A Survey of Large Language Models and Knowledge Graphs: Interconnections and Applications

Abstract

In the rapidly evolving landscape of artificial intelligence, the integration of large language models (LLMs) and knowledge graphs (KGs) represents a pivotal advancement, enhancing reasoning, representation, and application capabilities across various domains. This survey explores the synergy between LLMs and KGs, highlighting their role in improving natural language processing (NLP) tasks, semantic reasoning, and knowledge graph completion. The integration of structured knowledge from KGs into LLMs addresses limitations in context-awareness and factual reasoning, significantly reducing hallucination rates and enhancing decision-making processes in critical scenarios such as emergency management and biomedical research. Despite these advancements, challenges persist in data quality, scalability, and the integration of heterogeneous information sources. The survey underscores the need for robust data management practices and scalable solutions to enhance the reliability and effectiveness of AI systems. Future research directions include the development of more efficient algorithms, the integration of semantic reasoning with graph embeddings, and the exploration of multilingual and domain-specific knowledge graphs. By addressing these challenges, the AI community can drive the development of more sophisticated, reliable, and context-aware systems, ultimately advancing the field of artificial intelligence. This survey provides a comprehensive overview of the current state and future directions of LLM and KG integration, emphasizing the transformative potential of these technologies in enhancing AI applications.

1 Introduction

1.1 Interconnected Fields Overview

The interplay between large language models (LLMs), knowledge graphs (KGs), and related domains such as natural language processing (NLP), semantic reasoning, and graph embeddings creates a robust framework that significantly enhances artificial intelligence systems. LLMs, which excel in generating human-like text and understanding context, leverage the structured world knowledge provided by KGs. This integration is vital for improving reasoning and context-awareness in AI applications, as KGs systematically represent real-world entities and their interrelations, facilitating enhanced inferencing capabilities and complex semantic interpretation [1, 2, 3].

KGs serve as foundational structures for storing and retrieving interconnected data, which can enhance the performance of LLMs and improve search engines by offering structured context for more accurate information retrieval and reasoning. This synergy addresses the shortcomings of existing language representation models that primarily depend on unstructured corpora, which often struggle with polysemy and the diverse meanings of words [4].

The integration of KGs with LLMs spans various applications, including conversational recommender systems (CRS), where natural language interactions and dialogue history are utilized to infer user preferences and deliver precise recommendations [5]. Additionally, LLMs are advancing ontology

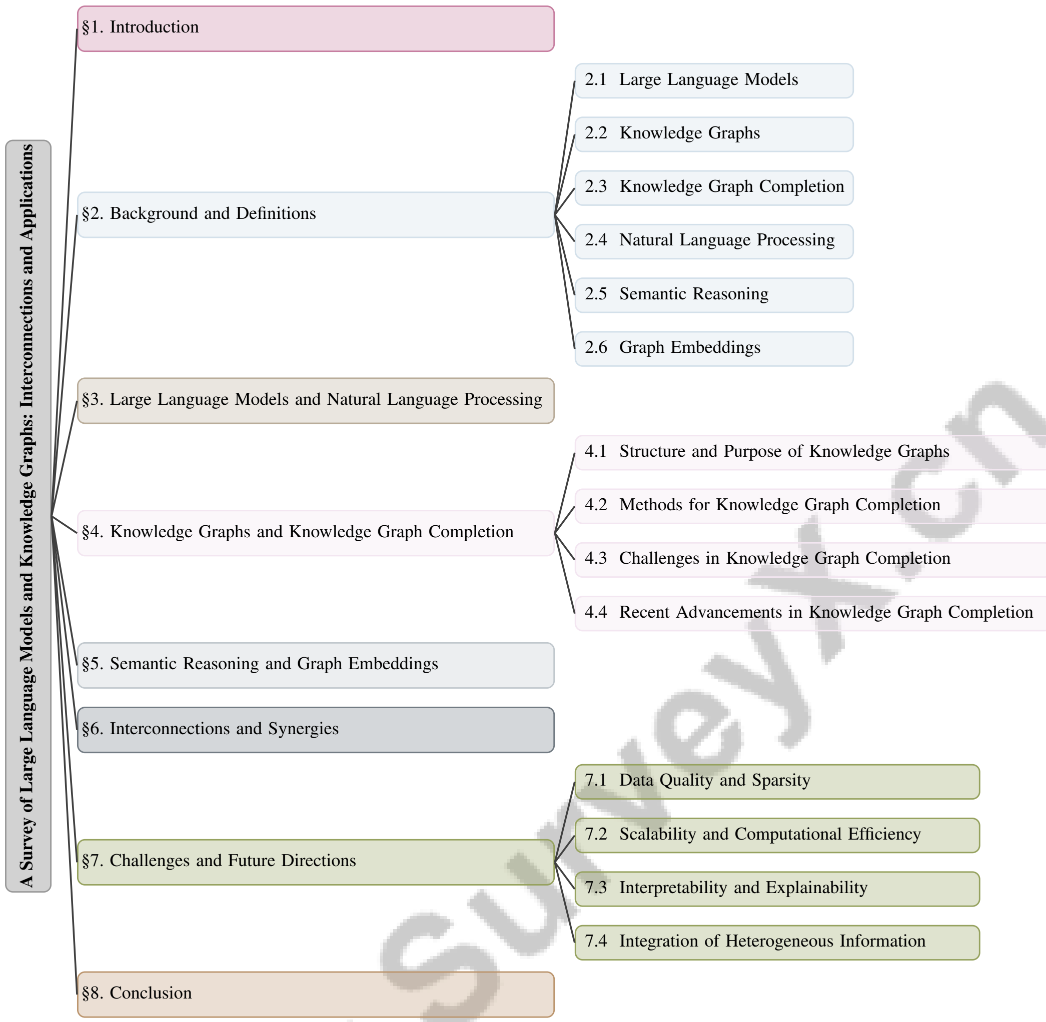


Figure 1: chapter structure

construction and knowledge graph engineering by automating the construction and enrichment of KGs through the extraction and structuring of information from extensive text corpora [6].

Despite challenges in integrating rich world knowledge from KGs into language models, ongoing research is dedicated to overcoming these obstacles, thereby enhancing AI systems' capabilities in understanding and interacting with human language [7]. The dynamic nature of knowledge representation through event knowledge graphs underscores the limitations of static KGs, which primarily cover unchanging facts about entities [8]. Furthermore, the evolving integration of LLMs with KGs necessitates benchmarks to evaluate LLM capabilities in handling complex queries, such as those involving SPARQL, which is critical for querying KGs [9].

The integration of structured and unstructured data sources, particularly in biomedical research, illustrates the interconnectedness of LLMs and KGs in enhancing applications like cancer diagnosis and treatment [10]. Moreover, KGs combined with LLMs improve decision-making processes in emergency management scenarios [11]. By addressing LLMs' limitations in recalling factual knowledge, KGs facilitate the generation of knowledge-grounded content [12]. The creation of domain-specific KGs, such as those for cybersecurity, is essential for organizing vast quantities of scientific literature and extracting actionable insights from unstructured text [13].

Additionally, the development of KGs for underrepresented languages, such as Bangla, presents both challenges and opportunities in constructing comprehensive datasets and tools for diverse linguistic contexts [14]. Integrating speech input with KGs to enhance recognition accuracy in speech interfaces emphasizes the need for innovative approaches to link spoken language with structured knowledge

[15]. Finally, benchmarking dialogue systems to ensure mutual understanding among participants addresses the challenge of bridging information gaps in conversations [16].

1.2 Significance of Each Field

The integration of large language models (LLMs) with external knowledge sources, particularly knowledge graphs (KGs), is pivotal in advancing natural language processing (NLP) tasks by enhancing LLMs' factual reasoning capabilities. Knowledge graph-enhanced pre-trained language models (KGPLMs) significantly bolster LLMs' ability to perform complex reasoning and understand context, thereby reducing reliance on extensive labeled datasets [12]. This synergy is especially beneficial in applications like conversational recommender systems (CRSs), where structured knowledge from KGs helps bridge knowledge gaps and improves interpretability [5].

In emergency management, LLMs present new avenues for enhancing targeted machine intelligence, facilitating more effective decision-making processes [11]. The ability of LLMs to incorporate structured knowledge from KGs enhances their performance in critical scenarios, where accuracy and timely responses are essential.

The significance of knowledge graphs also lies in their capacity to capture implicit and loosely structured knowledge, crucial for advancing AI systems' reasoning capabilities beyond traditional encyclopedic knowledge. Frameworks like BanglaAutoKG automate the construction of knowledge graphs from Bangla text, thereby improving information processing and reasoning in underrepresented linguistic contexts [14].

Recent advancements in LLMs demonstrate their potential to complement KG construction by automating resource-intensive processes such as ontology alignment and population. The integration of pre-trained language models (PLMs) with graph neural networks (GNNs) is particularly noteworthy, as it enhances knowledge sharing and reasoning in complex queries, contributing to the development of sophisticated AI systems capable of tackling increasingly intricate tasks [17].

Knowledge graphs are essential in enhancing AI systems, providing structured frameworks that improve applications across diverse sectors, including education, healthcare, and social media. Completing KGs is vital for inferring missing information, ensuring the comprehensiveness and accuracy of the knowledge stored within them. Innovative frameworks like SKarREC combine knowledge from LLMs with relationships derived from KGs to enhance concept recommendations, showcasing the transformative potential of these integrations [9].

The integration of LLMs, KGs, and associated technologies is crucial for advancing artificial intelligence and machine learning systems. This synergy enhances their ability to perform complex tasks by improving understanding and reasoning. For instance, LLMs facilitate the semi-automated construction of KGs, reducing reliance on human experts while ensuring data quality through methodologies like Retrieval-Augmented Generation (RAG). Empirical studies indicate that LLMs, such as GPT-4, excel in reasoning tasks and effectively assist in information extraction and query processing within KGs. This interconnected framework streamlines knowledge management and addresses challenges like information hallucinations and domain-specific limitations, paving the way for more sophisticated AI solutions capable of delivering accurate and contextually relevant outcomes [18, 19, 20, 21, 22].

1.3 Structure of the Survey

This survey is structured to provide a comprehensive exploration of large language models (LLMs) and knowledge graphs (KGs), detailing their definitions, applications, challenges, and future directions [23]. The paper begins with an Introduction that sets the stage by offering an overview of the interconnected fields and their significance in advancing artificial intelligence (AI) systems. Following this, the Background and Definitions section delves into the core concepts underlying LLMs, KGs, and related areas such as natural language processing (NLP), semantic reasoning, and graph embeddings.

Subsequent sections explore the role of LLMs in NLP, highlighting their contributions to generating human-like text and understanding context. The survey transitions to a discussion on knowledge graphs, examining their structure, purpose, and methods for knowledge graph completion. This is followed by an analysis of semantic reasoning and graph embeddings, emphasizing their importance in transforming data for AI applications.

AI-generated, for reference only.

The paper further investigates the synergies between LLMs, KGs, and related technologies, showcasing how their integration enhances reasoning, representation, and application across various domains. The Challenges and Future Directions section addresses key issues such as data quality, scalability, and interpretability, providing insights into potential advancements in the field. Finally, the Conclusion summarizes the key points discussed, underscoring the significance of continued research and collaboration in these rapidly evolving areas.The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Large Language Models

Large Language Models (LLMs) have revolutionized natural language processing (NLP) through their ability to generate human-like text and comprehend context, following the 'pre-train, prompt, predict' methodology [24]. Despite their proficiency in various language tasks, LLMs face challenges in integrating structured knowledge from knowledge graphs (KGs), which can lead to information loss [25]. Models like K-BERT illustrate efforts to infuse domain knowledge from KGs, enhancing sentence representations [26].

Advancements in LLMs focus on enhancing reasoning capabilities and minimizing erroneous or 'hallucinated' responses, particularly in complex reasoning tasks [11]. A notable challenge is their performance on domain-specific tasks requiring expert knowledge, as they often rely on general representations from large-scale corpora, leading to hallucinations and reduced interpretability [27]. Efforts to integrate LLMs with external knowledge sources, such as knowledge-enhanced pre-trained language models (PLMs), aim to incorporate structured knowledge from KGs to improve world knowledge understanding [28]. However, LLMs still generate incorrect answers due to reliance on potentially insufficient or outdated internalized knowledge [29].

Research continues to evaluate reasoning capabilities, debating whether LLMs genuinely reason or simulate reasoning processes [30]. Experiments with models like GPT-3.5-Turbo, GPT-4o, Llama-3-8B, and Llama-3-70B explore these capabilities and identify improvement areas [16]. Existing models struggle with understanding complex commonsense knowledge involving multi-token concepts, crucial for nuanced event relationships [31].

Current research addresses challenges by enhancing integration with structured knowledge sources like KGs and improving reasoning abilities. Efforts include combining LLMs with frameworks such as the Deep Document Model for nuanced document representation and Knowledge Graph-enhanced Query Processing for optimized query handling, ultimately aiming to enhance information retrieval accuracy and knowledge management effectiveness across domains, including academia and education [18, 12, 21, 22]. These advancements promise to elevate AI systems to new levels of sophistication and utility.

2.2 Knowledge Graphs

Knowledge Graphs (KGs) are structured representations of knowledge that significantly enhance decision-making systems' reliability across various domains [11]. Organized as multi-relational directed labeled graphs, KGs represent entities as nodes and relationships as edges, facilitating real-world facts encoding and retrieval [14]. This structure supports applications like web search, question answering, and commonsense reasoning by formalizing interconnected data representation and access [32].

KG construction employs diverse methodologies, including crowdsourced knowledge descriptions and LLMs for generating knowledge from unstructured sources, which is particularly valuable for creating domain-specific KGs for underrepresented languages like Bangla [14]. In fields like biomedical research and cybersecurity, KGs facilitate data integration from multiple sources to support tasks such as cancer biomarker discovery and comprehensive knowledge framework generation from scientific literature [11]. However, challenges related to data completeness impact their effectiveness in tasks like entity ranking and information retrieval.

KGs serve as a vital framework for advancing AI systems, providing structured, interpretable, and interconnected knowledge representations. They enable the organization and analysis of vast information amounts, particularly in academia, encapsulating metadata about research publications, authors, topics, and citations. By leveraging advanced NLP and machine learning techniques, KGs enhance entity and relationship extraction from scientific literature, addressing critical challenges in AI, including knowledge acquisition, reasoning, and representation [33, 34, 3]. Their integration with advanced technologies like pre-trained language models and graph neural networks continues to enrich AI's capability to derive insights from complex datasets, making them indispensable in the advancement of intelligent systems.

2.3 Knowledge Graph Completion

Knowledge Graph Completion (KGC) is crucial for addressing KGs' inherent incompleteness, which often lacks sufficient data for robust applications across domains. KGC involves predicting missing entities or relations within the graph, enhancing KGs' utility and comprehensiveness [35]. This task is essential for leveraging the rich semantic information embedded in KGs, often underutilized due to their incomplete nature.

Traditional KGC approaches primarily utilize Knowledge Graph Embeddings (KGEs) to transform entities and relationships into continuous vector spaces, capturing semantic and structural attributes while integrating structured and unstructured information. These embeddings are vital for applications like entity linking and question answering, addressing high computational and storage costs associated with large-scale KGs. Recent advancements focus on improving the interpretability of these embeddings, clarifying the semantic meaning of individual dimensions to enhance their effectiveness in reasoning tasks [36, 37, 3]. However, challenges persist in sparse KGs, where insufficient structural information weakens performance in link prediction tasks. Techniques such as hierarchy-aware embeddings and graph convolutional networks (GCNs) have been introduced to improve link prediction accuracy by preserving inherent hierarchical and logical data structures.

Integrating LLMs with KGs has shown promise in enhancing KGC tasks. For example, using pre-trained language models (PLMs) with graph neural networks (GNNs) has been explored to improve knowledge graph question answering (KGQA) by facilitating better knowledge sharing and reasoning for complex queries. Temporal Knowledge Graph Completion (TKGC) presents unique challenges by requiring reasoning at specific timestamps to fill in missing factual elements, necessitating an understanding of how temporal dynamics influence events. Recent advancements in TKGC methods, including the use of PLMs and innovative approaches to temporal information processing, aim to improve predictive capabilities, enhancing effectiveness in capturing temporal relationships within KGs [38, 39, 40]. This aspect of KGC is vital for applications dependent on temporal dynamics, such as event progression prediction.

Despite advancements, KGC remains complex, with ongoing challenges in acquiring accurate knowledge and enhancing reasoning capabilities. Addressing these challenges is essential for improving KGC reliability and effectiveness, thereby expanding the applicability of KGs across diverse domains. Leveraging advanced methodologies and addressing existing limitations are crucial for constructing organized and comprehensive knowledge structures. The scarcity of automatic KGC methods tailored for specific languages, like Bangla, underscores the need for multilingual approaches to enhance information processing and reasoning across linguistic contexts [14].

2.4 Natural Language Processing

Natural Language Processing (NLP) is a foundational aspect of artificial intelligence, enabling interaction between humans and machines through language comprehension and generation. The integration of LLMs with NLP has driven advancements in both Natural Language Understanding (NLU) and Natural Language Generation (NLG), as highlighted in recent surveys on knowledge-enhanced pre-trained language models (KE-PLMs) [41]. These models leverage extensive pre-training on large datasets, allowing for improved semantic understanding and generation across various linguistic tasks.

A significant challenge in NLP is disambiguating polysemous words and idiomatic expressions, which often confound pre-trained language models (PTLMs) due to their reliance on local context. Addressing these challenges necessitates integrating structured knowledge from KGs to enhance semantic parsing and context comprehension [42]. This integration is critical for converting natural language dialogue into executable SPARQL queries for KGs, as demonstrated by benchmarks focusing on semantic parsing in conversational interactions [43].

Knowledge Graph Augmented Information Extraction (KGAIE) exemplifies the application of NLP technologies in extracting task-specific information from regulatory documents, showcasing NLP's role in anchoring information extraction processes to structured knowledge [44]. The Text2KGBench benchmark evaluates language models' ability to generate KGs from natural language text guided by ontologies, highlighting the interplay between NLP and KG construction [45].

Frameworks like KELM enhance PLM representations by incorporating structured knowledge from KGs, creating unified knowledge-enhanced text graphs during fine-tuning. This approach significantly improves language models' comprehension and generation abilities through enriched contextual understanding [28]. Additionally, integrating KG relations into citation generation processes underscores the importance of contextual understanding in NLP tasks, enhancing relationships between cited and citing documents [8].

As NLP technology advances, integrating LLMs with KGs plays a crucial role in enhancing AI systems' capabilities. This synergy improves language processing and understanding while facilitating more accurate reasoning by leveraging structured factual knowledge from KGs. Recent research demonstrates that combining LLMs with KGs can optimize query processing and enhance knowledge utilization, enabling more effective interactions and knowledge management in academic contexts. This collaborative approach addresses traditional models' limitations by providing a more comprehensive framework for natural language understanding and reasoning [41, 21, 5]. This integration not only enhances AI applications' semantic understanding but also extends their utility across diverse domains, driving the development of sophisticated and context-aware AI solutions.

2.5 Semantic Reasoning

Semantic reasoning is integral to AI systems, enabling the derivation of meaning from complex data structures and narratives. It employs logical rules and external commonsense knowledge to understand and model dependencies among data points, essential for generating coherent and contextually relevant outputs. This process is particularly important in knowledge graphs, where accurately modeling relationships between entities is crucial for effective reasoning and insight extraction, especially in rapidly evolving fields like scientific research and domain-specific applications. Knowledge graphs encompass a variety of entities—authors, organizations, and research topics—and their interconnections, facilitating the analysis and management of vast amounts of data generated by scholarly publications and other complex information sources [33, 44].

A primary challenge in semantic reasoning is modeling entities based solely on their relationships within knowledge graphs, without relying on textual descriptions or sequence information. This limitation can hinder accurate meaning derivation [46]. Logical rules, fundamental for deriving meaning in knowledge graphs, also present challenges regarding computational intensity and scalability, complicating the reasoning process [47].

The significance of semantic reasoning extends to commonsense reasoning, where external commonsense knowledge is necessary for complex reasoning tasks. Current commonsense knowledge graphs often inadequately represent human reasoning complexities, particularly in modeling dependencies and varying belief likelihoods [48]. This inadequacy underscores the need for sophisticated methods to enhance representation and reasoning capabilities in AI systems.

Moreover, semantic reasoning is vital for deriving meaning from complex narratives, as evidenced by benchmarks that focus on explaining reasoning processes, such as identifying characters as criminals in narrative contexts [49]. These applications emphasize semantic reasoning's significance in advancing AI's ability to process and interpret complex data, ultimately contributing to more intelligent and context-aware systems.

2.6 Graph Embeddings

Graph embeddings represent a transformative approach in AI, bridging symbolic graph data and numerical machine learning models. By converting symbolic triples into real-valued vectors, knowledge graph embeddings (KGEs) facilitate the integration of graph-structured data with various machine learning algorithms, enhancing AI systems' capabilities in tasks like link prediction and node classification [50]. Generating KGEs often involves representing relationships among entities as sequences of words, as demonstrated by methods like KG2Vec, which employs a skip-gram model to learn embeddings from graph triples [51].

KGC methodologies are categorized into translation models, tensor factorization models, and neural network models, each utilizing specific scoring functions to predict missing links within graphs [52]. These models are pivotal in enhancing the comprehensiveness and accuracy of KGs, supporting effective knowledge representation and reasoning.

The integration of contextual embeddings with knowledge graph embeddings, as proposed in Combined Contextual and Knowledge Graph Embeddings (CKGE), exemplifies the synergy between different embedding techniques. By incorporating contextual embeddings from models like ELMo, CKGE allows for a richer representation of words and their relationships, leveraging both contextual and structural information [53]. This integration is crucial for applications requiring nuanced understanding and reasoning based on both textual and graph data.

Incorporating domain knowledge into language representation models, such as K-BERT, exemplifies the utility of graph embeddings in enhancing language models by injecting structured knowledge without extensive pre-training [27]. Additionally, the Structure Guided Multimodal Pretrained Transformer (SGMPT) demonstrates the application of structural information in processing multimodal data, highlighting the versatility of graph embeddings across various AI domains [54].

Graph embeddings are essential for transforming graph data into analyzable formats, enabling AI systems to efficiently process and reason with complex datasets. By representing entities and relationships as embedding vectors in a semantic space, these embeddings facilitate tasks like link prediction and knowledge graph completion, crucial for applications such as semantic search engines, question answering systems, and recommender systems. However, the diversity in embedding methods—such as Canonical decomposition, DistMult, and ComplEx—poses challenges in understanding their mechanisms and comparing effectiveness, particularly in handling incomplete knowledge graphs. Thus, advancements in multi-embedding interaction mechanisms and embedding interpretability are vital for enhancing performance and usability in AI applications relying on these complex data structures [55, 3]. Continued development and integration with contextual and multimodal embeddings promise to further advance intelligent systems' capabilities in understanding and interacting with the world.

3 Large Language Models and Natural Language Processing

The integration of Large Language Models (LLMs) into Natural Language Processing (NLP) has profoundly transformed the field, particularly in generating human-like text and understanding intricate linguistic structures. This section delves into the pivotal roles and contributions of LLMs, highlighting their impact on various NLP tasks and applications. As depicted in Figure 2, the figure illustrates the hierarchical structure of key concepts related to the integration of LLMs into NLP. It categorizes the LLMs' impact on NLP tasks, contextual embeddings, and semantic understanding, while also addressing the limitations and challenges associated with their application. This visual representation enhances our understanding of the multifaceted contributions of LLMs, providing a comprehensive overview of their advancements and the challenges that remain in the field.

3.1 Role of Large Language Models in NLP

LLMs have significantly advanced NLP by generating human-like text and understanding complex contexts through extensive dataset training [24]. They integrate structured knowledge from knowledge graphs (KGs) to enhance performance, as demonstrated by models like K-BERT, which leverage domain-specific knowledge [27]. The COMET framework and BERT-MK model illustrate LLMs' ability to generate knowledge-grounded text and improve medical NLP tasks by incorporating KG-based knowledge representations [6, 56].

Innovations like the Graph of Thoughts (GoT) and KELM frameworks integrate structured approaches and external knowledge into LLMs, enhancing reasoning and machine reading comprehension [57, 28]. Dialogue systems, exemplified by DiffKG, leverage KGs for human-like response generation [58]. T5-large's application in G2T tasks demonstrates LLMs' capability to utilize structured knowledge for improved text generation [59].

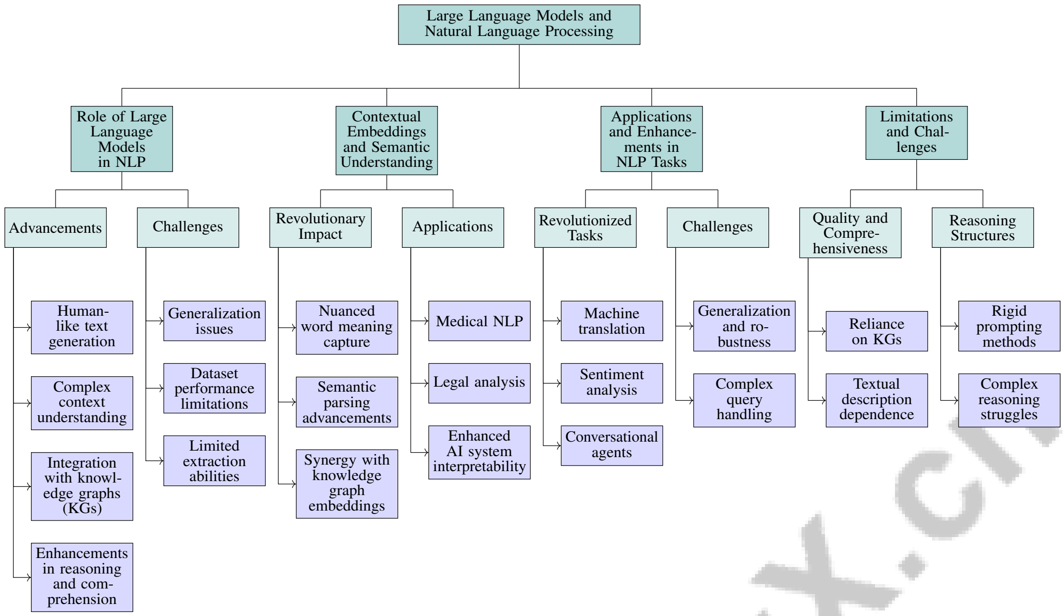


Figure 2: This figure illustrates the hierarchical structure of key concepts in the integration of Large Language Models (LLMs) into Natural Language Processing (NLP), highlighting their roles, advancements, applications, and challenges. It categorizes the LLMs' impact on NLP tasks, contextual embeddings, and semantic understanding while addressing limitations and challenges in their application.

Despite advancements, challenges in generalization and dataset performance persist. While GPT-4 excels in reasoning tasks, its extraction abilities are limited [18]. Models like CoCoLM enhance contextual understanding by integrating complex commonsense knowledge, while frameworks such as UAG improve reasoning reliability through uncertainty quantification [31, 60]. Techniques like Retrieval-Augmented Generation (RAG) and KG utilization continue to enhance LLM capabilities, addressing challenges like memory limitations and domain-specific issues [22, 41].

3.2 Contextual Embeddings and Semantic Understanding

Contextual embeddings have revolutionized semantic understanding in NLP by capturing nuanced meanings of words within their contexts, surpassing traditional word embeddings [42]. This advancement is critical for semantic parsing, transforming natural language into structured representations like SPARQL queries for effective KG interaction [43].

The synergy between contextual and knowledge graph embeddings, as seen in CKGE, leverages models like ELMo to provide richer word and relationship representations [53]. This integration is crucial for applications requiring deep semantic understanding, combining contextual and structural information strengths. Benchmarks like Text2KGBench highlight the interplay between contextual understanding and KG construction, showcasing models' ability to generate KGs from natural language text [45].

Frameworks like KELM enhance pre-trained language model representations through structured knowledge, improving semantic comprehension and generation capabilities [28]. Contextual embeddings also enhance AI systems' interpretability, facilitating applications in diverse fields, including medical NLP and legal analysis, by integrating advanced contextualized representations with structured knowledge [61, 56].

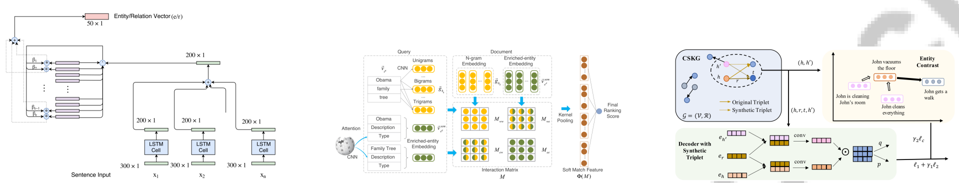
3.3 Applications and Enhancements in NLP Tasks

LLMs have revolutionized NLP tasks, advancing applications like machine translation, sentiment analysis, and conversational agents. In conversational recommender systems, LLMs use dialogue history to infer user preferences for personalized recommendations [5]. Their integration with

KGs enhances information extraction accuracy, crucial in biomedical research for tasks like cancer diagnosis [44, 10].

Frameworks like KELM improve machine reading comprehension and language generation by incorporating structured KG knowledge [28]. Models like COMET generate commonsense knowledge tuples, enriching contextual text understanding [6]. The integration of LLMs with graph embeddings advances semantic parsing, enhancing AI systems' interaction with KGs for effective information retrieval and reasoning [43]. Contextual embeddings with KG embeddings, exemplified by CKGE, offer richer semantic understanding [53].

Challenges remain in improving LLMs' generalization and robustness across datasets. Techniques like uncertainty quantification, as proposed by the UAG framework, aim to enhance reasoning reliability, addressing limitations in handling complex queries [60]. Ongoing development of embedding techniques and structured knowledge integration promises further advancements in NLP tasks, driving sophisticated, context-aware AI solutions.



(a) A neural network model for entity and relation recognition[62]

(b) Query-Document Interaction for Entity Matching[63]

(c) CSKG: A Novel Approach for Entity Contrastive Learning[64]

Figure 3: Examples of Applications and Enhancements in NLP Tasks

As illustrated in Figure 3, LLMs have significantly advanced NLP tasks. Three diverse models exemplify these advancements: a neural network for entity and relation recognition, a query-document interaction model for entity matching, and the CSKG model for entity contrastive learning. The first model identifies entities and relationships within text using a layered LSTM architecture. The second model employs query-document interaction, transforming both into embeddings and calculating interaction matrices for entity matching. The CSKG model introduces a graph-based structure for entity contrastive learning, generating synthetic triplets for enhanced learning. These models demonstrate LLMs' potential in improving information extraction, entity matching, and learning processes [62, 63, 64].

3.4 Limitations and Challenges

Despite LLMs' significant advancements in NLP, several limitations and challenges persist. A primary concern is the reliance on the quality and comprehensiveness of KGs, as incomplete or biased KGs can adversely affect semantic extensions and reasoning performance [65]. The dependence on textual descriptions for entities presents a barrier, as entities lacking descriptions are often excluded from predictions, limiting applicability in diverse contexts [66].

Existing prompting methods, like Chain-of-Thought (CoT) and Tree of Thoughts (ToT), impose rigid reasoning structures, potentially hindering advanced capabilities. The Graph of Thoughts (GoT) approach offers a more flexible framework [67], yet LLMs still struggle with complex reasoning, raising questions about their reasoning processes' depth [30]. Models like CoCoLM, integrating complex commonsense knowledge, face challenges related to noise in KGs and eventuality relations' complexity, impacting relationship modeling [31]. Addressing these limitations is crucial for advancing LLM capabilities in NLP and ensuring effective application across diverse domains.

4 Knowledge Graphs and Knowledge Graph Completion

The exploration of Knowledge Graphs (KGs) requires a detailed understanding of their structure and purpose, which are fundamental to their applications across various domains. KGs encode complex information through interconnected entities and relationships, enhancing the reasoning capabilities of artificial intelligence systems. Table 1 presents a detailed summary of the methods and

Category	Feature	Method
Structure and Purpose of Knowledge Graphs	Ontology and Integration	E-KELL[11]
Methods for Knowledge Graph Completion	Textual and Language Integration	BAKG[14]
Challenges in Knowledge Graph Completion	Scalability and Efficiency Textual Enhancement Type and Constraint Integration Interpretability and Transparency	SMORE[68], HSNMFk-SPLIT[13] TEE[69] TCRL[70] GRank[71], RBE-KGE[72], RDF-KGVS[73]
Recent Advancements in Knowledge Graph Completion	Text-Enhanced Models Matrix Factorization Techniques	SKGC[74], TE-KGE[75] PIE[76]

Table 1: This table provides a comprehensive overview of the current methodologies and advancements in Knowledge Graph Completion (KGC). It categorizes the methods based on their focus areas, such as structure and purpose, challenges, and recent advancements, along with the specific techniques employed in each category. The table serves as a reference for understanding the diverse approaches and their applications in enhancing the utility and comprehensiveness of Knowledge Graphs.

advancements in Knowledge Graph Completion, highlighting the various categories and techniques that contribute to the development and application of Knowledge Graphs. Additionally, Table 3 offers a comprehensive comparison of various methods employed in Knowledge Graph Completion, elucidating their distinct features and application contexts. This section delves into the structure and purpose of KGs, highlighting their role in effective information encoding and retrieval.

4.1 Structure and Purpose of Knowledge Graphs

Knowledge Graphs (KGs) are advanced data structures that represent information as a network of nodes (entities) and edges (relationships), forming a multi-relational directed labeled graph. This format encodes real-world facts in a machine-readable way, supporting applications like entity-oriented search, question-answering systems, and semantic parsing [73]. The integration of KGs with large language models (LLMs) enhances semantic understanding, as seen in Knowledge Graph Prompting (KGP), which constructs graphs with nodes representing document structures and edges denoting semantic relationships [11].

KGs are constructed using various methodologies, including their integration into speech recognition for improved error detection and correction through entity relationship leveraging [14]. ONO-KG, designed for cancer biomarker discovery, exemplifies the adaptability of KGs in representing complex biomedical information [32]. Beyond data representation, KGs enhance AI reasoning by facilitating efficient inference of new facts and integrating structured and unstructured information, crucial for applications like question answering, entity linking, and recommendation systems [36, 3]. For instance, the CoCoLM model employs the ASER knowledge graph for training, showcasing how KGs can enrich commonsense reasoning. The ReasoningLM framework further emphasizes KGs' role in improving semantic comprehension and language generation through multi-hop reasoning.

In specialized contexts, KGs encapsulate intricate information, such as causal relationships among medical concepts identified by LLMs, enhancing data interpretability and decision-making in fields like healthcare. Techniques like entity co-occurrence statistics and semantic query processing enable KGs to provide timely insights and optimize complex queries, enriching knowledge utilization in specialized domains [44, 36, 21, 3]. Frameworks like SPEDN further illustrate KGs' role in constructing semantic query graphs from natural language questions, underscoring their significance in structured information extraction.

KGs are essential for organizing and reasoning about complex information, offering a structured, interconnected representation that enhances intelligent systems' capabilities across various domains. In academia, KGs facilitate entity and relationship extraction from extensive scientific literature, addressing challenges posed by the volume of published research. Advancements in entity type recognition and KG extension improve the quality and interoperability of these graphs, making them valuable tools for AI applications such as semantic search engines and question-answering systems [77, 78, 34, 55, 33]. The continued development and integration of KGs with advanced technologies promise to enhance AI systems' understanding and interaction with the world.

Method Name	Methodologies Used	Integration Techniques	Application Domains
TE-KGE[75]	Textual Regularization	Textual Information Integration	Knowledge Graph Completion
GRank[71]	Graph Patterns	-	Link Prediction
PIE[76]	Tensor Decomposition Methods	Self-supervised Learning	Link Prediction Benchmarks
SKGC[74]	Bi-encoder Architecture	Language Model	Academic Research
BAKG[14]	Multilingual Llms	Dictionary-based Relation	Information Processing
RBE-KGE[72]	Horn Rules	Surrogate Model	Link Prediction

Table 2: This table provides a comprehensive overview of various methodologies and integration techniques employed in knowledge graph completion (KGC) methods. It highlights the specific application domains where these methods are applied, showcasing their diverse utility in enhancing knowledge graph accuracy and robustness.

4.2 Methods for Knowledge Graph Completion

Knowledge Graph Completion (KGC) addresses the incompleteness of KGs by inferring missing entities or relations, enhancing their utility and comprehensiveness. Various methodologies bolster the robustness and accuracy of KGs in representing complex information. Hierarchical and semantic non-negative matrix factorization extract structured ontologies from unstructured scientific literature, particularly in cybersecurity. Integrating LLMs with Scholarly Knowledge Graphs facilitates fine-grained document representation and optimized query processing, significantly improving research artifact retrieval. Cross-graph representation learning uses external, human-curated KGs to detect and correct errors in automatically constructed KGs, ensuring higher precision while accommodating new, domain-specific facts [79, 21, 13]. Table 2 presents an organized summary of different methods used in knowledge graph completion, detailing the methodologies, integration techniques, and application domains associated with each approach.

A prominent approach to KGC employs knowledge graph embeddings (KGEs), which map entities and relationships into continuous vector spaces, capturing semantic and structural properties to predict missing links based on geometric properties. Techniques incorporating textual information through regularization improve link prediction accuracy [75].

Neural network architectures significantly contribute to KGC, with models like GRank utilizing graph pattern matching for entity ranking in link prediction [71]. The PIE framework addresses inefficiencies in parameterization and inference in KGE methods, particularly with large-scale KGs and diverse entity distributions [76].

Moreover, integrating language models with KGs has shown promise in enhancing KGC tasks. The SimKGC method employs contrastive learning with multiple negative samples to improve entity representation [74]. The BanglaAutoKG framework exemplifies the use of multilingual LLMs for entity extraction and graph neural networks for semantic filtering, representing a novel method for knowledge graph completion in underrepresented languages [14].

Innovative frameworks like PHALM combine crowdsourced small-scale KGs with LLMs to generate larger-scale KGs, highlighting the importance of leveraging various data sources for KGC [32]. Rule-based approaches have also been explored to enhance the quality of explanations for embedding-based link predictors in KGs, improving the interpretability of the completion process [72].

These methods enhance the comprehensiveness and reliability of KGs, facilitating more effective knowledge representation and reasoning. They improve interpretability by addressing the semantics of individual dimensions in KG embeddings, validate facts using external human-curated KGs to detect errors, and optimize query processing through LLM integration with structured KGs. Collectively, these advancements support a nuanced understanding and utilization of knowledge, essential for AI applications such as academic research, information retrieval, and automated reasoning [80, 79, 8, 21, 3]. By leveraging advanced methodologies and addressing existing limitations, KGC plays an indispensable role in constructing organized and comprehensive knowledge structures, enhancing intelligent systems' capabilities in processing and reasoning with complex datasets.

4.3 Challenges in Knowledge Graph Completion

Knowledge Graph Completion (KGC) faces challenges due to the complexities of knowledge graphs and the limitations of current methodologies. A major challenge is the scalability and complexity of multi-hop reasoning tasks, which require traversing multiple relations within a KG, complicating the effective materialization of training instances [68]. The incompleteness of type constraints in KGs further complicates the task, as these constraints are vital for ensuring accurate predictions of entities and relations [70].

The unseen-entities problem, where entities in the test set do not appear in the training data, leads to ineffective scoring by traditional methods [69]. This issue is exacerbated by the black-box nature of KG embedding models, which produce outputs without clear explanations, complicating the understanding of their predictions [71]. Furthermore, redundancy in parameterization, due to uniform embedding dimensions for all entities, and high inference time complexity when querying large KGs present substantial obstacles to efficient KGC [76].

The high computational cost associated with contrastive learning in text-based KGC limits the effective use of larger negative sample sizes, restricting improvements in prediction accuracy [74]. Existing methods often rely predominantly on intrinsic KG structure, limiting their ability to incorporate external information for enhanced prediction accuracy [75]. Additionally, the increasing size and complexity of KGs pose visualization challenges, potentially overwhelming users [73].

Recent advancements have attempted to address these challenges through innovative approaches. For instance, the RBE-KGE method generates explanations for link predictions by mining rules with bounded atoms learned from localized contexts, enhancing KGC model interpretability [72]. Approaches in cybersecurity knowledge graph generation demonstrate the ability to automatically determine topic numbers and extract coherent, domain-specific information from large corpora, thereby enhancing KGC scalability and adaptability [13].

Despite these advancements, challenges remain in enhancing interpretability and efficiency in KGC models. Existing benchmarks often rely on random sampling of entities, which may overestimate KGC performance due to easy negative candidates that do not accurately reflect true performance [81]. The focus on single-language KGs in existing benchmarks leads to suboptimal performance due to a lack of cross-lingual information sharing and ID proliferation issues [35]. Addressing these challenges is crucial for advancing the field of KGC and unlocking the full potential of KGs for a wide range of AI applications.

4.4 Recent Advancements in Knowledge Graph Completion

Recent advancements in Knowledge Graph Completion (KGC) have focused on enhancing the accuracy, efficiency, and interpretability of predicting missing entities and relations within KGs. A notable development is the Parameter and Inference Efficient (PIE) solution, which combines low-rank decomposition of entity embeddings with a fine-grained entity typing approach, significantly improving parameter efficiency and inference speed in large-scale KGC tasks [76].

In text-based KGC, SimKGC has emerged as a leading framework, demonstrating substantial improvements over existing state-of-the-art methods by effectively addressing inefficiencies in contrastive learning, thus boosting performance in link prediction tasks [74]. Experiments with the FB60K dataset have highlighted the efficacy of textual regularization in enhancing methods like TransE and its variants, underscoring the importance of incorporating textual information in KGC [75].

Moreover, the development of efficient benchmarking frameworks has been pivotal in advancing KGC methodologies. The benchmark framework proposed by Cornell et al. allows for fast and accurate evaluation of KGC models, significantly improving performance estimation speed and reliability, particularly in large-scale settings [81].

There has also been a growing focus on multilingual KGC, exemplified by the ALIGNKGC framework, which has shown significant improvements in KGC accuracy and alignment scores compared to baseline models, highlighting the effectiveness of joint training in enhancing multilingual KGC performance [35].

These advancements reflect concerted efforts to refine KGC methodologies, improve interpretability, and enhance scalability. The integration of innovative techniques such as KGExplainer, which utilizes connected subgraph explanations, and multi-perspective frameworks like MPIKGC, leveraging LLMs for enhanced entity descriptions, is expected to improve the transparency and interpretability of KGC tasks. This progress will enhance the quality of explanations and predictions, broadening the applicability of KGs across diverse domains and facilitating more accountable AI applications in areas like web search, recommendation systems, and natural language processing [82, 83, 84, 52, 3].

Feature	Hierarchical and Semantic Non-Negative Matrix Factorization	Cross-Graph Representation Learning	Knowledge Graph Embeddings (KGEs)
Optimization Technique	Matrix Factorization	Representation Learning	Vector Space Mapping
Application Domain	Cybersecurity	General	General
Integration Method	Structured Ontologies	External Kgs	Geometric Properties

Table 3: This table provides a comparative analysis of three methodologies for Knowledge Graph Completion: Hierarchical and Semantic Non-Negative Matrix Factorization, Cross-Graph Representation Learning, and Knowledge Graph Embeddings (KGEs). It highlights key features such as optimization techniques, application domains, and integration methods, offering insights into their respective strengths and applicability within the field of Knowledge Graphs.

5 Semantic Reasoning and Graph Embeddings

5.1 Concept and Role of Semantic Reasoning

Semantic reasoning is pivotal in AI, enabling the extraction of meaning from complex datasets through logical rules and external knowledge sources like knowledge graphs (KGs). This process enhances AI systems' interpretability and reliability by elucidating intricate relationships among entities and concepts. The Semantically Enhanced Knowledge Graph Embedding (SE-KGE) method exemplifies this integration, improving commonsense knowledge acquisition by embedding semantic resources within the knowledge graph framework [85]. The SPEDN framework further highlights the importance of semantic reasoning by effectively parsing complex natural language questions, demonstrating proficiency in handling intricate queries that necessitate a deep understanding and contextual interpretation [86]. Additionally, the SMORE system illustrates the efficiency of semantic reasoning in knowledge graph completion tasks by leveraging multi-GPU systems to dynamically generate training data [68].

In dialogue systems, benchmarks proposed by Schneider et al. provide structured evaluation methods for conversational grounding, enhancing semantic reasoning capabilities [16]. The integration of pre-trained language models with temporal and relational information optimizes the use of semantic reasoning in tasks like temporal knowledge graph completion [40]. Semantic reasoning is also critical in applications such as cancer diagnosis, where real-time updates to knowledge graphs are vital for accurate information extraction and decision-making [10]. Knowledge Graph-Enhanced Pre-trained Language Models (KGPLMs) exhibit improved performance in natural language processing tasks by leveraging structured knowledge, thereby enhancing factual accuracy and reasoning capabilities [12].

Visualization systems described by Kerdjoudj et al. enhance user interaction with semantic reasoning by providing clearer views of extracted knowledge, supporting more effective decision-making and knowledge interpretation [73]. Thus, semantic reasoning is essential for advancing AI systems' capabilities in processing, interpreting, and reasoning with complex datasets, fostering the development of more sophisticated and reliable intelligent systems.

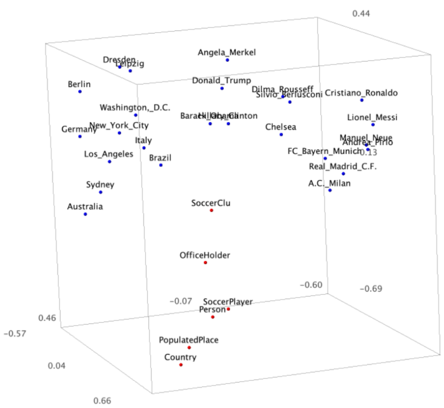
5.2 Graph Embeddings: Transformation and Applications

Graph embeddings are crucial in transforming complex graph data into formats suitable for analysis and diverse applications in AI. By converting nodes, edges, and their relationships into continuous vector spaces, graph embeddings facilitate the integration of graph-structured data with machine learning algorithms, enabling tasks such as anomaly detection and knowledge representation [87]. Methods like LineaRE preserve knowledge graph semantics through linear mappings that model complex relationships [88].

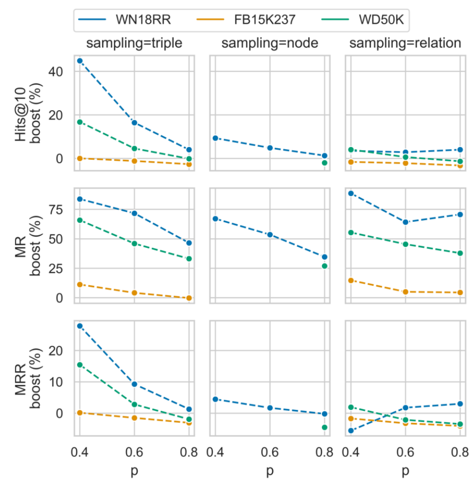
Generating graph embeddings involves capturing both semantic and structural information. The Joint Language and Semantic Structure (LASS) framework demonstrates this by obtaining semantic embeddings through language models while reconstructing structural information via a structured loss function [89]. This dual approach ensures that embeddings accurately represent entities and relationships while retaining structural integrity for precise data analysis.

Advanced applications of graph embeddings involve modeling interactions within knowledge graphs through complex mathematical operations. Tran et al. propose representing each entity and relation with multiple embedding vectors and computing interaction scores using trilinear products, combined through a weight vector to enhance representation richness and accuracy [55]. This method is particularly effective in applications demanding nuanced understanding and reasoning based on both textual and graph data.

Graph embeddings are essential for converting complex graph data into structured formats, allowing AI systems to efficiently process, reason, and derive insights from intricate datasets. Representing entities and relationships as vectors in a semantic space facilitates knowledge inference, semantic search, and recommendation systems. Various knowledge graph embedding methods—such as DistMult and ComplEx—enhance interpretability and effectiveness, improving the performance of AI applications in knowledge-intensive domains [90, 55, 3]. The continued development and integration of contextual and multimodal embeddings promise to advance the capabilities of intelligent systems in understanding and interacting with the world.



(a) The image represents a 3D scatter plot with a set of points in a cube-shaped space.[51]



(b) Effect of Sampling Strategy on Hits@10, MR, and MRR Boost in Graph Neural Networks[90]

Figure 4: Examples of Graph Embeddings: Transformation and Applications

As shown in Figure 4, the exploration of semantic reasoning and graph embeddings illustrates their transformation and applications through visual representations. The first image depicts a 3D scatter plot with points marked by blue dots and labeled with names such as "Angela$_M erkel,""Donald_T rump,"and"Cristiano_R onaldo,"exemplifyinghowentitiesandrelationshipscanberepres$
$axisdenotingthehyperparameterp.Collectively,theseimagesemphasizethesignificanceofgraphembeddingsinenhan$
90]

5.3 Integration of Contextual and Knowledge Graph Embeddings

The integration of contextual and knowledge graph embeddings significantly enhances AI capabilities, particularly for tasks requiring a deep understanding of both textual and structural information. By combining the strengths of contextual embeddings, which capture dynamic semantics of language, with knowledge graph embeddings that encode structured relationships among entities, AI systems can achieve a more comprehensive representation of information [2].

A key advantage of this integration is its ability to effectively represent long-tail entities, often underrepresented in large datasets. The Language Model Knowledge Embeddings (LMKE) framework exemplifies this by leveraging both structural and textual information to provide rich representations of these entities, thereby enhancing AI performance in tasks like information extraction and recommendation [91]. This dual approach ensures embeddings capture semantic nuances while maintaining structural integrity for accurate data analysis.

The incorporation of logics within these embeddings further enhances their applicability in areas such as recommender systems, where understanding intricate relationships between user preferences and item characteristics is crucial [2]. By embedding logical rules within the knowledge graph framework, AI systems can perform more sophisticated reasoning tasks, leading to improved accuracy and personalization in recommendations.

AI-generated, for reference only.

The integration of contextual and knowledge graph embeddings holds considerable promise for advancing AI capabilities across various domains. By offering a unified representation of information through integrated embeddings, AI systems can more effectively process and reason with complex datasets. This advancement enhances their ability to perform tasks such as link prediction and entity resolution, facilitating the development of more intelligent and context-aware applications. Leveraging hypercomplex algebra to integrate structural knowledge from knowledge graphs with textual information from language models enables a deeper understanding of the interplay between different modalities, ultimately improving performance across various applications [61, 92].

6 Interconnections and Synergies

6.1 Integration of Large Language Models and Knowledge Graphs

The fusion of large language models (LLMs) with knowledge graphs (KGs) significantly advances AI by enhancing reasoning and contextual understanding. This integration combines the generative prowess of LLMs with the structured semantic depth of KGs, fostering advanced AI applications across various domains. The ReasoningLM model exemplifies this synergy by boosting performance in knowledge graph question answering (KGQA), highlighting the potential of LLMs and KGs in enhancing reasoning and comprehension [17]. In emergency management, frameworks like E-KELL utilize LLMs and KGs to enhance decision-making accuracy and interpretability, crucial in time-sensitive scenarios [11]. Similarly, BanglaAutoKG demonstrates the efficacy of multilingual LLMs in constructing KGs from Bangla text, improving the processing of underrepresented languages [14].

Projects such as PHALM integrate human contextual knowledge with LLM scalability, creating high-quality KGs that enrich AI performance by providing structured data for reasoning tasks [32]. This approach underscores the importance of merging human insights with automated processes to produce comprehensive knowledge representations. Overall, integrating LLMs and KGs establishes a robust framework that enhances reasoning abilities, representation of complex information, and applications in fields like academic research and knowledge management. This synergy improves query processing, knowledge extraction, and semi-automated KG construction, mitigating information hallucinations and empowering users to engage more effectively with knowledge [18, 19, 20, 93, 21].

6.2 Enhancing Reasoning and Representation

Integrating LLMs with KGs significantly boosts AI systems' reasoning and representation capabilities. This synergy leverages LLMs' generative and contextual understanding with KGs' structured knowledge. Frameworks like KG-RAG enable LLMs to access KGs, reducing information hallucinations and enhancing accuracy in knowledge-intensive tasks. Systems like the ANU Scholarly Knowledge Graph facilitate nuanced semantic query processing, enhancing knowledge utilization and natural language understanding [19, 21, 18]. This integration enhances the ability to perform complex reasoning tasks, synthesizing structured data such as scholarly metadata with unstructured data from research papers, optimizing query processing for accurate information retrieval [33, 21].

Moreover, integrating semantic reasoning with graph embeddings enhances AI systems' representation capabilities. By embedding logical rules and external knowledge, these systems can perform sophisticated reasoning tasks, resulting in more accurate and personalized applications. The ability to dynamically generate and update KGs in real-time, as seen in critical applications like cancer diagnosis, underscores the significance of this integration in enhancing decision-making processes [10]. This advancement promotes sophisticated, context-aware applications by leveraging cutting-edge NLP and machine learning techniques, streamlining the management and analysis of extensive scientific literature, and driving AI technologies' progression across various domains [77, 44, 94, 33, 21].

6.3 Applications in Knowledge Graph Construction and Completion

The integration of LLMs with KGs has significantly advanced knowledge graph construction and completion, offering transformative applications across various domains. In language modeling, knowledge representation learning (KRL) enhances AI systems' semantic understanding and generation capabilities. By utilizing the rich structural information in KGs, LLMs improve language modeling tasks, enhancing the accuracy and contextual relevance of generated text [80]. In question answering, the synergy between LLMs and KGs enhances systems' capabilities to comprehend and

Benchmark	Size	Domain	Task Format	Metric
FB-Test-S-C[95]	22,492	Knowledge Graph Completion	Entity Ranking	MRR, Hits@K
KGR[49]	500	Criminal Investigation	Criminal Identification	Correctness, Explainability
BMKGQA[96]	201,000	Biomedical Question Answering	Multi-hop Question Answering	Hits@10
LLM-KG-Bench[97]	5,000	Knowledge Graphs	Sparql Select Query Generation	F1-score, accuracy
KHGE[92]	1,000	Knowledge Graphs	Link Prediction	MRR, Hits@N
KGC-LLM[98]	150	Knowledge Graphs	Knowledge Graph Completion	Accuracy, F1
RPLLM[99]	483,142	Knowledge Graph Completion	Relation Prediction	Mean Rank, Hits@1
KG-HRD[100]	121	Human Rights Violations	Knowledge Graph Construction	Precision, Recall

Table 4: Table showcasing various benchmarks used in the evaluation of knowledge graph construction and completion tasks. The benchmarks cover diverse domains such as knowledge graph completion, criminal investigation, and biomedical question answering, with task formats including entity ranking, multi-hop question answering, and link prediction. Metrics for evaluation include MRR, Hits@K, accuracy, and F1-score, reflecting the comprehensive assessment of systems integrating large language models with knowledge graphs.

respond to complex queries. Advanced frameworks, such as the KG-enhanced Query Processing system, improve accuracy and efficiency in retrieving information from structured repositories like the ANU Scholarly Knowledge Graph [19, 21, 24]. Table 4 provides a detailed overview of the benchmarks utilized in the study of knowledge graph construction and completion applications, highlighting the diversity of domains, task formats, and evaluation metrics.

KRL extends to information retrieval and recommendation systems, where LLM-KG integration enhances the processing and analysis of large data volumes. Leveraging structured knowledge within KGs enables these systems to provide personalized and relevant recommendations, improving user experience [80]. Moreover, multilingual knowledge graph completion techniques, such as KEnS, demonstrate the potential of cross-lingual knowledge to enhance KG completion performance. By effectively utilizing information from multiple languages, these techniques improve KG comprehensiveness and accuracy, supporting applications requiring cross-lingual information sharing [101].

The integration of LLMs and KGs in knowledge graph construction and completion applications underscores these technologies' transformative potential in advancing AI systems. By employing advanced NLP and machine learning techniques to enhance processing, interpretation, and reasoning with complex datasets, this approach fosters the development of intelligent, context-aware applications across various fields, including academia. This integration facilitates comprehensive knowledge graphs that effectively represent entities and relationships in scientific literature, enabling researchers and policymakers to analyze and manage vast data efficiently [33, 21, 44].

6.4 Improving NLP and Information Retrieval

The integration of LLMs with KGs significantly enhances natural language processing (NLP) and information retrieval by leveraging structured and unstructured data. This synergy enables more accurate semantic parsing and contextual understanding, crucial for converting natural language into structured queries, thus improving interaction with knowledge graphs [43]. In NLP, incorporating KGs into LLMs facilitates knowledge-grounded content generation, enhancing models' ability to produce contextually rich and semantically coherent text [6]. This integration is particularly beneficial for machine translation and sentiment analysis, where nuanced meanings in context are essential [28].

In information retrieval, LLM-KG integration improves search result accuracy and relevance by providing structured context for precise information extraction and reasoning [44]. This is particularly evident in conversational recommender systems, where inferring user preferences from dialogue history is enhanced by the structured knowledge provided by KGs [5]. Moreover, integrating contextual embeddings with knowledge graph embeddings allows for richer representations of words and their relationships, facilitating more effective information retrieval and reasoning [53]. This approach enhances AI systems' semantic understanding, enabling them to process and interpret complex datasets more effectively.

The integration of LLMs with KGs significantly boosts NLP and information retrieval capabilities. This synergy fosters sophisticated, context-aware AI applications by improving query accuracy and efficiency, as demonstrated by frameworks like KG-RAG and the semi-automated construction of KGs using LLMs. These advancements enhance knowledge utilization, reduce reliance on latent knowledge, and improve reasoning abilities, revolutionizing scholarly knowledge management and enabling the development of intelligent systems adept at handling complex, knowledge-intensive tasks [18, 19, 20, 41, 21]. This synergy not only enhances the semantic understanding of language models but also extends their utility across various domains, ultimately improving the quality and relevance of information retrieval processes.

7 Challenges and Future Directions

Advancing artificial intelligence requires tackling the intricate challenges associated with integrating large language models (LLMs) and knowledge graphs (KGs). This section delves into the critical factors affecting AI systems' efficacy and reliability, focusing on data quality, sparsity, scalability, computational efficiency, interpretability, explainability, and the integration of heterogeneous information.

7.1 Data Quality and Sparsity

Data quality and sparsity are pivotal in the integration and performance of LLMs and KGs. The accuracy and completeness of KGs are crucial, as inaccuracies can lead to erroneous outputs, particularly when retrieval processes lack precision [11]. The abstract nature of commonsense knowledge complicates inferring missing information, posing challenges for models like CoCoLM, which excel in multi-relational inference but struggle with complex commonsense knowledge [31].

These challenges are compounded by the need for high-quality datasets and significant computational resources for training large models [40]. Incomplete type constraints can impair data representation and reasoning [70]. Variability in model performance based on entity complexity further underscores the need for enhanced data quality in KGs [69].

Benchmarks often inadequately assess models' abilities to learn logical rules, potentially leading to misleading evaluations [102]. The SMORE system may falter with complex queries or large graphs beyond computational limits [68], and current models' focus on smaller graph patterns restricts the complexity of relationships they can capture [71].

Scalability challenges arise from increased complexity and noise when mining longer rules [72]. Data quality and computational challenges are noted limitations in cybersecurity knowledge graph generation [13], and visualization of large RDF knowledge graphs suffers from data quality and sparsity issues [73].

Addressing these challenges requires robust data management practices and scalable solutions to enhance semantic query processing's accuracy, efficiency, and interpretability. This integration will improve factual knowledge retrieval and facilitate developing advanced frameworks leveraging LLMs and KGs, transforming knowledge management and discovery across domains, including academic research and education [5, 21, 22].

7.2 Scalability and Computational Efficiency

Scalability and computational efficiency are crucial for deploying LLMs and KGs, especially as these systems grow in size and complexity. Despite the potential of LLM-KG integration to enhance AI capabilities, scalability remains a significant challenge. Integrating probabilistic logic with neural networks may struggle with large KGs or intricate relational structures, impacting scalability and computational feasibility [103].

Generating embeddings for large datasets is essential for scalability. KG2Vec demonstrates notable speed and efficiency in embedding generation compared to existing methods, providing a valuable tool for large-scale knowledge graphs [51]. However, as KG size increases, challenges related to retrieval latency and computational efficiency may arise, affecting real-time performance in applications requiring rapid information retrieval [24].

Adapting methods to diverse document types beyond academic papers presents additional scalability challenges. While leveraging LLMs offers substantial benefits, proposed methods may struggle to scale effectively across varied document types, necessitating further research to enhance adaptability and scalability in diverse contexts [21].

Developing more efficient algorithms and scalable frameworks capable of managing the computational demands of large-scale LLM and KG integration is essential. Prioritizing computational efficiency and scalability will enable AI systems to adeptly navigate real-world application complexities, enhancing performance and utility across diverse domains. Integrating LLMs with structured KGs allows for more accurate and efficient information retrieval, as demonstrated in scholarly research contexts. Employing advanced Natural Language Processing (NLP) techniques within these frameworks facilitates effective analysis of vast scientific literature, ultimately supporting improved decision-making and knowledge management [33, 94, 62, 21].

7.3 Interpretability and Explainability

Interpretability and explainability are vital for AI systems, especially in integrating LLMs with KGs. These components ensure transparency and trustworthiness, allowing users to understand decision-making processes and underlying mechanisms. The UAG framework exemplifies efforts to enhance interpretability through rigorous uncertainty quantification, crucial for reliable predictions in high-stakes scenarios [60]. The SMORE system highlights potential research avenues to optimize sampling techniques and integrate complex reasoning capabilities, improving knowledge graph completion model interpretability [68].

Models like SGMPT exhibit higher complexity and running time than baseline models due to additional structural fusion mechanisms, illustrating the trade-off between interpretability and computational efficiency [54]. The PIE framework emphasizes reducing parameters in knowledge graph embedding (KGE) models, enhancing interpretability by simplifying architecture without sacrificing performance [76].

For large-scale AI systems, interpretability and explainability are crucial, particularly in integrating question understanding and graph reasoning [17]. Approaches that provide relevant interpretations of predictions made by embedding-based models enhance user trust and transparency [72].

Visualization systems, as described by Kerdjoudj et al., significantly improve user understanding of complex RDF knowledge graphs, underscoring the importance of interpretability in AI systems [73]. However, current knowledge graph-enhanced pre-trained language models (KGPLMs) may require additional computational resources and time for training, posing challenges for widespread adoption and interpretability [12].

Future work should focus on enhancing interpretability and managing false negatives, critical for explainability in AI systems [74]. Despite filtering processes, methods discussed by Ide et al. may still yield incorrect knowledge, leading to potentially harmful applications [32]. Addressing these challenges is vital for developing effective, interpretable, and explainable AI systems, fostering user trust and transparency.

7.4 Integration of Heterogeneous Information

Integrating heterogeneous information sources into LLMs and KGs presents challenges and opportunities for enhancing AI systems' expressivity and utility. A primary challenge is the semantic alignment of diverse data types into KG embeddings, necessitating effective methodologies for harmonizing these varied sources [31]. Incorporating additional semantics, such as contextual descriptions of entities into dynamic KGs, complicates the task, requiring innovative approaches for automating prompt generation and integrating node semantics [40].

Future research should explore diverse forms of semantic evidence and refine modeling techniques to enhance extrapolation capabilities [70]. Integrating geometrical spaces or hybrid models combining hyperbolic and Euclidean approaches could further improve knowledge base completion capabilities [76]. Enhancing scalability, implementing advanced error-handling mechanisms, and accommodating complex domain-specific knowledge are critical areas for future improvement [81].

Opportunities include developing tools and frameworks that facilitate the adoption of semantic units in various KG applications, addressing interoperability challenges with existing systems [75]. Further research could refine HAKE's framework for better performance on complex datasets and integrate additional data sources for enhanced hierarchical modeling [71]. Developing more effective models and exploring adaptive integration of incremental learning into the PKGC framework are promising directions [35].

In historical data integration, benchmarks supporting interdisciplinary research by merging fragmented historical information highlight challenges and opportunities. Future work could enhance rule mining processes and explore advanced models for better integration of rules and embeddings, improving KG construction methods [14]. Optimizing context integration processes, investigating hyperparameter effects, and extending CaQR applicability to other reasoning models are also of interest [86].

Improving training data for knowledge graph completion and reasoning tasks is essential for advancing heterogeneous information integration [81]. Future research will focus on incorporating larger ontologies, addressing bias and fairness in KG generation, and enhancing reasoning capabilities in LLMs [75]. Exploring DualE applications in different domains and enhancing computational efficiency are also promising directions [76].

By addressing these challenges and leveraging opportunities, future research can drive the development of sophisticated and context-aware AI applications, improving the capacity of AI systems to process and reason with complex datasets. Developing robust frameworks for real-time integration of KGs with LLMs will be crucial for advancing the field [35]. Future work will also focus on enhancing prompting strategies, exploring closed-domain applications, and comparing proposed methods with existing state-of-the-art tools, alongside addressing the integration challenge of heterogeneous information sources in G2T generation [14].

8 Conclusion

The integration of large language models (LLMs) with knowledge graphs (KGs) marks a pivotal evolution in artificial intelligence, enhancing both reasoning and representation capabilities across diverse applications. This survey underscores the synergistic potential of these technologies to transform AI systems by providing structured context and reducing hallucination rates. The ability of LLMs to generate citation text exemplifies the improvements achieved through the incorporation of knowledge graphs. Despite this progress, significant technical challenges remain, particularly in areas such as scalability, data quality, and the integration of diverse information sources. The enhancement of knowledge graph embeddings with semantic information has proven to be advantageous for commonsense knowledge acquisition, highlighting the ongoing need for research in this area. The development of domain-specific knowledge graphs, such as those for cybersecurity, further demonstrates the practical benefits of extracting valuable insights from large datasets. Future research directions should focus on integrating advanced semantic knowledge techniques, improving model resilience to noise and errors, and exploring the capabilities of generative models for knowledge graph completion. The proposed framework for academic knowledge graphs significantly advances scholarly knowledge management, showcasing the transformative potential of these integrations. Additionally, evaluating knowledge graph extension frameworks highlights their practicality and effectiveness, indicating the importance of continued research to enhance these methodologies.

References

[1] Dat Quoc Nguyen. A survey of embedding models of entities and relationships for knowledge graph completion, 2020.

[2] Wen Zhang, Jiaoyan Chen, Juan Li, Zezhong Xu, Jeff Z. Pan, and Huajun Chen. Knowledge graph reasoning with logics and embeddings: Survey and perspective, 2022.

[3] Chandrahas, Tathagata Sengupta, Cibi Pragadeesh, and Partha Pratim Talukdar. Inducing interpretability in knowledge graph embeddings, 2017.

[4] Hiroaki Hayashi, Zecong Hu, Chenyan Xiong, and Graham Neubig. Latent relation language models, 2019.

[5] Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[6] Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. Comet: Commonsense transformers for automatic knowledge graph construction, 2019.

[7] Gabriel R. Rosenbaum, Lavender Yao Jiang, Ivaxi Sheth, Jaden Stryker, Anton Alyakin, Daniel Alexander Alber, Nicolas K. Goff, Young Joon Fred Kwon, John Markert, Mustafa Nasir-Moin, Jan Moritz Niehues, Karl L. Sangwon, Eunice Yang, and Eric Karl Oermann. Medg-krp: Medical graph knowledge representation probing, 2024.

[8] Avinash Anand, Mohit Gupta, Kritarth Prasad, Ujjwal Goel, Naman Lal, Astha Verma, and Rajiv Ratn Shah. Kg-ctg: Citation generation through knowledge graph-guided large language models, 2024.

[9] Qingyao Li, Wei Xia, Kounianhua Du, Qiji Zhang, Weinan Zhang, Ruiming Tang, and Yong Yu. Learning structure and knowledge aware representation with large language models for concept recommendation, 2024.

[10] Md. Rezaul Karim, Lina Molinas Comet, Md Shajalal, Oya Deniz Beyan, Dietrich Rebholz-Schuhmann, and Stefan Decker. From large language models to knowledge graphs for biomarker discovery in cancer, 2023.

[11] Minze Chen, Zhenxiang Tao, Weitong Tang, Tingxin Qin, Rui Yang, and Chunli Zhu. Enhancing emergency decision-making with knowledge graphs and large language models, 2023.

[12] 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, 2024.

[13] Ryan Barron, Maksim E. Eren, Manish Bhattarai, Selma Wanna, Nicholas Solovyev, Kim Rasmussen, Boian S. Alexandrov, Charles Nicholas, and Cynthia Matuszek. Cyber-security knowledge graph generation by hierarchical nonnegative matrix factorization, 2024.

[14] Azmine Toushik Wasi, Taki Hasan Rafi, Raima Islam, and Dong-Kyu Chae. Banglaautokg: Automatic bangla knowledge graph construction with semantic neural graph filtering, 2024.

[15] Ashwini Jaya Kumar, Sören Auer, Christoph Schmidt, and Joachim köhler. Towards a knowledge graph based speech interface, 2017.

[16] Phillip Schneider, Nektarios Machner, Kristiina Jokinen, and Florian Matthes. Bridging information gaps in dialogues with grounded exchanges using knowledge graphs, 2024.

[17] Jinhao Jiang, Kun Zhou, Wayne Xin Zhao, Yaliang Li, and Ji-Rong Wen. Reasoninglm: Enabling structural subgraph reasoning in pre-trained language models for question answering over knowledge graph, 2023.

[18] 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, 2024.

[19] Diego Sanmartin. Kg-rag: Bridging the gap between knowledge and creativity, 2024.

[20] Vamsi Krishna Kommineni, Birgitta König-Ries, and Sheeba Samuel. From human experts to machines: An llm supported approach to ontology and knowledge graph construction, 2024.

[21] Runsong Jia, Bowen Zhang, Sergio J. Rodríguez Méndez, and Pouya G. Omran. Leveraging large language models for semantic query processing in a scholarly knowledge graph, 2024.

[22] Tuan Bui, Oanh Tran, Phuong Nguyen, Bao Ho, Long Nguyen, Thang Bui, and Tho Quan. Cross-data knowledge graph construction for llm-enabled educational question-answering system: A case study at hcmut, 2024.

[23] Sakher Khalil Alqaaidi and Krzysztof Kochut. Knowledge graph completion using structural and textual embeddings, 2024.

[24] Yu Wang, Nedim Lipka, Ryan A. Rossi, Alexa Siu, Ruiyi Zhang, and Tyler Derr. Knowledge graph prompting for multi-document question answering, 2023.

[25] Fedor Moiseev, Zhe Dong, Enrique Alfonseca, and Martin Jaggi. Skill: Structured knowledge infusion for large language models, 2022.

[26] Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. K-bert: Enabling language representation with knowledge graph, 2019.

[27] Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. K-bert: Enabling language representation with knowledge graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2901–2908, 2020.

[28] Yinquan Lu, Haonan Lu, Guirong Fu, and Qun Liu. Kelm: Knowledge enhanced pre-trained language representations with message passing on hierarchical relational graphs, 2022.

[29] Jinheon Baek, Alham Fikri Aji, and Amir Saffari. Knowledge-augmented language model prompting for zero-shot knowledge graph question answering, 2023.

[30] Jie Huang and Kevin Chen-Chuan Chang. Towards reasoning in large language models: A survey. *arXiv preprint arXiv:2212.10403*, 2022.

[31] Changlong Yu, Hongming Zhang, Yangqiu Song, and Wilfred Ng. Cocolm: Complex commonsense enhanced language model with discourse relations, 2022.

[32] Tatsuya Ide, Eiki Murata, Daisuke Kawahara, Takato Yamazaki, Shengzhe Li, Kenta Shinzato, and Toshinori Sato. Phalm: Building a knowledge graph from scratch by prompting humans and a language model, 2023.

[33] Danilo Dessì, Francesco Osborne, Diego Reforgiato Recupero, Davide Buscaldi, and Enrico Motta. Generating knowledge graphs by employing natural language processing and machine learning techniques within the scholarly domain, 2020.

[34] Ciyuan Peng, Feng Xia, Mehdi Naseriparsa, and Francesco Osborne. Knowledge graphs: Opportunities and challenges. *Artificial Intelligence Review*, 56(11):13071–13102, 2023.

[35] Harkanwar Singh, Prachi Jain, Mausam, and Soumen Chakrabarti. Multilingual knowledge graph completion with joint relation and entity alignment, 2021.

[36] Genet Asefa Gesese, Russa Biswas, Mehwish Alam, and Harald Sack. A survey on knowledge graph embeddings with literals: Which model links better literal-ly?, 2020.

[37] Yang Liu, Chuan Zhou, Peng Zhang, Yanan Cao, Yongchao Liu, Zhao Li, and Hongyang Chen. Cl4kge: A curriculum learning method for knowledge graph embedding, 2024.

[38] Jiapu Wang, Boyue Wang, Meikang Qiu, Shirui Pan, Bo Xiong, Heng Liu, Linhao Luo, Tengfei Liu, Yongli Hu, Baocai Yin, and Wen Gao. A survey on temporal knowledge graph completion: Taxonomy, progress, and prospects, 2023.

[39] Jiayi Li, Ruilin Luo, Jiaqi Sun, Jing Xiao, and Yujiu Yang. Progressive knowledge graph completion, 2024.

[40] Wenjie Xu, Ben Liu, Miao Peng, Xu Jia, and Min Peng. Pre-trained language model with prompts for temporal knowledge graph completion, 2024.

[41] Linmei Hu, Zeyi Liu, Ziwang Zhao, Lei Hou, Liqiang Nie, and Juanzi Li. A survey of knowledge enhanced pre-trained language models, 2023.

[42] Ziheng Zeng, Kellen Tan Cheng, Srihari Venkat Nanniyur, Jianing Zhou, and Suma Bhat. Iekg: A commonsense knowledge graph for idiomatic expressions, 2023.

[43] Phillip Schneider, Manuel Klettner, Kristiina Jokinen, Elena Simperl, and Florian Matthes. Evaluating large language models in semantic parsing for conversational question answering over knowledge graphs, 2024.

[44] Vivek Khetan, Annervaz K M, Erin Wetherley, Elena Eneva, Shubhashis Sengupta, and Andrew E. Fano. Knowledge graph anchored information-extraction for domain-specific insights, 2021.

[45] Nandana Mihindukulasooriya, Sanju Tiwari, Carlos F. Enguix, and Kusum Lata. Text2kgbench: A benchmark for ontology-driven knowledge graph generation from text, 2023.

[46] Zhangchi Qiu, Ye Tao, Shirui Pan, and Alan Wee-Chung Liew. Knowledge graphs and pre-trained language models enhanced representation learning for conversational recommender systems, 2024.

[47] Linhao Luo, Jiaxin Ju, Bo Xiong, Yuan-Fang Li, Gholamreza Haffari, and Shirui Pan. Chatrule: Mining logical rules with large language models for knowledge graph reasoning, 2024.

[48] Shantanu Jaiswal, Liu Yan, Dongkyu Choi, and Kenneth Kwok. A probabilistic-logic based commonsense representation framework for modelling inferences with multiple antecedents and varying likelihoods, 2022.

[49] Takahiro Kawamura, Shusaku Egami, Koutarou Tamura, Yasunori Hokazono, Takanori Ugai, Yusuke Koyanagi, Fumihito Nishino, Seiji Okajima, Katsuhiko Murakami, Kunihiko Takamatsu, Aoi Sugiura, Shun Shiramatsu, Shawn Zhang, and Kouji Kozaki. Report on the first knowledge graph reasoning challenge 2018 – toward the explainable ai system, 2019.

[50] Han Xiao, Minlie Huang, and Xiaoyan Zhu. Ksr: A semantic representation of knowledge graph within a novel unsupervised paradigm, 2020.

[51] Tommaso Soru, Stefano Ruberto, Diego Moussallem, André Valdestilhas, Alexander Bigerl, Edgard Marx, and Diego Esteves. Expeditious generation of knowledge graph embeddings, 2018.

[52] Satvik Garg and Dwaipayan Roy. A birds eye view on knowledge graph embeddings, software libraries, applications and challenges, 2022.

[53] Lea Dieudonat, Kelvin Han, Phyllicia Leavitt, and Esteban Marquer. Exploring the combination of contextual word embeddings and knowledge graph embeddings, 2020.

[54] Ke Liang, Sihang Zhou, Yue Liu, Lingyuan Meng, Meng Liu, and Xinwang Liu. Structure guided multi-modal pre-trained transformer for knowledge graph reasoning, 2023.

[55] Hung Nghiep Tran and Atsuhiro Takasu. Analyzing knowledge graph embedding methods from a multi-embedding interaction perspective, 2023.

[56] Bin He, Di Zhou, Jinghui Xiao, Xin jiang, Qun Liu, Nicholas Jing Yuan, and Tong Xu. Integrating graph contextualized knowledge into pre-trained language models, 2021.

[57] Jie Ma, Zhitao Gao, Qi Chai, Wangchun Sun, Pinghui Wang, Hongbin Pei, Jing Tao, Lingyun Song, Jun Liu, Chen Zhang, and Lizhen Cui. Debate on graph: a flexible and reliable reasoning framework for large language models, 2024.

[58] Yi-Lin Tuan, Sajjad Beygi, Maryam Fazel-Zarandi, Qiaozi Gao, Alessandra Cervone, and William Yang Wang. Towards large-scale interpretable knowledge graph reasoning for dialogue systems, 2022.

[59] Daehee Kim, Deokhyung Kang, Sangwon Ryu, and Gary Geunbae Lee. Ontology-free general-domain knowledge graph-to-text generation dataset synthesis using large language model, 2024.

[60] Bo Ni, Yu Wang, Lu Cheng, Erik Blasch, and Tyler Derr. Towards trustworthy knowledge graph reasoning: An uncertainty aware perspective, 2024.

[61] Mojtaba Nayyeri, Zihao Wang, Mst. Mahfuja Akter, Mirza Mohtashim Alam, Md Rashad Al Hasan Rony, Jens Lehmann, and Steffen Staab. Integrating knowledge graph embedding and pretrained language models in hypercomplex spaces, 2023.

[62] K M Annervaz, Somnath Basu Roy Chowdhury, and Ambedkar Dukkipati. Learning beyond datasets: Knowledge graph augmented neural networks for natural language processing, 2018.

[63] Zhenghao Liu, Chenyan Xiong, Maosong Sun, and Zhiyuan Liu. Entity-duet neural ranking: Understanding the role of knowledge graph semantics in neural information retrieval, 2018.

[64] Ying Su, Tianqing Fang, Huiru Xiao, Weiqi Wang, Yangqiu Song, Tong Zhang, and Lei Chen. Entaile: Introducing textual entailment in commonsense knowledge graph completion, 2024.

[65] Xunzhu Tang, Tiezhu Sun, Rujie Zhu, and Shi Wang. Ckg: Dynamic representation based on context and knowledge graph, 2022.

[66] Haseeb Shah, Johannes Villmow, Adrian Ulges, Ulrich Schwanecke, and Faisal Shafait. An open-world extension to knowledge graph completion models, 2019.

[67] Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, Michal Podstawski, Lukas Gianinazzi, Joanna Gajda, Tomasz Lehmann, Hubert Niewiadomski, Piotr Nyczyk, et al. Graph of thoughts: Solving elaborate problems with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17682–17690, 2024.

[68] Hongyu Ren, Hanjun Dai, Bo Dai, Xinyun Chen, Denny Zhou, Jure Leskovec, and Dale Schuurmans. Smore: Knowledge graph completion and multi-hop reasoning in massive knowledge graphs, 2021.

[69] Cunxiang Wang, Jinhang Wu, Luxin Liu, and Yue Zhang. Commonsense knowledge graph reasoning by selection or generation? why?, 2020.

[70] Denis Krompaß, Stephan Baier, and Volker Tresp. Type-constrained representation learning in knowledge graphs, 2015.

[71] Takuma Ebisu and Ryutaro Ichise. Graph pattern entity ranking model for knowledge graph completion, 2019.

[72] Luis Galárraga. Effects of locality and rule language on explanations for knowledge graph embeddings, 2023.

[73] Fadhela Kerdjoudj and Olivier Curé. Rdf knowledge graph visualization from a knowledge extraction system, 2015.

[74] Liang Wang, Wei Zhao, Zhuoyu Wei, and Jingming Liu. Simkgc: Simple contrastive knowledge graph completion with pre-trained language models, 2022.

[75] Tong Chen, Sirou Zhu, Yiming Wen, and Zhaomin Zheng. Knowledge graph completion with text-aided regularization, 2021.

[76] Linlin Chao, Xiexiong Lin, Taifeng Wang, and Wei Chu. Pie: a parameter and inference efficient solution for large scale knowledge graph embedding reasoning, 2022.

[77] Daqian Shi. Knowledge graph extension by entity type recognition, 2024.

[78] Lingfeng Zhong, Jia Wu, Qian Li, Hao Peng, and Xindong Wu. A comprehensive survey on automatic knowledge graph construction, 2023.

[79] Yaqing Wang, Fenglong Ma, and Jing Gao. Efficient knowledge graph validation via cross-graph representation learning, 2020.

[80] Yankai Lin, Xu Han, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. Knowledge representation learning: A quantitative review, 2018.

[81] Filip Cornell, Yifei Jin, Jussi Karlgren, and Sarunas Girdzijauskas. Are we wasting time? a fast, accurate performance evaluation framework for knowledge graph link predictors, 2024.

[82] Mohamad Zamini, Hassan Reza, and Minou Rabiei. A review of knowledge graph completion, 2022.

[83] Tengfei Ma, Xiang song, Wen Tao, Mufei Li, Jiani Zhang, Xiaoqin Pan, Jianxin Lin, Bosheng Song, and xiangxiang Zeng. Kgexplainer: Towards exploring connected subgraph explanations for knowledge graph completion, 2024.

[84] Derong Xu, Ziheng Zhang, Zhenxi Lin, Xian Wu, Zhihong Zhu, Tong Xu, Xiangyu Zhao, Yefeng Zheng, and Enhong Chen. Multi-perspective improvement of knowledge graph completion with large language models, 2024.

[85] Ikhlas Alhussien, Erik Cambria, and Zhang NengSheng. Semantically enhanced models for commonsense knowledge acquisition, 2018.

[86] Sijia Wei, Wenwen Zhang, Qisong Li, and Jiang Zhao. Semantic parsing for question answering over knowledge graphs, 2024.

[87] Fan Lu, Quan Qi, and Huaibin Qin. Edge-enabled anomaly detection and information completion for social network knowledge graphs, 2024.

[88] Yanhui Peng and Jing Zhang. Lineare: Simple but powerful knowledge graph embedding for link prediction, 2021.

[89] Jianhao Shen, Chenguang Wang, Linyuan Gong, and Dawn Song. Joint language semantic and structure embedding for knowledge graph completion, 2022.

[90] Albert Sawczyn, Jakub Binkowski, Piotr Bielak, and Tomasz Kajdanowicz. Empowering small-scale knowledge graphs: A strategy of leveraging general-purpose knowledge graphs for enriched embeddings, 2024.

[91] Xintao Wang, Qianyu He, Jiaqing Liang, and Yanghua Xiao. Language models as knowledge embeddings, 2023.

[92] Théo Trouillon and Maximilian Nickel. Complex and holographic embeddings of knowledge graphs: A comparison, 2017.

[93] Hanrong Zhang, Xinyue Wang, Jiabao Pan, and Hongwei Wang. Saka: An intelligent platform for semi-automated knowledge graph construction and application, 2024.

[94] Aman Kumar and Swathi Dinakaran. Textbook to triples: Creating knowledge graph in the form of triples from ai textbook, 2021.

[95] Ying Zhou, Xuanang Chen, Ben He, Zheng Ye, and Le Sun. Re-thinking knowledge graph completion evaluation from an information retrieval perspective, 2022.

[96] Dattaraj J. Rao, Shraddha S. Mane, and Mukta A. Paliwal. Biomedical multi-hop question answering using knowledge graph embeddings and language models, 2022.

[97] Lars-Peter Meyer, Johannes Frey, Felix Brei, and Natanael Arndt. Assessing sparql capabilities of large language models, 2024.

[98] Vasile Ionut Remus Iga and Gheorghe Cosmin Silaghi. Assessing llms suitability for knowledge graph completion, 2024.

[99] Sakher Khalil Alqaaidi and Krzysztof Kochut. Relations prediction for knowledge graph completion using large language models, 2024.

[100] Camila Díaz, Jocelyn Dunstan, Lorena Etcheverry, Antonia Fonck, Alejandro Grez, Domingo Mery, Juan Reutter, and Hugo Rojas. Automatic knowledge-graph creation from historical documents: The chilean dictatorship as a case study, 2024.

[101] Xuelu Chen, Muhao Chen, Changjun Fan, Ankith Uppunda, Yizhou Sun, and Carlo Zaniolo. Multilingual knowledge graph completion via ensemble knowledge transfer, 2020.

[102] Michael R. Douglas, Michael Simkin, Omri Ben-Eliezer, Tianqi Wu, Peter Chin, Trung V. Dang, and Andrew Wood. What is learned in knowledge graph embeddings?, 2021.

[103] Meng Qu and Jian Tang. Probabilistic logic neural networks for reasoning, 2019.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

AI-generated, for reference only.