A Survey on Unifying Large Language Models and Knowledge Graphs for Biomedicine and Healthcare

Ran Xu
Department of Computer Science,
Emory Universit
Atlanta, USA
ran.xu@emory.edu

Patrick Jiang
Department of Computer Science,
UIUC
Urbana, USA
pj20@illinois.edu

Linhao Luo
Department of Computer Science,
Monash University
Melbourne, Australia
linhao.luo@monash.edu

Cao Xiao
GE HealthCare
Seattle, USA
cao.xiao@gehealthcare.com

Adam Cross
Department of Pediatrics, UIC
Chicago, USA
arcross@uic.edu

Shirui Pan
School of ICT, Griffith University
Brisbane, Australia
s.pan@griffith.edu.au

Jimeng Sun
Department of Computer Science,
UIUC
Urbana, USA
jimeng@illinois.edu

Carl Yang*
Department of Computer Science,
Emory University
Atlanta, USA
j.carlyang@emory.edu

Abstract

In recent years, the landscape of digital biomedicine and healthcare has been reshaped due to the disruptive breakthroughs in AI–facilitated by tremendous data and high-performance computers, large language models (LLMs) have transformed information technology from accessing data to performing analytical tasks. While demonstrating unprecedented capabilities, LLMs have been found unreliable in tasks requiring factual knowledge and rigorous reasoning. Biomedicine and healthcare, as an important vertical domain rapidly benefitting from progress in AI, necessitates strict requirements on the accuracy, controllability, and interpretability of analytical models, posing critical challenges for LLMs. Despite recent studies addressing the hallucination problem of LLMs, research on empowering LLMs with the ability to plan, reason, and ground with explicit knowledge has also started to prosper, especially in the biomedicine and healthcare domain. On the other hand, biomedical data are enormous and notoriously complex, coming from various sources (*e.g.*, biomedical knowledge bases, online literature, and hospitals) and bearing various modalities (*e.g.*, tables, texts, images and time-series). Healthcare professionals have spent decades collecting, cleaning, and curating various types of data. The processes are extremely costly, producing various datasets with different data schemas, coding systems, and quality standards, many privately owned by the creators, making their integrative analysis and utilization through unified AI techniques still rather challenging. The generalizability of LLMs across different types of data endow them strong promises in automating the processing of large-scale complex healthcare data such as into unified knowledge graphs (KGs). Our goal in this survey is to systematically investigate and summarize recent studies on the unification of LLMs and KGs, towards fully utilizing the value of complex data, unleashing the power of generative AI, and expediting next-generation AI for biomedicine and healthcare applications.

CCS Concepts

• **Applied computing** → **Life and medical sciences**.

Keywords

large language model, knowledge graph, biomedical sciences, health informatics

ACM Reference Format:

Ran Xu, Patrick Jiang, Linhao Luo, Cao Xiao, Adam Cross, Shirui Pan, Jimeng Sun, and Carl Yang. 2025. A Survey on Unifying Large Language Models and Knowledge Graphs for Biomedicine and Healthcare. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25), August 3–7, 2025, Toronto, ON, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3711896.3736556

1 Introduction

Large language models (LLMs) have reshaped AI research and implementations, with unprecedented capabilities widely shown in various text-related tasks, bringing humans ever close to general AI. Recent research on multi-agent systems have further magnified LLMs' advantages of *language comprehension*, *broad knowledge* and *generalizability* through conversations, showing strong promises for deep human-model collaboration for critical applications [1]. In biomedicine and healthcare, extensive enthusiasm

*Corresponding author

KDD '25, Toronto, ON, Canada

ACM ISBN 979-8-4007-1454-2/2025/08
https://doi.org/10.1145/3711896.3736556

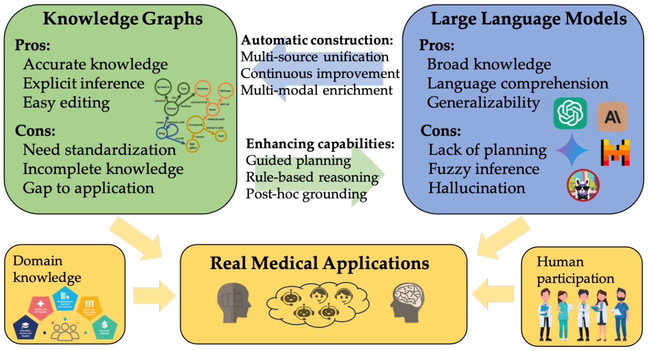

Figure 1: Overview of the proposed biomedical knowledge language models framework.

has been witnessed on the exploration and evaluation of LLMs in answering medical questions [2], extracting clinical information [3] and assisting clinical decisions [4]. Studies have also revealed the limitations of LLMs regarding their *lack of knowledge* [5], *fuzzy inference* [6], and *hallucination* [7]. Specifically, in biomedicine and healthcare, the lack of knowledge can be caused by the lack of access to high-quality data about various biomedical concepts and patient conditions, as well as the rapidly evolving new biomedical knowledge; the fuzzy inference nature can lead to difficulties in conducting reliable comprehension and stable predictions for complex medical questions; and hallucination creating factual errors and misinformation can cause fatal and life-threatening problems in the healthcare workflows [8].

Knowledge graph (KG) has been widely studied across academia and industry, due to its advantages in storing *accurate*, *explicit* and *easily-modifiable* knowledge [9]. In biomedicine and healthcare, researchers and professionals have spent decades collecting, processing, and curating various types of biomedical and clinical data towards the construction of medical KGs [10], which are widely used to support basic science research [11], pharmaceutical research [12], clinical decisions [13] and policy making [14]. However, biomedical data are notoriously noisy and complex, where datasets about specific concepts and conditions come from various sources such as institutions using different data schemas and coding conventions [15], and the data can also include multiple modalities such as tables, texts, images, and time-series [16]. While such *multi-source* and *multi-modality* data hold great promises in integrative and comprehensive biomedical analysis, extracting and unifying high-quality knowledge from them is non-trivial.

Recently, significant research attention has been drawn to the synergies between KGs and LLMs [17], due to their naturally complementary advantages (Figure 1). The construction and modeling of KGs have always relied on advances in natural language processing (NLP) tools, and nowadays, researchers have intensively explored language models towards the embedding [18], completion [19] and construction [20] 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 [21] or inference [22], and enabling knowledge-based interpretation and evaluation [23]. In very recent years, pioneering studies have also started to explore the combination of KGs and LLMs for biomedicine and healthcare [24]. Most of these studies have focused on specific healthcare applications and only implemented shallow and straightforward technical designs, without fundamentally improving the KGs and LLMs.

To expedite LLM-based research towards next-generation AI for health, this survey will comprehensively investigate and summarize recent works addressing the *data*, *model* and *application* challenges of unifying KGs and LLMs for biomedicine and healthcare, through a systematic conceptual framework of `BioMedKLM` (BioMedical Knowledge Language Models, Figure 1). We present `BioMedKLM` to discuss major functionality needed to build high-quality KGs that integrate complex *biomedical data*, enhance LLMs to obtain reliable *biomedical models*, and properly employ the data and model to enable critical and novel *biomedical applications*. This survey, as illustrated in Figure 1, consists of three distinct but interrelated research perspectives that address the data, model and application challenges blocking the transformation of AI for health.

- **Perspective P1: LLM-aided KG construction from multi-source multi-modality biomedical data.** We will discuss LLM-based methods for unifying existing biomedical KGs collectable from different sources (P1.1), continuously enhance biomedical KGs by extracting concepts and relations from evolving biomedical literature, and (P1.3) comprehensively augment biomedical KGs through the integration of multi-modal biomedical data.
- **Perspective P2: KG-guided LLM enhancement towards reliable biomedical models.** We will discuss the utilization of biomedical KGs to enhance the capabilities of LLMs through providing biomedical knowledge to enhance LLM planning (P2.1), enabling biomedical neural symbolic reasoning with LLMs (P2.2), and enforcing post-hoc biomedical error detection to verify LLMs' knowledge and reasoning (P2.3).
- **Perspective P3: Knowledge language co-modeling to facilitate reliable biomedical applications.** We will discuss how the integration of KGs and LLMs enables critical biomedical applications including clinical decision support for enhanced diagnostics (P3.1), drug discovery and development for accelerated pharmaceutical research (P3.2), and biomedical knowledge management for organizing the vast body of biomedical information (P3.3).

This survey establishes a systematic and comprehensive framework of `BioMedKLM` that aims to enable the synergistic and progressive improvements of KGs and LLMs for biomedicine and healthcare. Techniques related to `BioMedKLM` reflect methodological innovations in data mining and generative AI, fundamentally improving KGs via automatic extraction and integration of multi-source multi-modality knowledge, enhancing LLMs towards knowledge-guided planning, reasoning and grounding, and unleashing the power of generative AI towards ethical, trustworthy and human-centered biomedical applications.

2 LLM-aided KG Construction

In this section, we examine the benefits of leveraging LLMs for medical KG construction, particularly in *accuracy*, *consistency*, *coverage*, and *freshness* of knowledge. Traditional approaches, as surveyed in [10], often focus on specific disease areas or entity types. Efforts to build comprehensive biomedical KGs primarily rely on integrating existing sources [25], using coding systems for entity alignment [26]. However, these methods struggle with terminological variations, leading to redundancy and inconsistency. A few studies have attempted to construct general-purpose medical KGs through integrating existing ones [25, 27], but they heavily rely on existing coding systems and thesauruses [26] for entity alignment across KGs, which often fail in front of varying terminologies such as due to different conventions or abbreviations, leading to high degrees of duplication and inconsistency. Recently, several studies started to explore the potential of LLMs to automate KG construction [20, 28]. In the following, we discuss key applications of LLMs and multi-modal foundation models (MMFMs) toward constructing high-quality KGs using multi-modality biomedical data from existing KGs, biomedical literature and medical institutions.

2.1 Integrating existing KGs

KG integration, often referred to as knowledge fusion or alignment, involves merging KGs from diverse sources and formats. It is a core challenge in knowledge engineering, requiring effective strategies to ensure consistency and interoperability [29]. 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 [29]. 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 [27]. 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; (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 neural approaches have been proposed for entity alignment on KGs, 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. Second, LLMs can draw on their extensive knowledge acquired during pre-training to aid in disambiguating entities and mapping relationships across different KGs. Third, LLMs possess robust few-shot learning abilities, making them particularly valuable for specialized domain applications where labeled data are limited.

Prior works [30, 31] explore the potential of linking biomedical entities across KGs. Specifically, HiPrompt [30] aligns entities between biomedical KGs and standardized ontologies via a two-stage approach: traditional information retrieval techniques (BM25) followed by a LLM-based re-ranking using hierarchy-oriented prompts. PromptLink [31] further improves this two-stage framework 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. The improvements are especially significant for weaker LLMs, which is intuitive and useful since not every medical institution can always (safely) access the strongest LLMs. Besides, AutoAlign [32] utilizes off-the-shelf LLMs to capture relationships between entity types with a predicate-proximity graph and then aligns entities across KGs by computing similarity in the embeddings space.

The above advances in LLM-aided KG integration suggest several promising future directions. For example, future LLM-driven KG integration could focus on enabling *evolving knowledge graph updates* by resolving inconsistencies between new and existing knowledge. Another key challenge in leveraging LLMs to enrich KGs is the risk for misinformation [33]. To alleviate this, human-in-the-loop frameworks play a critical role in verifying and refining LLM-generated outputs [34].

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.

The key advantage of LLMs in this domain lies in their ability to generate novel, semantically coherent information that can supplement and enrich existing KGs. Unlike rule-based or supervised machine learning approaches, LLMs can leverage their extensive understanding of language and the world to infer missing connections, identify new entities, and uncover implicit relationships - all without being constrained by the limitations of manually curated training data.

As demonstrated in recent works, LLM-based approaches have shown strong potential in KG construction and completion. Zhu et al. [20] exploit in-context learning to predict missing entities and relations, generating new triplets to expand existing KGs. Meanwhile, KC-GenRE [35] frames KG completion as a retrieval and ranking problem. Their approaches first retrieve candidate entities and then employ LLMs to refine and reorder them, ultimately forming additional knowledge triplets. Zhang et al. [36] further incorporate the schema elements relevant to the prompt for standardizing the triplets and improving the generation quality of LLMs. Inspired by recent progress of LLM reasoning, Nie et al. [37] further chain-of-thought promoting techniques [38] to better guiding LLMs in understanding triple knowledge in unstructured data with improved triplet extraction accuracy.

Beyond traditional text-based prompting, code-based instructions offer an alternative approach for guiding LLMs in structured knowledge generation. Code LLMs, designed for processing structured data like programming code, naturally align with the hierarchical and relational nature of KGs. Their training on structured inputs enables them to better capture and manipulate graph-based representations, making them well-suited for tasks requiring logical consistency and precision [39, 40]. For ontology expansion task, CodeTaxo [41] represents entity relationships with a hierarchical structure inspired by programming syntax for better harnessing LLMs' ability to interpret structured patterns. By leveraging code-like representations, it enables LLMs to systematically construct taxonomies, improving the organization and completeness of KGs. CodeKGC [42] encoded the schema of KGs by modeling code definitions for capture the structural information inherent in the data. By leveraging chain-of-thought prompting, it systematically generates precise knowledge triples. To summarize, code-driven prompting enables LLMs to structure and categorize concepts more efficiently by leveraging syntax-based reasoning, improving KG organization and consistency. This method outperforms conventional natural language prompts by providing a more structured and logic-driven way to enhance KG completion.

Apart from prompting, several studies explored fine-tuning to adapt LLMs for KG completion. KG-LLM [43] directly performs instruction tuning on KG completion tasks including *triplet classification*, *relation prediction* and *link prediction* and outperform frontier models using lightweight backbones only. KOPA [44] first conducts pretraining to obtain entity and relation embeddings, which transforms them into virtual knowledge tokens within a unified textual space. These tokens then act as prefixes in LLM prompts that enable structure-aware reasoning by combining LLM generation with KG-based retrieval. MKGL [45] structures knowledge as three-word sentences (entity-relation-entity triplets), and then finetunes LLMs to generate and complete KG triplets by leveraging real-time KG context retrieval and token embedding augmentation to enhance factual consistency. KG-FIT [46] attempts to exploit open-world knowledge from LLMs to enhance KG embeddings. It first constructs a semantically coherent, hierarchical structure of entity clusters, then fine-tunes these embeddings by integrating the hierarchical structure within textual embeddings. Such a hybrid approach combines structural KG information with semantic depth from LLMs for richer representations.

2.3 Enriching KGs with multi-modality data

Another property of biomedical KGs is the diverse sources and modalities of useful knowledge. Besides existing KGs and online literature, medical institutions including hospitals, clinics and medical centers generate vast amounts of patient data daily, through patient visits, clinical trials, health monitoring devices, and so on. Traditionally, specialized models and algorithms have been developed to process and analyze various modalities of patient data such as electronic health records (EHRs), medical images, physiological waveforms, clinical notes and health insurance claims. These methods can hardly perform integrative analysis across data modalities and generalize across different medical institutions. Recently, LLM-based multimodality foundation models (MMFMs) have shown strong promises in analyzing multi-modality data through the unified interface of languages [47]. However, studies on aligning general MMFMs to real patient data have shown this task to be rather challenging due to the lack of high-quality fine-grained pairs of *X-and-text* labeled data for instruction tuning– let X be *chest x-rays* and *text* be *radiology reports*, most reports do not include accurate annotations (bounding boxes) of lesions in the lung [48]; let X be *spatiotemporal EEG/SEEG signals* and *text* be *nursing notes*, most notes do not indicate accurate time ranges and spatial locations of epilepsies in the brain [49]. Besides the development of biomedical MMFMs, how to utilize them for the extraction of novel knowledge from multi-modality patient data that can be properly incorporated into the biomedical KGs for further integrative utilization towards empowering LLMs and facilitating various downstream biomedical applications can highly impact AI practices in health informatics, but this remains much under-explored.

To adapt MMFMs with domain-specific knowledges, BioMed-VITAL [48] is designed for efficiently aligning biomedical MMFMs with clinician preferences. Through novel designs in the three steps of *data generation*, *selection* and *instruction tuning*, BioMed-VITAL demonstrates significant improvements in medical visual chat and VQA datasets, showing strong promises for further utilization towards open-ended knowledge discovery from medical images. On the other hand, OpenVik [50] explores prompting MMFMs to generate format-free visual knowledge from the detected regions. OpenVik further gathered novel relational knowledge generated from images from Visual Genome [51] to form a novel KG, and demonstrated its advantages over existing commonsense KGs such as ConceptNet [52] and COMET [53] for various KG-supported downstream tasks. It opens up the new arena of utilizing MMFMs for automatic relational knowledge discovery. PromptCap [54] leverages pre-trained multimodal large language models (MLLMs) to understand both visual and textual information in a unified framework for improved visual relation extraction. Yang et al. [55] introduce an automated pipeline for building product KGs in e-commerce directly from raw images. It leverages MMFMs to extract visual details and utilizes an LLM to infer missing attributes and relationships. By hierarchically structuring and linking entities, this pipeline enables scalable KG construction without requiring manual intervention.

3 KG-guided LLM Enhancement

LLMs have shown impressive communication and question answering capabilities, demonstrating strong promises in various healthcare applications [2, 56]. However, to reliably model biomedical data and generate factual and accurate answers, LLMs still face the challenges of lacking domain knowledge, fuzzy inferences, and hallucination [8]. Retrieval augmented generation (RAG) [57], which aims at retrieving question-relevant evidences and generating evidence-based answers, has strong promises in evidence-critical domains including biomedicine and healthcare. However, effective and efficient RAG for biomedicine and healthcare is seldom studied and requires solutions to practical challenges including (1) how to find relevant evidences from complex biomedical data, (2) how to conduct biomedically valid complex inference, and (3) how to reliably guarantee the removal of biomedical errors. In this section, we will comprehensively investigate these challenges and demonstrate the advantages of utilizing KGs to enable reliable LLMs for biomedicine and healthcare.

3.1 Planning with domain knowledge

Modeling biomedical data is challenging because diseases are often complicated and heterogeneous such as regarding the causes, symptoms, and treatment effects. Given a question seemingly as simple as *'Will patient 0315 with type 2 diabetes develop cardiovascular comorbidities in the next 5 years'*, an experienced doctor will naturally expand the question into a set of related questions– *e.g.*, *'Does the patient suffer from obesity'*, *'Does the patient take metformin'*, *'Does the patient smoke'*, *'What is the patient's current and historical blood pressure'*, *etc.*, so as to find and retrieve necessary evidences for answering the original question. This process is referred to as query planning in LLM research [58], and LLMs are known to perform poorly and hallucinate when planning for complex queries even with explicit and deliberately designed prompts [59]. Moreover, LLMs also lack concrete knowledge about biomedicine and healthcare to generate faithful retrieval plans– *e.g.*, for calling tools to retrieve existing patient data, asking the patient through a chatting interface, or ordering new tests for the patient, especially facing the complexity of diseases and biomedical data.

Early works have primarily relied on external retrievers to acquire relevant factual knowledge to enhance LLM reasoning. For instance, Baek et al. [60] proposed a direct retrieval method to extract pertinent triples from KGs. However, the retriever, being a shallow embedding model, may not consistently retrieve the most relevant facts, particularly in the complex medical domain. KGs encompass a wealth of domain-specific knowledge, posing challenges for LLMs with limited domain expertise in comprehending and utilizing this information for answering medical questions. To further harness the potential of LLMs in leveraging domain knowledge, the *plan-and-solve* paradigm [61] has been introduced, where LLMs are first prompted to generate a plan. Based on the generated plan, LLMs can retrieve the relevant domain knowledge and perform reasoning to generate answers [62]. However, existing approaches fall short in handling the complex structured knowledge within KGs to enable effective planning and reasoning. To address this limitation, Luo et al. [63] proposed a *planning-retrieval-reasoning* framework, RoG, which empowers LLMs to plan and reason over KGs . The overall framework is depicted in Figure 2.

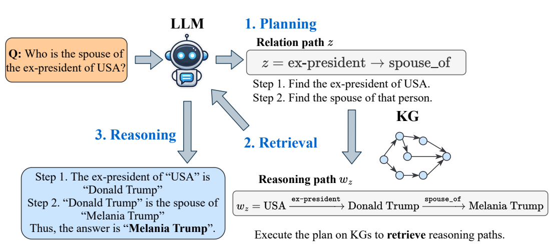

Figure 2: The overall framework of planning and reasoning on KGs (RoG).

RoG first generates multiple relation paths, grounded in knowledge graphs (KGs), which are used as structured plans. Relation paths, which represent semantic relationships between entities, have been widely applied in various reasoning tasks on KGs [64] by decomposing a complex reasoning into multiple simple steps. Leveraging these relation paths, one can efficiently retrieve up-to-date knowledge from KGs through a constrained breadth-first search. Consequently, relation paths act as reliable plans that guide both the retrieval and reasoning processes on domain-specific KGs. Furthermore, by treating relation paths as plans, one ensures that these plans are firmly grounded in KGs, enabling large language models (LLMs) to retrieve pertinent knowledge and perform accurate reasoning. To formalize this approach, RoG is framed as an optimization problem aimed at maximizing the probability of deriving an answer from a KG $\mathcal{G}$ with respect to a given question q, by generating relation paths z as the guiding plan:

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

where θ denotes the parameters of LLMs and a denotes the final answer. To enable accurate planning with domain knowledge, two instruction tuning tasks are designed: 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

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

where the shortest paths $\mathcal{Z}^* \subseteq \mathcal{Z}$ between q and a in KGs are used as supervision signals. The probability of LLMs generating faithful relation paths is maximized 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

After retrieving evidences, LLMs need to follow them and generate answers. While evidence following has been shown achievable through prompt designs [65, 66], LLMs will struggle to generate the correct answers when important supporting evidences are missing, which is very common in biomedicine and healthcare due to the sparse and incomplete patient data especially regarding rare diseases. With strong prior knowledge of LLMs and available partial evidences, it is promising to infer missing evidences with LLMs between the retrieval and generation steps, which ideally requires the LLMs to be able to conduct accurate and efficient biomedically valid reasoning. Recently, extensive research has been conducted on enabling LLMs for knowledge-based reasoning, which can be roughly divided into embedding-based (neural) approaches [67] and rule-based (symbolic) approaches [68]. However, the neural approaches loose interpretability and can hardly work with sparse KGs, while the symbolic approaches are resource-consuming due to large amounts of paths that do not encode valid logic rules. LLMs are neural models that can also naturally work with embeddings, making it an ideal backbone for combining neural and symbolic approaches towards accurate and efficient reasoning, which, however, is hardly studied.

Recent efforts have focused on enabling LLMs to perform reasoning on structured KGs through retrieval-based methods and prompting techniques [68]. CoK [69] and KD-CoT [70] retrieve facts from external KGs to guide the chain-of-thought (CoT) reasoning process conducted by LLMs. To capture graph structures, GNN-RAG [71] utilizes a lightweight graph neural network to efficiently retrieve knowledge from KGs, formatted as sentence paths to stimulate the reasoning process in LLMs. Mindmap [72] introduces a prompt-based approach that equips LLMs with the ability to comprehend and reason over KGs. Despite the success of these methods, challenges persist in the development of principled prompts for KG representation and reasoning. Furthermore, LLMs continue to face limitations in their understanding of graph structures and reasoning with text-based graph prompts [73].

Unlike previous approaches that necessitate a computationally intensive fine-tuning phase or the design of ad-hoc prompts for LLMs, Luo et al. [74] recently proposed the KG-constrained reasoning (GCR) paradigm. GCR integrates unstructured reasoning in LLMs with structured knowledge in knowledge graphs (KGs), aiming to enable efficient and effective reasoning over structured knowledge. The overall framework is depicted in Figure 3.

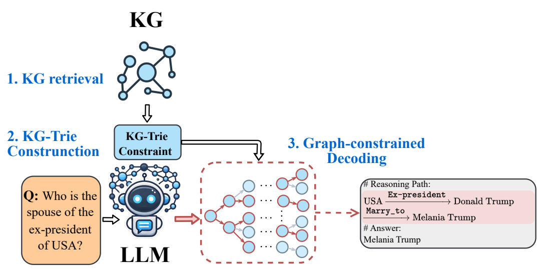

Figure 3: The overall framework of KG-constrained reasoning (GCR).

Graph-constrained reasoning, inspired by the concept that LLMs reason through decoding [38], 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, a retrieval module is first adopted to find a relevant KG that is helpful for reasoning. Then, the KG is converted into a structured index, KG-Trie, to facilitate efficient reasoning on KG using LLMs. Trie is also known as the prefix tree [75] 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, graph-constrained decoding is proposed that employs a lightweight KG-specialized LLM to generate multiple KG-grounded reasoning paths and answers. With the constraints from KG-Trie, one ensures faithful reasoning while leveraging the strong reasoning capabilities of LLMs to efficiently explore paths on 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

Biomedicine and healthcare have high requirements on factuality and accuracy. While the previous subtasks can improve factuality via planning and accuracy via reasoning, errors can still occur due to LLM's failure to follow evidences, misunderstanding and hallucination. Therefore, we propose to utilize KGs to add post-hoc knowledge grounding, further ensuring the LLM reliability. While existing works on KG-based RAG mostly use KGs as additional resources of factual knowledge [76], post-hoc error detection will be conducted on the LLM outputs. This is fundamentally different and more challenging, as the outputs often blend facts from diverse sources and involve multiple reasoning steps.

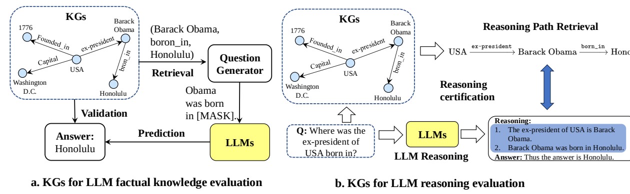

Figure 4: The illustration of LLM reflection with KGs. (a) The evaluation of the factual knowledge inside LLMs. (b) The evaluation of the reaosning process of LLMs with KGs.

The hallucination phenomenon in LLMs is commonly attributed to their limited factual knowledge. To systematically evaluate the factual knowledge embedded in LLMs and enable faithful respond in solving medical problem, we propose a novel framework, as illustrated in Figure 4a, which leverages Knowledge Graphs (KGs) [77]. In contrast to traditional approaches that rely on human-annotated question-answer datasets, our method generates valid and diverse questions from KGs at varying levels of difficulty while ensuring comprehensive knowledge coverage. Specifically, we extract atomic knowledge from KGs in the form of sets of triples. These triples are then converted into question-answer pairs using various question generation techniques, such as template-based and LLM-based methods. The generated question-answer pairs are subsequently used to assess the factual knowledge of LLMs by comparing the model-generated answers with the corresponding ground-truth answers. The evaluation outcomes provide insights into the factual accuracy of LLMs and can be used to better understand their hallucination behaviors. This approach facilitates a systematic evaluation of LLMs' factual knowledge and offers valuable guidance for enhancing their reliability across a range of high-stakes applications.

In addition to factual knowledge, the structure of knowledge graphs (KGs) can be leveraged to justify the reasoning process of large language models (LLMs). Minh-Vuong et al. [78] developed a framework that explores the chain-of-thought (CoT) reasoning capabilities of LLMs in multi-hop question answering by utilizing KGs, as illustrated in Figure 4b. The framework includes two evaluation modules: discriminative evaluation and generative evaluation. The discriminative evaluation assesses whether LLMs possess sufficient knowledge to conduct faithful reasoning. It inputs both valid and invalid reasoning paths, retrieved from KGs, into LLMs and requests them to predict the validity of these paths. Conversely, the generative evaluation aims to assess the faithfulness of the LLMs' reasoning process by grounding it in KGs. For a reasoning process generated by LLMs, the generative evaluation module retrieves relevant facts from KGs and compares them with the ground-truth reasoning paths. The evaluation results serve to reflect the reasoning capabilities of LLMs and offer insights into the faithfulness of their reasoning. Despite demonstrating impressive reasoning abilities, LLMs face ongoing challenges in ensuring faithful reasoning, particularly in multi-hop question answering.

4 Applications in Biomedicine and Healthcare

Biomedicine and healthcare represent ideal domains for the integration of KGs and LLMs due to their critical need for both structured knowledge and natural language understanding. The motivation for applying MedKLM in these domains stems from the complexity of biomedical data, the high stakes of healthcare decisions, and the necessity for interpretable AI systems that clinicians can trust. Healthcare applications face unique challenges including data fragmentation across institutions, strict requirements for accuracy and reliability, privacy concerns, and the need to incorporate domain-specific expertise. Despite these challenges, `BioMedKLM` offers promising directions for transforming healthcare through: (1) enhancing clinical decision support with reasoning grounded in biomedical knowledge; (2) accelerating drug discovery and development by connecting molecular structures with biomedical insights; and (3) improving biomedical knowledge management to organize, integrate, and apply the vast amount of healthcare information available. The following subsections explore these specific applications where the synergy between KGs and LLMs has shown particular promise in addressing complex healthcare needs.

4.1 `BioMedKLM` for clinical decision support

Clinical decision support represents a key application domain for `BioMedKLM`, where the synergy between KGs and LLMs enables more accurate and interpretable clinical decisions. By combining structured biomedical knowledge with advanced language understanding capabilities, these systems help clinicians analyze complex patient data and make informed medical decisions. We categorize existing approaches into two categories:

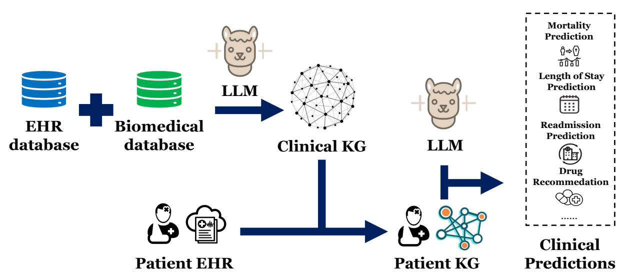

Figure 5: An example application of `BioMedKLM` for clinical decision support.

(1) Integrated KG+LLM approaches. Recent developments demonstrate how the integration of KGs and LLMs can enhance clinical tasks through complementary strengths. GraphCare [79] uses LLMs to construct patient-specific KGs, enabling more personalized and accurate clinical predictions. RAM-EHR [80] integrates a hypergraph that captures complex biomedical concept relationships and LLM-summarized medical context to improve clinical predictions. EMERGE [81] demonstrates effective fusion of KGs with clinical notes for enhanced prediction accuracy. MedIKAL [24] reimagines the biomedical KG as an assistant that guides LLM decision-making, whereas KARE [82] introduces a novel framework combining LLM's reasoning capability and precise biomedical knowledge retrieval with its constructed comprehensive biomedical KG community summaries. ComLLM [83] introduces a method prompting LLMs with a disease-specific KG for disease progression prediction, while MedTok [84] introduces multi-modal (KG and LLM) medical code learning to enhance clinical predictions. MedRAG [85] combines a four-tier hierarchical diagnostic KG with RAG to enable accurate diagnosis of diseases with similar manifestations, while also providing personalized treatment recommendations and proactive follow-up questions. We show an example of using `BioMedKLM` for clinical predictions in Figure 5.

(2) KG/Graph-only and LLM-only foundations. Earlier approaches focused on either structured knowledge or language understanding independently. Graph-based methods [86] established the fundamental importance of structured biomedical knowledge representation for clinical predictions. These methods could be significantly enhanced by incorporating LLMs for automated graph construction and natural language explanation generation. Conversely, pure LLM approaches [87, 88] and domain-specific language models [3] demonstrated remarkable natural language understanding capabilities in the medical domain, but their potential could be further realized through integration with KGs for improved factual accuracy and reasoning capabilities.

4.2 `BioMedKLM` for drug discovery and development

Drug discovery and development represents another key application domain for `BioMedKLM`, where the synergy between KGs and LLMs enables more accurate and efficient drug design and evaluation.

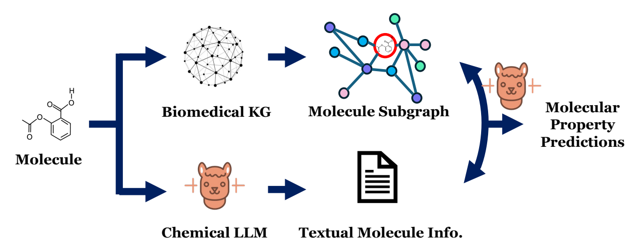

Figure 6: An example application of `BioMedKLM` on molecular property prediction tasks for drug discovery.

(1) Drug discovery. Molecules inherently possess graph structures, making them natural candidates for graph-based representation learning. Recent work has explored various approaches to enhance molecular representations [89], with a particularly promising direction being the integration of knowledge graphs [90]. For instance, Ye et al. [91] pioneered the incorporation of static KG embeddings into molecular fingerprints to improve drug-target interaction prediction. KANO [92] constructed Element-KG that augments molecular representations at the chemical element level through contrastive learning. Taking a different approach, Gode [90] developed a comprehensive biomedical KG that treats molecules as graph-in-graph structures, enabling deeper integration of biomedical knowledge into molecular representations. The field has recently seen exciting developments in leveraging LLMs' reasoning capabilities to enhance molecular property predictions [93], opening new opportunities for BioMedKLM to synergistically combine KGs and LLMs in drug discovery applications. Figure 6 illustrates an example of applying BioMedKLM to drug discovery, where we generate molecular representation by integrating structured knowledge from biomedical KGs [90] with textual understanding from chemical LLMs [94]. This knowledge-enriched representation is then learned by LLM to make accurate and interpretable property prediction for highly reliable drug discovery.

(2) Drug development. Drug development requires deep understanding of biological mechanisms, disease pathways, and drug effects. Recent advances leverage both KGs and language models to accelerate this process. KGWAS [95] and TxGNN [96] exemplified this trend as graph foundation models: KGWAS used functional genomics KGs to detect disease-associated genetic variants in rare diseases, while TxGNN learned from biomedical KGs for drug repurposing. These systems could be enhanced by using LLMs to enrich their knowledge graphs with literature-derived relationships and translate graph reasoning into natural language explanations. Malas et al. [97] demonstrated KG-based drug prioritization by extracting semantic features between drugs and diseases to train a classifier that successfully identified promising candidates for polycystic kidney disease. In clinical trials, LINT [98] combined medical code KGs with LLM embeddings to predict trial outcomes. While TrialGPT [99] demonstrated strong performance using LLMs alone for trial matching, its retrieval-matching-ranking pipeline presents promising opportunities for incorporating KGs to further enhance matching accuracy and interpretability in drug evaluation studies.

4.3 BioMedKLM for biomedical knowledge management

Biomedical knowledge management is a critical aspect of modern healthcare, requiring systematic organization, curation, and integration of information from diverse sources such as clinical guidelines, research literature, and patient data. The combination of LLMs and KGs offers a powerful framework for automating knowledge extraction, integration, and maintenance while ensuring accuracy, consistency, and interpretability. This section explores the application of BioMedKLM for biomedical knowledge management.

Biomedical Knowledge Graph Construction and Updates. Construction and maintenance of biomedical KGs are fundamental to many healthcare applications. BioMedKLM facilitates automatic construction and continuous updating of KGs through several approaches: Graphusion [100] utilizes LLMs for KG construction with a global perspective, employing retrieval-augmented generation (RAG) to extract and integrate knowledge from diverse sources. KG-RAG [101] optimizes prompt generation for LLMs to enhance accuracy and relevance of extracted biomedical knowledge. Domain-specific KGs have been developed using LLMs, such as KGs for diabetes [102], traditional Chinese medicine [103], heart failure [104], rare disease [105], and Alzheimer's disease research [106].

Data Generation and Knowledge Application. BioMedKLM extends beyond knowledge organization to practical applications: (1) ClinGen [107] leverages LLMs to generate synthetic clinical text data while ensuring accuracy and alignment with real-world scenarios. (2) Clinical decision support systems integrate KGs and

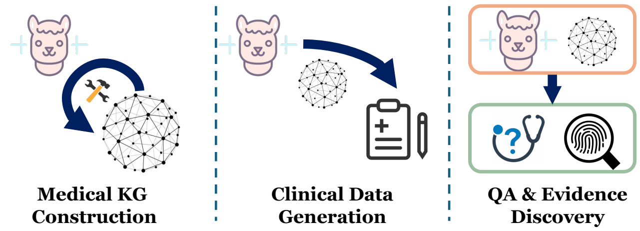

Figure 7: An example application of BioMedKLM for biomedical knowledge management.

LLMs to enhance prediction accuracy and interpretability [82, 108]. (3) Risk prediction models utilizing KGs have been developed for various conditions such as macular edema [109], adverse drug reactions [110], and cancer [111].

Question Answering and Evidence Discovery. BioMedKLM plays a crucial role in medical question answering and knowledge validation: (1) Studies [112, 113] demonstrate effective QA systems using LLM-KG combinations, including specialized systems for hepatitis B [114] and hypertension [115]. KGARevion [116] combines LLMs with KGs for medical QA through an iterative generate-verify-refine pipeline, improving accuracy over existing methods by dynamically refining generated answers using structured knowledge from KGs. This integration enables extraction of complex relationships and evidence from biomedical literature. (2) The framework shows potential in detecting medical fraud and waste, analyzing patterns in healthcare data to identify potential cases of fraud [117, 118]. BioMedKLM also leverages LLMs and KGs to automate the process of literature synthesis [119], enabling researchers and clinicians to efficiently identify and integrate key findings from vast amounts of text.

This integration of LLMs and KGs in biomedical knowledge management demonstrates significant potential for improving healthcare information organization, accessibility, and application, while maintaining high standards of accuracy and reliability.

5 Conclusions and Future Directions

In this paper, we discuss the trending efforts of knowledge and language co-modeling. 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. We also provide a detailed review of knowledge and language co-modeling for real-world applications in biomedicine and healthcare. While this nascent research direction holds great potential, we envision several specific problems worthy of further exploration including the development of more powerful and generic KGs, the trade-off between unified and specialized KGs, the evaluation of knowledge in LLMs, and the resolution of conflicts between KGs and LLMs for biomedicine and healthcare.

References

[1] Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, Carol Chen, Catherine Olsson, Christopher Olah, Danny Hernandez, Dawn Drain, Deep Ganguli, Dustin Li, Eli Tran-Johnson, Ethan Perez, Jamie Kerr, Jared Mueller, Jeffrey Ladish, Joshua Landau, Kamal Ndousse, Kamile Lukosuite, Liane

Lovitt, Michael Sellitto, Nelson Elhage, Nicholas Schiefer, Noemi Mercado, Nova DasSarma, Robert Lasenby, Robin Larson, Sam Ringer, Scott Johnston, Shauna Kravec, Sheer El Showk, Stanislav Fort, Tamera Lanham, Timothy Telleen-Lawton, Tom Conerly, Tom Henighan, Tristan Hume, Samuel R Bowman, Zac Hatfield-Dodds, Ben Mann, Dario Amodei, Nicholas Joseph, Sam McCandlish, Tom Brown, and Jared Kaplan. Constitutional ai: Harmlessness from ai feedback. arXiv preprint arXiv:2212.08073, 2022.

[2] Karan Singhal, Shekoofeh Azizi, Tao Tu, S Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, Perry Payne, Martin Seneviratne, Paul Gamble, Chris Kelly, Nathaneal Scharli, Aakanksha Chowdhery, Philip Mansfield, Blaise Aguera y Arcas, Dale Webster, Greg S Corrado, Yossi Matias, Katherine Chou, Juraj Gottweis, Nenad Tomasev, Yun Liu, Alvin Rajkomar, Joelle Barral, Christopher Semturs, Alan Karthikesalingam, and Vivek Natarajan. Large language models encode clinical knowledge. Nature, 620(7972):172–180, 2023.

[3] Xi Yang, Aokun Chen, Nima PourNejatian, Hoo Chang Shin, Kaleb E Smith, Christopher Parisien, Colin Compas, Cheryl Martin, Anthony B Costa, Mona G Flores, Ying Zhang, Tanja Magoc, Christopher A Harle, Gloria Lipori, Duane A Mitchell, William R Hogan, Elizabeth A Shenkman, Jiang Bian, and Yonghui Wu. A large language model for electronic health records. NPJ Digital Medicine, 5(1):194, 2022.

[4] Nikita Mehandru, Brenda Y Miao, Eduardo Rodriguez Almaraz, Madhumita Sushil, Atul J Butte, and Ahmed Alaa. Evaluating large language models as agents in the clinic. NPJ Digital Medicine, 7(1):84, 2024.

[5] Xuming Hu, Junzhe Chen, Xiaochuan Li, Yufei Guo, Lijie Wen, Philip S Yu, and Zhijiang Guo. Do large language models know about facts? arXiv preprint arXiv:2310.05177, 2023.

[6] Hanmeng Liu, Ruoxi Ning, Zhiyang Teng, Jian Liu, Qiji Zhou, and Yue Zhang. Evaluating the logical reasoning ability of chatgpt and gpt-4. arXiv preprint arXiv:2304.03439, 2023.

[7] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12):1–38, 2023.

[8] Michael Wornow, Yizhe Xu, Rahul Thapa, Birju Patel, Ethan Steinberg, Scott Fleming, Michael A Pfeffer, Jason Fries, and Nigam H Shah. The shaky foundations of large language models and foundation models for electronic health records. NPJ Digital Medicine, 6(1):135, 2023.

[9] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia D'amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, and Sebastian Neumaier. Knowledge graphs. ACM Computing Surveys, 54(4):1–37, 2021.

[10] Hejie Cui, Jiaying Lu, Shiyu Wang, Ran Xu, Wenjing Ma, Shaojun Yu, Yue Yu, Xuan Kan, Tianfan Fu, Chen Ling, Joyce Ho, Fei Wang, and Carl 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.

[11] Fan Feng, Feitong Tang, Yijia Gao, Dongyu Zhu, Tianjun Li, Shuyuan Yang, Yuan Yao, Yuanhao Huang, and Jie Liu. Genomickb: a knowledge graph for the human genome. Nucleic Acids Research, 51(D1):D950–D956, 2023.

[12] Finlay MacLean. Knowledge graphs and their applications in drug discovery. Expert Opinion on Drug Discovery, 16(9):1057–1069, 2021.

[13] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F Stewart, and Jimeng Sun. Gram: graph-based attention model for healthcare representation learning. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pages 787–795, 2017.

[14] Zhihan Gao, Min Gao, Chun-hua Chen, Yifan Zhou, Zhi-Hui Zhan, and Yuan Ren. Knowledge graph of wastewater-based epidemiology development: A data-driven analysis based on research topics and trends. Environmental Science and Pollution Research, 30(11):28373–28382, 2023.

[15] Chuan Hong, Everett Rush, Molei Liu, Doudou Zhou, Jiehuan Sun, Aaron Sonabend, Victor M Castro, Petra Schubert, Vidul A Panickan, Tianrun Cai, Lauren Costa, Zeling He, Nicholas Link, Ronald Hauser, J Michael Gaziano, Shawn N Murphy, George Ostrouchov, Yuk-Lam Ho, Edmon Begoli, Junwei Lu, Kelly Cho, Katherine P Liao, and Tianxi Cai. Clinical knowledge extraction via sparse embedding regression (keser) with multi-center large scale electronic health record data. NPJ Digital Medicine, 4(1):151, 2021.

[16] Julián N Acosta, Guido J Falcone, Pranav Rajpurkar, and Eric J Topol. Multimodal biomedical ai. Nature Medicine, 28(9):1773–1784, 2022.

[17] Jeff Pan, Simon Razniewski, Jan-Christoph Kalo, Sneha Singhania, Jiaoyan Chen, Stefan Dietze, Hajira Jabeen, Janna Omeliyanenko, Wen Zhang, Matteo Lissandrini, Russa Biswas, Gerard de Melo, Angela Bonifati, Edlira Vakaj, Mauro Dragoni, and Damien Graux. Large language models and knowledge graphs: Opportunities and challenges. Transactions on Graph Data and Knowledge, 2023.

[18] Xintao Wang, Qianyu He, Jiaqing Liang, and Yanghua Xiao. Language models as knowledge embeddings. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 2022.

[19] Liang Yao, Chengsheng Mao, and Yuan Luo. Kg-bert: Bert for knowledge graph completion. arXiv preprint arXiv:1909.03193, 2019.

[20] Yuqi Zhu, Xiaohan Wang, Jing Chen, Shuofei Qiao, Yixin Ou, Yunzhi Yao, Shumin Deng, Huajun Chen, and Ningyu Zhang. Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities. arXiv preprint arXiv:2305.13168, 2023.

[21] Linmei Hu, Zeyi Liu, Ziwang Zhao, Lei Hou, Liqiang Nie, and Juanzi Li. A survey of knowledge enhanced pre-trained language models. IEEE Transactions on Knowledge and Data Engineering (TKDE), 36:1413–1430, 2023.

[22] Shiyang Li, Yifan Gao, Haoming Jiang, Qingyu Yin, Zheng Li, Xifeng Yan, Chao Zhang, and Bing Yin. Graph reasoning for question answering with triplet retrieval. In Findings of the Association for Computational Linguistics: ACL (ACL-Findings), 2023.

[23] Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller. Language models as knowledge bases? In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019.

[24] Mingyi Jia, Junwen Duan, Yan Song, and Jianxin Wang. medikal: Integrating knowledge graphs as assistants of llms for enhanced clinical diagnosis on emrs. arXiv preprint arXiv:2406.14326, 2024.

[25] Su Chang, Yu Hou, Suraj Rajendran, Jacqueline R M A Maasch, Zehra Abedi, Haotan Zhang, Zilong Bai, Anthony Cuturrufo, Winston Guo, Fayzan F Chaudhry, Gregory Ghahramani, Jian Tang, Feixiong Cheng, Yue Li, Rui Zhang, Jiang Bian, and Fei Wang. Biomedical discovery through the integrative biomedical knowledge hub (ibkh). iScience, 26(4):106460, 2023.

[26] James E Harrison, Stefanie Weber, Robert Jakob, and Christopher G Chute. Icd-11: an international classification of diseases for the twenty-first century. BMC Medical Informatics and Decision Making, 21:1–10, 2021.

[27] Alberto Santos, Ana R Colaço, Annelaura B Nielsen, Lili Niu, Philipp E Geyer, Fabian Coscia, Nicolai J Wewer Albrechtsen, Filip Mundt, Lars Juhl Jensen, and Matthias Mann. Clinical knowledge graph integrates proteomics data into clinical decision-making. bioRxiv, pages 2020–05, 2020.

[28] Hongbin Ye, Ningyu Zhang, Hui Chen, and Huajun Chen. Generative knowledge graph construction: A review. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1–17, 2022.

[29] Youfu Yan, Yu Hou, Yongkang Xiao, Rui Zhang, and Qianwen Wang. Knownet: Guided health information seeking from llms via knowledge graph integration. IEEE Transactions on Visualization and Computer Graphics, 2024.

[30] Jiaying Lu, Jiaming Shen, Bo Xiong, Wenjing Ma, Steffen Staab, and Carl 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.

[31] Yuzhang Xie, Jiaying Lu, Joyce Ho, Fadi Nahab, Xiao Hu, and Carl 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, pages 2589–2593, 2024.

[32] Rui Zhang, Yixin Su, Bayu Distiawan Trisedya, Xiaoyan Zhao, Min Yang, Hong Cheng, and Jianzhong Qi. Autoalign: fully automatic and effective knowledge graph alignment enabled by large language models. IEEE Transactions on Knowledge and Data Engineering, 2023.

[33] Linyao Yang, Hongyang Chen, Zhao Li, Xiao Ding, and Xindong 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.

[34] Ran Xu, Hui Liu, Sreyashi Nag, Zhenwei Dai, Yaochen Xie, Xianfeng Tang, Chen Luo, Yang Li, Joyce C Ho, Carl Yang, and Qi He. Simrag: Self-improving retrieval-augmented generation for adapting large language models to specialized domains. arXiv preprint arXiv:2410.17952, 2024.

[35] Yilin Wang, Minghao Hu, Zhen Huang, Dongsheng Li, Dong Yang, and Xicheng 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, pages 9668–9680, 2024.

[36] Bowen Zhang and Harold 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), pages 9820–9836, 2024.

[37] Jixuan Nie, Xia Hou, Wenfeng Song, Xuan Wang, Xinyu Zhang, Xingliang Jin, Shuozhe Zhang, and Jiaqi Shi. Knowledge graph efficient construction: Embedding chain-of-thought into llms. Proceedings of the VLDB Endowment. ISSN, 2150:8097, 2024.

[38] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2022.

[39] Wenhu Chen, Xueguang Ma, Xinyi Wang, and William W. Cohen. Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks. Transactions on Machine Learning Research, 2023.

[40] Wenqi Shi, Ran Xu, Yuchen Zhuang, Yue Yu, Jieyu Zhang, Hang Wu, Yuanda Zhu, Joyce Ho, Carl Yang, and May Dongmei Wang. Ehragent: Code empowers large language models for few-shot complex tabular reasoning on electronic health records. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pages 22315–22339, 2024.

[41] Qingkai Zeng, Yuyang Bai, Zhaoxuan Tan, Zhenyu Wu, Shangbin Feng, and Meng Jiang. Codetaxo: Enhancing taxonomy expansion with limited examples via code language prompts. arXiv preprint arXiv:2408.09070, 2024.

[42] Zhen Bi, Jing Chen, Yinuo Jiang, Feiyu Xiong, Wei Guo, Huajun Chen, and Ningyu Zhang. Codekgc: Code language model for generative knowledge graph construction. ACM Transactions on Asian and Low-Resource Language Information Processing, 23(3):1–16, 2024.

[43] Liang Yao, Jiazhen Peng, Chengsheng Mao, and Yuan Luo. Exploring large language models for knowledge graph completion. arXiv preprint arXiv:2308.13916, 2023.

[44] Yichi Zhang, Zhuo Chen, Lingbing Guo, Yajing Xu, Wen Zhang, and Huajun Chen. Making large language models perform better in knowledge graph completion. In Proceedings of the 32nd ACM International Conference on Multimedia, page 233–242. Association for Computing Machinery, 2024.

[45] Lingbing Guo, Zhongpu Bo, Zhuo Chen, Yichi Zhang, Jiaoyan Chen, Lan Yarong, Mengshu Sun, Zhiqiang Zhang, Yangyifei Luo, Qian Li, Qiang Zhang, Wen Zhang, and Huajun Chen. MKGL: Mastery of a three-word language. In The Thirty-eighth Annual Conference on Neural Information Processing Systems, 2024.

[46] Pengcheng Jiang, Lang Cao, Cao Xiao, Parminder Bhatia, Jimeng Sun, and Jiawei Han. KG-FIT: Knowledge graph fine-tuning upon open-world knowledge. In The Thirty-eighth Annual Conference on Neural Information Processing Systems, 2024.

[47] Zhengyuan Yang, Linjie Li, Kevin Lin, Jianfeng Wang, Chung-Ching Lin, Zicheng Liu, and Lijuan Wang. The dawn of lmms: Preliminary explorations with gpt-4v (ision). arXiv preprint arXiv:2309.17421, 2023.

[48] Hejie Cui, Lingjun Mao, Xin Liang, Jieyu Zhang, Hui Ren, Quanzheng Li, Xiang Li, and Carl Yang. Biomedical visual instruction tuning with clinician preference alignment. arXiv preprint arXiv:2406.13173, 2024.

[49] Junru Chen, Yang Yang, Tao Yu, Yingying Fan, Xiaolong Mo, and Carl Yang. Brainnet: Epileptic wave detection from seeg with hierarchical graph diffusion learning. In Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD), 2022.

[50] Hejie Cui, Xinyu Fang, Zihan Zhang, Ran Xu, Xuan Kan, Xin Liu, Yue Yu, Manling Li, Yangqiu Song, and Carl Yang. Open visual knowledge extraction via relation-oriented multimodality model prompting. In Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2024.

[51] Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A Shamma, et al. Visual genome: Connecting language and vision using crowdsourced dense image annotations. International journal of computer vision, 123:32–73, 2017.

[52] Robyn Speer, Joshua Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge. In Proceedings of the AAAI conference on artificial intelligence, volume 31, 2017.

[53] Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. Comet: Commonsense transformers for knowledge graph construction. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2019.

[54] Yushi Hu, Hang Hua, Zhengyuan Yang, Weijia Shi, Noah A Smith, and Jiebo Luo. Promptcap: Prompt-guided image captioning for vqa with gpt-3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 2963–2975, 2023.

[55] Zhantao Yang, Han Zhang, Fangyi Chen, Anudeepsekhar Bolimera, and Marios Savvides. Hierarchical knowledge graph construction from images for scalable e-commerce. arXiv preprint arXiv:2410.21237, 2024.

[56] Debadutta Dash, Rahul Thapa, Juan M Banda, Akshay Swaminathan, Morgan Cheatham, Mehr Kashyap, Nikesh Kotecha, Jonathan H Chen, Saurabh Gombar, Lance Downing, Rachel Pedreira, Ethan Goh, Angel Arnaout, Garret Kenn Morris, Honor Magon, Matthew P Lungren, Eric Horvitz, and Nigam 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.

[57] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2020.

[58] Chan Hee Song, Jiaman Wu, Clayton Washington, Brian M Sadler, Wei-Lun Chao, and Yu Su. Llm-planner: Few-shot grounded planning for embodied agents with large language models. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pages 2998–3009, 2023.

[59] Subbarao Kambhampati, Karthik Valmeekam, Lin Guan, Kaya Stechly, Mudit Verma, Siddhant Bhambri, Lucas Saldyt, and Anil Murthy. Llms can't plan, but can help planning in llm-modulo frameworks. arXiv preprint arXiv:2402.01817, 2024.

[60] Jinheon Baek, Alham Fikri Aji, Jens Lehmann, and Sung Ju Hwang. Direct fact retrieval from knowledge graphs without entity linking. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2023.

[61] Lei Wang, Wanyu Xu, Yihuai Lan, Zhiqiang Hu, Yunshi Lan, Roy Ka-Wei Lee, and Ee-Peng 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.

[62] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. In Proceedings of the International Conference on Learning Representations (ICLR), 2022.

[63] Linhao Luo, Yuan-Fang Li, Reza Haf, and Shirui Pan. Reasoning on graphs: Faithful and interpretable large language model reasoning. In Proceedings of the International Conference on Learning Representations (ICLR), 2024.

[64] Hongwei Wang, Hongyu Ren, and Jure Leskovec. Relational message passing for knowledge graph completion. In Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD), 2021.

[65] Wenpeng Yin, Qinyuan Ye, Pengfei Liu, Xiang Ren, and Hinrich Schütze. Llm-driven instruction following: Progresses and concerns. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.

[66] Tobias Schimanski, Jingwei Ni, Mathias Kraus, Elliott Ash, and Markus Leippold. Towards faithful and robust llm specialists for evidence-based question-answering. arXiv preprint arXiv:2402.08277, 2024.

[67] Siyuan Wang, Zhongyu Wei, Jiarong Xu, Taishan Li, and Zhihao Fan. Unifying structure reasoning and language model pre-training for complex reasoning. arXiv preprint arXiv:2301.08913, 2023.

[68] Jinhao Jiang, Kun Zhou, Xin Zhao, and Ji-Rong 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.

[69] Jianing Wang, Qiushi Sun, Xiang Li, and Ming Gao. Boosting language models reasoning with chain-of-knowledge prompting. arXiv preprint arXiv:2306.06427, 2023.

[70] Keheng Wang, Feiyu Duan, Sirui Wang, Peiguang Li, Yunsen Xian, Chuantao Yin, Wenge Rong, and Zhang Xiong. Knowledge-driven cot: Exploring faithful reasoning in llms for knowledge-intensive question answering. arXiv preprint arXiv:2308.13259, 2023.

[71] Costas Mavromatis and George Karypis. Gnn-rag: Graph neural retrieval for large language model reasoning. arXiv preprint arXiv:2405.20139, 2024.

[72] Yilin Wen, Zifeng Wang, and Jimeng Sun. Mindmap: Knowledge graph prompting sparks graph of thoughts in large language models. arXiv preprint arXiv:2308.09729, 2023.

[73] Jin Huang, Xingjian Zhang, Qiaozhu Mei, and Jiaqi Ma. Can llms effectively leverage graph structural information through prompts, and why? Transactions on Machine Learning Research, 2024.

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

[75] Edward Fredkin. Trie memory. Communications of the ACM, 3(9):490–499, 1960.

[76] Jiuzhou Han, Nigel Collier, Wray Buntine, and Ehsan Shareghi. Pive: Prompting with iterative verification improving graph-based generative capability of llms. arXiv preprint arXiv:2305.12392, 2023.

[77] Linhao Luo, Trang Vu, Dinh Phung, and Reza Haf. Systematic assessment of factual knowledge in large language models. In Findings of the Association for Computational Linguistics: EMNLP (EMNLP-Findings), 2023.

[78] Thi Nguyen, Linhao Luo, Fatemeh Shiri, Dinh Phung, Yuan-Fang Li, Thuy-Trang Vu, and Gholamreza 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.

[79] Pengcheng Jiang, Cao Xiao, Adam Richard Cross, and Jimeng Sun. Graphcare: Enhancing healthcare predictions with personalized knowledge graphs. In The Twelfth International Conference on Learning Representations, 2024.

[80] Ran Xu, Wenqi Shi, Yue Yu, Yuchen Zhuang, Bowen Jin, May Dongmei Wang, Joyce Ho, and Carl Yang. RAM-EHR: Retrieval augmentation meets clinical predictions on electronic health records. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 754–765, Bangkok, Thailand, August 2024. Association for Computational Linguistics.

[81] Yinghao Zhu, Changyu Ren, Zixiang Wang, Xiaochen Zheng, Shiyun Xie, Junlan Feng, Xi Zhu, Zhoujun Li, Liantao Ma, and Chengwei Pan. Emerge: Enhancing multimodal electronic health records predictive modeling with retrieval-augmented generation. In Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, CIKM '24, page 3549–3559, New York, NY, USA, 2024. Association for Computing Machinery.

[82] Pengcheng Jiang, Cao Xiao, Minhao Jiang, Parminder Bhatia, Taha Kass-Hout, Jimeng Sun, and Jiawei Han. Reasoning-enhanced healthcare predictions with knowledge graph community retrieval. The Thirteenth International Conference on Learning Representations, 2025.

[83] Haohui Lu and Usman Naseem. Can large language models enhance predictions of disease progression? investigating through disease network link prediction. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pages 17703–17715, Miami, Florida, USA, November 2024. Association for Computational Linguistics.

[84] Xiaorui Su, Shvat Messica, Yepeng Huang, Ruth Johnson, Lukas Fesser, Shanghua Gao, Faryad Sahneh, and Marinka Zitnik. Multimodal medical code tokenizer, 2025.

[85] Xuejiao Zhao, Siyan Liu, Su-Yin Yang, and Chunyan Miao. Medrag: Enhancing retrieval-augmented generation with knowledge graph-elicited reasoning for healthcare copilot. arXiv preprint arXiv:2502.04413, 2025.

[86] Hejie Cui, Xinyu Fang, Ran Xu, Xuan Kan, Joyce C Ho, and Carl Yang. Multimodal fusion of ehr in structures and semantics: Integrating clinical records and notes with hypergraph and llm. arXiv preprint arXiv:2403.08818, 2024.

[87] Zixiang Wang, Yinghao Zhu, Junyi Gao, Xiaochen Zheng, Yuhui Zeng, Yifan He, Bowen Jiang, Wen Tang, Ewen M Harrison, Chengwei Pan, et al. Retcare: Towards interpretable clinical decision making through llm-driven medical knowledge retrieval. In Artificial Intelligence and Data Science for Healthcare: Bridging Data-Centric AI and People-Centric Healthcare, 2024.

[88] Hejie Cui, Zhuocheng Shen, Jieyu Zhang, Hui Shao, Lianhui Qin, Joyce C Ho, and Carl Yang. Llms-based few-shot disease predictions using ehr: A novel approach combining predictive agent reasoning and critical agent instruction. arXiv preprint arXiv:2403.15464, 2024.

[89] Ana Sanchez-Fernandez, Elisabeth Rumetshofer, Sepp Hochreiter, and Günter Klambauer. Contrastive learning of image-and structure-based representations in drug discovery. In ICLR2022 Machine Learning for Drug Discovery, 2022.

[90] Pengcheng Jiang, Cao Xiao, Tianfan Fu, Jimeng Sun, and Jiawei Han. Bi-level contrastive learning for knowledge-enhanced molecule representations. In Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence, 2025.

[91] Qing Ye, Chang-Yu Hsieh, Ziyi Yang, Yu Kang, Jiming Chen, Dongsheng Cao, Shibo He, and Tingjun Hou. A unified drug–target interaction prediction framework based on knowledge graph and recommendation system. Nature Communications, 12(1):6775, 2021.

[92] Yin Fang, Qiang Zhang, Ningyu Zhang, Zhuo Chen, Xiang Zhuang, Xin Shao, Xiaohui Fan, and Huajun Chen. Knowledge graph-enhanced molecular contrastive learning with functional prompt. Nature Machine Intelligence, 5(5):542–553, 2023.

[93] Yinhan He, Zaiyi Zheng, Patrick Soga, Yaochen Zhu, Yushun Dong, and Jundong Li. Explaining graph neural networks with large language models: A counterfactual perspective on molecule graphs. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, Findings of the Association for Computational Linguistics: EMNLP 2024, pages 7079–7096, Miami, Florida, USA, November 2024. Association for Computational Linguistics.

[94] Taicheng Guo, Bozhao Nan, Zhenwen Liang, Zhichun Guo, Nitesh Chawla, Olaf Wiest, Xiangliang Zhang, et al. What can large language models do in chemistry? a comprehensive benchmark on eight tasks. Advances in Neural Information Processing Systems, 36:59662–59688, 2023.

[95] Kexin Huang, Tony Zeng, Soner Koc, Alexandra Pettet, Jingtian Zhou, Mika Jain, Dongbo Sun, Camilo Ruiz, Hongyu Ren, Laurence J Howe, et al. Small-cohort gwas discovery with ai over massive functional genomics knowledge graph. medRxiv, pages 2024–12, 2024.

[96] Kexin Huang, Payal Chandak, Qianwen Wang, Shreyas Havaldar, Akhil Vaid, Jure Leskovec, Girish N Nadkarni, Benjamin S Glicksberg, Nils Gehlenborg, and Marinka Zitnik. A foundation model for clinician-centered drug repurposing. Nature Medicine, 30(12):3601–3613, 2024.

[97] Tareq B Malas, Wytze J Vlietstra, Roman Kudrin, Sergey Starikov, Mohammed Charrout, Marco Roos, Dorien JM Peters, Jan A Kors, Rein Vos, Peter AC 't Hoen, et al. Drug prioritization using the semantic properties of a knowledge graph. Scientific reports, 9(1):6281, 2019.

[98] Chufan Gao, Tianfan Fu, and Jimeng Sun. Language interaction network for clinical trial approval estimation. arXiv preprint arXiv:2405.06662, 2024.

[99] Qiao Jin, Zifeng Wang, Charalampos S Floudas, Fangyuan Chen, Changlin Gong, Dara Bracken-Clarke, Elisabetta Xue, Yifan Yang, Jimeng Sun, and Zhiyong Lu. Matching patients to clinical trials with large language models. Nature communications, 15(1):9074, 2024.

[100] Rui Yang, Boming Yang, Aosong Feng, Sixun Ouyang, Moritz Blum, Tianwei She, Yuang Jiang, Freddy Lecue, Jinghui Lu, and Irene Li. Graphusion: A rag framework for knowledge graph construction with a global perspective. arXiv preprint arXiv:2410.17600, 2024.

[101] Karthik Soman, Peter W Rose, John H Morris, Rabia E Akbas, Brett Smith, Braian Peetoom, Catalina Villouta-Reyes, Gabriel Cerono, Yongmei Shi, Angela Rizk-Jackson, et al. Biomedical knowledge graph-optimized prompt generation for large language models. Bioinformatics, 40(9):btae560, 2024.

[102] Duy H. Ho, Udiptaman Das, Regina Ho, and Yugyung Lee. Leveraging multi-agent systems and large language models for diabetes knowledge graphs. In 2024 IEEE International Conference on Big Data (BigData), pages 3401–3410, 2024.

[103] Yichong Zhang and Yongtao Hao. Traditional chinese medicine knowledge graph construction based on large language models. Electronics, 13(7):1395, 2024.

[104] Tianhan Xu, Yixun Gu, Mantian Xue, Renjie Gu, Bin Li, and Xiang Gu. Knowledge graph construction for heart failure using large language models with prompt engineering. Frontiers in Computational Neuroscience, 18:1389475, 2024.

[105] Lang Cao, Jimeng Sun, and Adam Cross. Autord: An automatic and end-to-end system for rare disease knowledge graph construction based on ontologies-enhanced large language models. arXiv preprint arXiv:2403.00953, 2024.

[106] Joseph D Romano, Van Truong, Rachit Kumar, Mythreye Venkatesan, Britney E Graham, Yun Hao, Nick Matsumoto, Xi Li, Zhiping Wang, Marylyn D Ritchie, et al. The alzheimer's knowledge base: A knowledge graph for alzheimer disease research. Journal of Medical Internet Research, 26:e46777, 2024.

[107] Ran Xu, Hejie Cui, Yue Yu, Xuan Kan, Wenqi Shi, Yuchen Zhuang, May Dongmei Wang, Wei Jin, Joyce Ho, and Carl Yang. Knowledge-infused prompting: Assessing and advancing clinical text data generation with large language models. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, Findings of the Association for Computational Linguistics: ACL 2024, pages 15496–15523, Bangkok, Thailand, August 2024. Association for Computational Linguistics.

[108] Pengcheng Jiang, Cao Xiao, Adam Richard Cross, and Jimeng Sun. Graphcare: Enhancing healthcare predictions with personalized knowledge graphs. In The Twelfth International Conference on Learning Representations, 2024.

[109] Zhi-Qing Li, Zi-Xuan Fu, Wen-Jun Li, Hao Fan, Shu-Nan Li, Xi-Mo Wang, and Peng Zhou. Prediction of diabetic macular edema using knowledge graph. Diagnostics, 13(11):1858, 2023.

[110] Pratik Joshi, V Masilamani, and Anirban Mukherjee. A knowledge graph embedding based approach to predict the adverse drug reactions using a deep neural network. Journal of Biomedical Informatics, 132:104122, 2022.

[111] Shike Wang, Fan Xu, Yunyang Li, Jie Wang, Ke Zhang, Yong Liu, Min Wu, and Jie Zheng. Kg4sl: knowledge graph neural network for synthetic lethality prediction in human cancers. Bioinformatics, 37(Supplement_1):i418–i425, 2021.

[112] Xiaofeng Huang, Jixin Zhang, Zisang Xu, Lu Ou, and Jianbin Tong. A knowledge graph based question answering method for medical domain. PeerJ Computer Science, 7:e667, 2021.

[113] Rui Yang, Haoran Liu, Edison Marrese-Taylor, Qingcheng Zeng, Yu He Ke, Wanxin Li, Lechao Cheng, Qingyu Chen, James Caverlee, Yutaka Matsuo, et al. Kg-rank: Enhancing large language models for medical qa with knowledge graphs and ranking techniques. arXiv preprint arXiv:2403.05881, 2024.

[114] Yating Yin, Lei Zhang, Yiguo Wang, Mingqiang Wang, Qiming Zhang, and Guozheng Li. Question answering system based on knowledge graph in traditional chinese medicine diagnosis and treatment of viral hepatitis b. BioMed research international, 2022(1):7139904, 2022.

[115] Gengxian Zhou, E Haihong, Zemin Kuang, Ling Tan, Xiaoxuan Xie, Jundi Li, and Haoran Luo. Clinical decision support system for hypertension medication based on knowledge graph. Computer Methods and Programs in Biomedicine, 227:107220, 2022.

[116] Xiaorui Su, Yibo Wang, Shanghua Gao, Xiaolong Liu, Valentina Giunchiglia, Djork-Arné Clevert, and Marinka Zitnik. Knowledge graph based agent for complex, knowledge-intensive qa in medicine. arXiv preprint arXiv:2410.04660, 2024.

[117] Haixia Sun, Jin Xiao, Wei Zhu, Yilong He, Sheng Zhang, Xiaowei Xu, Li Hou, Jiao Li, Yuan Ni, Guotong Xie, et al. Medical knowledge graph to enhance fraud, waste, and abuse detection on claim data: Model development and performance evaluation. JMIR Medical Informatics, 8(7):e17653, 2020.

[118] Theodora S Brisimi, Vanessa Lopez, Valentina Rho, Marco Sbodio, Gabriele Picco, Morten Kristiansen, John Segrave-Daly, and Conor Cullen. Ontology-guided policy information extraction for healthcare fraud detection. In Digital Personalized Health and Medicine, pages 879–883. IOS Press, 2020.

[119] Timofey V Ivanisenko, Pavel S Demenkov, and Vladimir A Ivanisenko. An accurate and efficient approach to knowledge extraction from scientific publications using structured ontology models, graph neural networks, and large language models. International Journal of Molecular Sciences, 25(21):11811, 2024.