391, 2024.
- [263] L. Sun, K. Zhang, Q. Li, and R. Lou, “Umie: Unified multimodal information extraction with instruction tuning,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, no. 17, 2024, pp. 19 062–19 070.
- [264] J. Li, H. Li, Z. Pan, D. Sun, J. Wang, W. Zhang, and G. Pan, “Prompting chatgpt in mner: Enhanced multimodal named entity recognition with auxiliary refined knowledge,” in Findings of the Association for Computational Linguistics: EMNLP 2023, 2023, pp. 2787–2802.
- [265] X. Hu, J. Chen, A. Liu, S. Meng, L. Wen, and P. S. Yu, “Prompt me up: Unleashing the power of alignments for multimodal entity and relation extraction,” in Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 5185–5194.
- [266] Q. Yu, J. Li, Y. Wu, S. Tang, W. Ji, and Y. Zhuang, “Visually-prompted language model for fine-grained scene graph generation in an open world,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 21 560–21 571.
- [267] L. Li, J. Xiao, G. Chen, J. Shao, Y. Zhuang, and L. Chen, “Zero-shot visual relation detection via composite visual cues from large language models,” Advances in Neural Information Processing Systems, vol. 36, 2024.
- [268] R. Li, S. Zhang, D. Lin, K. Chen, and X. He, “From pixels to graphs: Open-vocabulary scene graph generation with vision-language models,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 28 076–28 086.
- [269] S. Chen, J. Tracey, A. Bies, and S. Strassel, “Schema learning corpus: Data and annotation focused on complex events,” in Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 2024, pp. 14 393–14 399.
- [270] H. Rasheed, M. Maaz, S. Shaji, A. Shaker, S. Khan, H. Cholakkal, R. M. Anwer, E. Xing, M.-H. Yang, and F. S. Khan, “Glamm: Pixel grounding large multimodal model,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 13 009–13 018.
- [271] H. Cui, X. Fang, Z. Zhang, R. Xu, X. Kan, X. Liu, Y. Yu, M. Li, Y. Song, and C. Yang, “Open visual knowledge extraction via relation-oriented multimodality model prompting,” in Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2024.

- [272] Z. Yang, H. Zhang, F. Chen, A. Bolimera, and M. Savvides, “Hierarchical knowledge graph construction from images for scalable e-commerce,” arXiv preprint arXiv:2410.21237, 2024.
- [273] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin et al., “A survey on large language model based autonomous agents,” Frontiers of Computer Science, vol. 18, no. 6, p. 186345, 2024.
- [274] S. Wu, O. Irsoy, S. Lu, V. Dabravolski, M. Dredze, S. Gehrmann, P. Kambadur, D. Rosenberg, and G. Mann, “Bloomberggpt: A large language model for finance,” arXiv preprint arXiv:2303.17564, 2023.
- [275] J. Cui, Z. Li, Y. Yan, B. Chen, and L. Yuan, “Chatlaw: Open-source legal large language model with integrated external knowledge bases,” arXiv preprint arXiv:2306.16092, 2023.
- [276] Y. Chen and J. Zou, “Genept: A simple but effective foundation model for genes and cells built from chatgpt,” bioRxiv, 2024.
- [277] K. Singhal, S. Azizi, T. Tu, S. S. Mahdavi, J. Wei, H. W. Chung, N. Scales, A. Tanwani, H. Cole-Lewis, S. Pfohl, P. Payne, M. Seneviratne, P. Gamble, C. Kelly, N. Scharli, A. Chowdhery, P. Mansfield, B. Aguera y Arcas, D. Webster, G. S. Corrado, Y. Matias, K. Chou, J. Gottweis, N. Tomasev, Y. Liu, A. Rajkomar, J. Barral, C. Sementurs, A. Karthikesalingam, and V. Natarajan, “Large language models encode clinical knowledge,” Nature, vol. 620, no. 7972, pp. 172–180, 2023.
- [278] K. Singhal, T. Tu, J. Gottweis, R. Sayres, E. Wulczyn, L. Hou, K. Clark, S. Pfohl, H. Cole-Lewis, D. Neal, M. Schaeckermann, A. Wang, M. Amin, S. Lachgar, P. Mansfield, S. Prakash, B. Green, E. Dominowska, B. Aguera y Arcas, N. Tomasev, Y. Liu, R. Wong, C. Sementurs, S. S. Mahdavi, J. Barral, D. Webster, G. S. Corrado, Y. Matias, S. Azizi, A. Karthikesalingam, and V. Natarajan, “Towards expert-level medical question answering with large language models,” arXiv preprint arXiv:2305.09617, 2023.
- [279] C. E. Haupt and M. Marks, “Ai-generated medical advice? gpt and beyond,” JAMA, vol. 329, no. 16, pp. 1349–1350, 2023.
- [280] H. Nori, N. King, S. M. McKinney, D. Carignan, and E. Horvitz, “Capabilities of gpt-4 on medical challenge problems,” arXiv preprint arXiv:2303.13375, 2023.
- [281] P. Lee, S. Bubeck, and J. Petro, “Benefits, limits, and risks of gpt-4 as an ai chatbot for medicine,” New England Journal of Medicine, vol. 388, no. 13, pp. 1233–1239, 2023.
- [282] S. L. Fleming, K. Morse, A. Kumar, C.-C. Chiang, B. Patel, E. Brunskill, and N. Shah, “Assessing the potential of usmle-like exam questions generated by gpt-4,” medRxiv, pp. 2023–04, 2023.
- [283] Z. Chen, A. H. Cano, A. Romanou, A. Bonnet, K. Matoba, F. Salvi, M. Pagliardini, S. Fan, A. Köpf, A. Mohtashami, A. Sallinen, A. Sakhaeirad, V. Swamy, I. Krawczuk, D. Bayazit, A. Marmet, S. Montariol, M.-A. Hartley, M. Jaggi, and A. Bosselut, “Meditron-70b: Scaling medical pretraining for large language models,” arXiv preprint arXiv:2311.16079, 2023.
- [284] X. Yang, A. Chen, N. PourNejatian, H. C. Shin, K. E. Smith, C. Parisien, C. Compas, C. Martin, A. B. Costa, M. G. Flores, Y. Zhang, T. Magoc, C. A. Harle, G. Lipori, D. A. Mitchell, W. R. Hogan, E. A. Shenkman, J. Bian, and Y. Wu, “A large language model for electronic health records,” NPJ Digital Medicine, vol. 5, no. 1, p. 194, 2022.

- [285] R. Luo, L. Sun, Y. Xia, T. Qin, S. Zhang, H. Poon, and T.-Y. Liu, “Biogpt: generative pre-trained transformer for biomedical text generation and mining,” Briefings in Bioinformatics, vol. 23, no. 6, p. bbac409, 2022.
- [286] N. Mehandru, B. Y. Miao, E. R. Almaraz, M. Sushil, A. J. Butte, and A. Alaa, “Evaluating large language models as agents in the clinic,” NPJ Digital Medicine, vol. 7, no. 1, p. 84, 2024.
- [287] S. S. Biswas, “Role of chat gpt in public health,” Annals of Biomedical Engineering, vol. 51, no. 5, pp. 868–869, 2023.
- [288] D. Dash, R. Thapa, J. M. Banda, A. Swaminathan, M. Cheatham, M. Kashyap, N. Kotecha, J. H. Chen, S. Gombar, L. Downing, R. Pedreira, E. Goh, A. Arnaout, G. K. Morris, H. Magon, M. P. Lungren, E. Horvitz, and N. H. Shah, “Evaluation of gpt-3.5 and gpt-4 for supporting real-world information needs in healthcare delivery,” arXiv preprint arXiv:2304.13714, 2023.
- [289] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, “Retrieval-augmented generation for knowledge-intensive nlp tasks,” in Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2020.
- [290] Y. Bang, S. Cahyawijaya, N. Lee, W. Dai, D. Su, B. Wilie, H. Lovenia, Z. Ji, T. Yu, W. Chung et al., “A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity,” in Proceedings of the International Joint Conference on Natural Language Processing and Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics, 2023.
- [291] X. Kang, L. Qu, L.-K. Soon, Z. Li, and A. Trakic, “Bridging law and data: Augmenting reasoning via a semi-structured dataset with irac methodology,” arXiv preprint arXiv:2406.13217, 2024.
- [292] Y. Liu, Q. Zeng, J. Ordieres Meré, and H. Yang, “Anticipating stock market of the renowned companies: a knowledge graph approach,” Complexity, vol. 2019, 2019.
- [293] Y.-L. Tuan, S. Beygi, M. Fazel-Zarandi, Q. Gao, A. Cervone, and W. Y. Wang, “Towards large-scale interpretable knowledge graph reasoning for dialogue systems,” in Findings of the Association for Computational Linguistics (ACL-Findings), 2022.
- [294] J. Baek, A. F. Aji, J. Lehmann, and S. J. Hwang, “Direct fact retrieval from knowledge graphs without entity linking,” in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2023.
- [295] L. Wang, W. Xu, Y. Lan, Z. Hu, Y. Lan, R. K.-W. Lee, and E.-P. Lim, “Plan-and-solve prompting: Improving zero-shot chain-of-thought reasoning by large language models,” in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2023.
- [296] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. R. Narasimhan, and Y. Cao, “React: Synergizing reasoning and acting in language models,” in Proceedings of the International Conference on Learning Representations (ICLR), 2022.
- [297] H. Wang, H. Ren, and J. Leskovec, “Relational message passing for knowledge graph completion,” in Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD), 2021.

- [298] X. Xu, P. Zhang, Y. He, C. Chao, and C. Yan, “Subgraph neighboring relations infomax for inductive link prediction on knowledge graphs,” in Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 2022.
- [299] K. Wang, F. Duan, S. Wang, P. Li, Y. Xian, C. Yin, W. Rong, and Z. Xiong, “Knowledge-driven cot: Exploring faithful reasoning in llms for knowledge-intensive question answering,” arXiv preprint arXiv:2308.13259, 2023.
- [300] J. Huang, X. Zhang, Q. Mei, and J. Ma, “Can llms effectively leverage graph structural information through prompts, and why?” Transactions on Machine Learning Research, 2024.
- [301] L. Luo, Z. Zhao, C. Gong, G. Haffari, and S. Pan, “Graph-constrained reasoning: Faithful reasoning on knowledge graphs with large language models,” arXiv preprint arXiv:2410.13080, 2024.
- [302] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, and D. Zhou, “Chain-of-thought prompting elicits reasoning in large language models,” in Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2022.
- [303] E. Fredkin, “Trie memory,” Communications of the ACM, vol. 3, no. 9, pp. 490–499, 1960.
- [304] L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin et al., “A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions,” arXiv preprint arXiv:2311.05232, 2023.
- [305] G. Agrawal, T. Kumarage, Z. Alghamdi, and H. Liu, “Can knowledge graphs reduce hallucinations in llms?: A survey,” arXiv preprint arXiv:2311.07914, 2023.
- [306] T. Nguyen, L. Luo, F. Shiri, D. Phung, Y.-F. Li, T.-T. Vu, and G. Haffari, “Direct evaluation of chain-of-thought in multi-hop reasoning with knowledge graphs,” in Findings of the Association for Computational Linguistics: ACL (ACL-Findings), 2024.
- [307] D. C. Nguyen, Q.-V. Pham, P. N. Pathirana, M. Ding, A. Seneviratne, Z. Lin, O. Dobre, and W.-J. Hwang, “Federated learning for smart healthcare: A survey,” ACM Computing Surveys, vol. 55, no. 3, pp. 1–37, 2022.
- [308] R. S. Antunes, C. André da Costa, A. Küderle, I. A. Yari, and B. Eskofier, “Federated learning for healthcare: Systematic review and architecture proposal,” ACM Transactions on Intelligent Systems and Technology (TIST), vol. 13, no. 4, pp. 1–23, 2022.
- [309] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, “Federated learning for healthcare informatics,” Journal of Healthcare Informatics Research, vol. 5, pp. 1–19, 2021.
- [310] N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, S. Ourselin, M. Sheller, R. M. Summers, A. Trsk, D. Xu, M. Baust, and M. J. Cardoso, “The future of digital health with federated learning,” NPJ Digital Medicine, vol. 3, no. 1, pp. 1–7, 2020.
- [311] A. Wei, N. Haghtalab, and J. Steinhardt, “Jailbroken: How does llm safety training fail?” in Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2024.
- [312] B. Wang, W. Chen, H. Pei, C. Xie, M. Kang, C. Zhang, C. Xu, Z. Xiong, R. Dutta, R. Schaeffer, S. T. Truong, S. Arora, M. Mazeika, D. Hendrycks, Z. Lin, Y. Cheng, S. Koyejo, D. Song, and B. Li, “Decodingtrust: A comprehensive assessment of trustworthiness in gpt models,” in Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2023.

- [313] Z. Xu, Y. Liu, G. Deng, Y. Li, and S. Picek, “Llm jailbreak attack versus defense techniques—a comprehensive study,” arXiv preprint arXiv:2402.13457, 2024.
- [314] C. He, K. Balasubramanian, E. Ceyani, C. Yang, H. Xie, L. Sun, L. He, L. Yang, P. S. Yu, Y. Rong, P. Zhao, J. Huang, M. Annavaram, and S. Avestimehr, “Fedgraphnn: A federated learning system and benchmark for graph neural networks,” in ICLR Workshop on Distributed and Private Machine Learning, 2021.
- [315] H. Xie, J. Ma, L. Xiong, and C. Yang, “Federated graph classification over non-iid graphs,” in Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS), 2021.
- [316] K. Zhang, C. Yang, X. Li, L. Sun, and S. M. Yiu, “Subgraph federated learning with missing neighbor generation,” in Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS), 2021.
- [317] K. Zhang, L. Sun, B. Ding, S. M. Yiu, and C. Yang, “Deep efficient private neighbor generation for subgraph federated learning,” in Proceedings of the 2024 SIAM International Conference on Data Mining (SDM), 2024.
- [318] H. Xie, L. Xiong, and C. Yang, “Federated node classification over graphs with latent link-type heterogeneity,” in Proceedings of the International World Wide Web Conference (WWW), 2023.
- [319] K. Zhang, Y. Wang, H. Wang, L. Huang, C. Yang, X. Chen, and L. Sun, “Efficient federated learning on knowledge graphs via privacy-preserving relation embedding aggregation,” in Findings of the Association for Computational Linguistics: EMNLP (EMNLP-Findings), 2022.
- [320] Z. Gu, K. Zhang, G. Bai, L. Chen, L. Zhao, and C. Yang, “Dynamic activation of clients and parameters for federated learning over heterogeneous graphs,” in Proceedings of the IEEE International Conference on Data Engineering (ICDE), 2023.
- [321] H. Xie, L. Xiong, and C. Yang, “Federated node classification over distributed ego-networks with secure contrastive embedding sharing,” in Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM), 2024.
- [322] T. Rebedea, R. Dinu, M. N. Sreedhar, C. Parisien, and J. Cohen, “Nemo guardrails: A toolkit for controllable and safe llm applications with programmable rails,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.
- [323] Z. Xiang, L. Zheng, Y. Li, J. Hong, Q. Li, H. Xie, J. Zhang, Z. Xiong, C. Xie, C. Yang, D. Song, and B. Li, “Guardagent: Safeguard llm agents by a guard agent via knowledge-enabled reasoning,” arXiv preprint arXiv:2406.09187, 2024.
- [324] S. Artiga, K. Orgera, and O. Pham, “Disparities in health and health care: Five key questions and answers,” Kaiser Family Foundation, 2020.
- [325] S. Artiga and E. Hinton, “Beyond health care: the role of social determinants in promoting health and health equity,” Kaiser Family Foundation, 2018.
- [326] W. C. on Social Determinants of Health and W. H. Organization, Closing the gap in a generation: health equity through action on the social determinants of health: Commission on Social Determinants of Health final report. World Health Organization, 2008.

- [327] M. Marmot and R. Wilkinson, Social determinants of health. Oup Oxford, 2005.
- [328] J. C. Phelan, B. G. Link, and P. Tehranifar, “Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications,” Journal of Health and Social Behavior, vol. 51, no. 1_suppl, pp. S28–S40, 2010.
- [329] R. Yang, H. Liu, Q. Zeng, Y. H. Ke, W. Li, L. Cheng, Q. Chen, J. Caverlee, Y. Matsuo, and I. Li, “Kg-rank: Enhancing large language models for medical qa with knowledge graphs and ranking techniques,” arXiv preprint arXiv:2403.05881, 2024.
- [330] M. Dehghan, M. Alomrani, S. Bagga, D. Alfonso-Hermelo, K. Bibi, A. Ghaddar, Y. Zhang, X. Li, J. Hao, Q. Liu, J. Lin, B. Chen, P. Parthasarathi, M. Biparva, and M. Rezagholizadeh, “EWEK-QA : Enhanced web and efficient knowledge graph retrieval for citation-based question answering systems,” in Findings of the Association for Computational Linguistics: ACL (ACL-Findings), 2024.
- [331] R. Xu, Z. Qi, C. Wang, H. Wang, Y. Zhang, and W. Xu, “Knowledge conflicts for llms: A survey,” arXiv preprint arXiv:2403.08319, 2024.
- [332] Y. Zhao, A. Devoto, G. Hong, X. Du, A. P. Gema, H. Wang, K.-F. Wong, and P. Minervini, “Steering knowledge selection behaviours in llms via sae-based representation engineering,” arXiv preprint arXiv:2410.15999, 2024.
- [333] F. Wang, X. Wan, R. Sun, J. Chen, and S. Ö. Arik, “Astute rag: Overcoming imperfect retrieval augmentation and knowledge conflicts for large language models,” arXiv preprint arXiv:2410.07176, 2024.
- [334] X. Zhu, Z. Li, X. Wang, X. Jiang, P. Sun, X. Wang, Y. Xiao, and N. J. Yuan, “Multi-modal knowledge graph construction and application: A survey,” IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 2, pp. 715–735, 2022.
- [335] X. Long, J. Zeng, F. Meng, Z. Ma, K. Zhang, B. Zhou, and J. Zhou, “Generative multi-modal knowledge retrieval with large language models,” in Proceedings of the AAAI International Conference on Artificial Intelligence (AAAI), 2024.
- [336] A. Kumar, C. Agarwal, S. Srinivas, A. J. Li, S. Feizi, and H. Lakkaraju, “Certifying llm safety against adversarial prompting,” arXiv preprint arXiv:2309.02705, 2023.
- [337] P. Cheng, Y. Ding, T. Ju, Z. Wu, W. Du, P. Yi, Z. Zhang, and G. Liu, “Trojanrag: Retrieval-augmented generation can be backdoor driver in large language models,” arXiv preprint arXiv:2405.13401, 2024.