Agentic AI, Context Engineering and Knowledge Graphs: Current Approaches, Challenges and Opportunities

Niraj Karki[1], Manjila Pandey[1], Sanju Tiwari[2], Nandana Mihindukulasooriya[3], Sven Groppe[4] and Daniel Dobriy[5]

[1]*Pulchowk Engineering Campus, Nepal*
[2]*Sharda University, Delhi-NCR, India*
[3]*IBM Research, NYC, USA*
[4]*Universität zu Lübeck, Germany, & TU Bergakademie Freiberg, Germany*
[5]*WU Vienna, Austria, & Dobriy AI, Austria*

Abstract

With the recent advancements in Large Language Models (LLMs) and Agentic AI, Context Engineering (CE) has emerged as a novel research area. CE aims to fill the prompts for LLM Agents with relevant contextual knowledge required to perform complex tasks, where the quality of this context is paramount for reliability. Knowledge Graphs (KGs) offer a promising approach to integrate diverse contextual knowledge based on Semantic Web and Knowledge Representation approaches. In this paper, we study current approaches to identify challenges and opportunities for utilising KGs in CE and explore their limitations and strategic future research directions. The findings illustrate inconsistencies in methodologies and limited understanding of scalability and quality assurance challenges, which slow down the development of robust, context-aware AI systems capable of dealing with real-world complexity and multi-domain reasoning tasks.

Keywords

Context Engineering, Knowledge Graphs, Large Language Models, Ontology, Knowledge Representation, Quality

1. Introduction

Large Language Models (LLMs) have demonstrated impressive performance across a wide variety of natural language tasks, including machine translation [1], question answering [2], summarization [3], and dialogue generation [4]. Their growing influence spans domains such as cybersecurity, education, and healthcare due to their ability to generalise well on language tasks [5]. However, the performance and efficiency of these models are fundamentally governed by the context they receive. They still face significant challenges such as difficulty in handling structured knowledge, especially in the case of smaller models [6], lack of explicit memory, and hallucinations, where models produce plausible sounding but factually incorrect responses [7]. These limitations directly affect the quality, factual reliability, and robustness of LLM-based agentic systems. To address these shortcomings, the emerging domain of Context Engineering (CE) focuses on providing high-quality context to guide LLMs more effectively [8]. Within this landscape, Knowledge Graphs (KGs) provide a promising solution to tackle many persistent challenges in LLM-based agentic systems [9]. As structured representations of entities and their interrelations, KGs help bridge the gap between unstructured language and symbolic reasoning [10]. KGs encode factual information about real-world objects in a machine-readable format [11] and overcome LLM limitations by grounding context, using multi-hop reasoning and serving as a validator for LLM outputs and response quality [12].

QuaLLM-KG 2026, 1st International Workshop on Quality in Large Language Models and Knowledge Graphs, Published in the Proceedings of the Workshops of the EDBT/ICDT 2026 Joint Conference (March 24-27, 2026), Tampere, Finland

*Corresponding author: Niraj Karki.

†These authors contributed equally: Niraj Karki and Manjila Pandey.

nirajkarki416124@gmail.com (N. Karki); 078bei020.manjila@pcampus.edu.np (M. Pandey); tiwarisanju18@ieee.org (S. Tiwari); nandana.cse@gmail.com (N. Mihindukulasooriya); sven.groppe@uni-luebeck.de (S. Groppe); daniel.dobriy@wu.ac.at (D. Dobriy)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

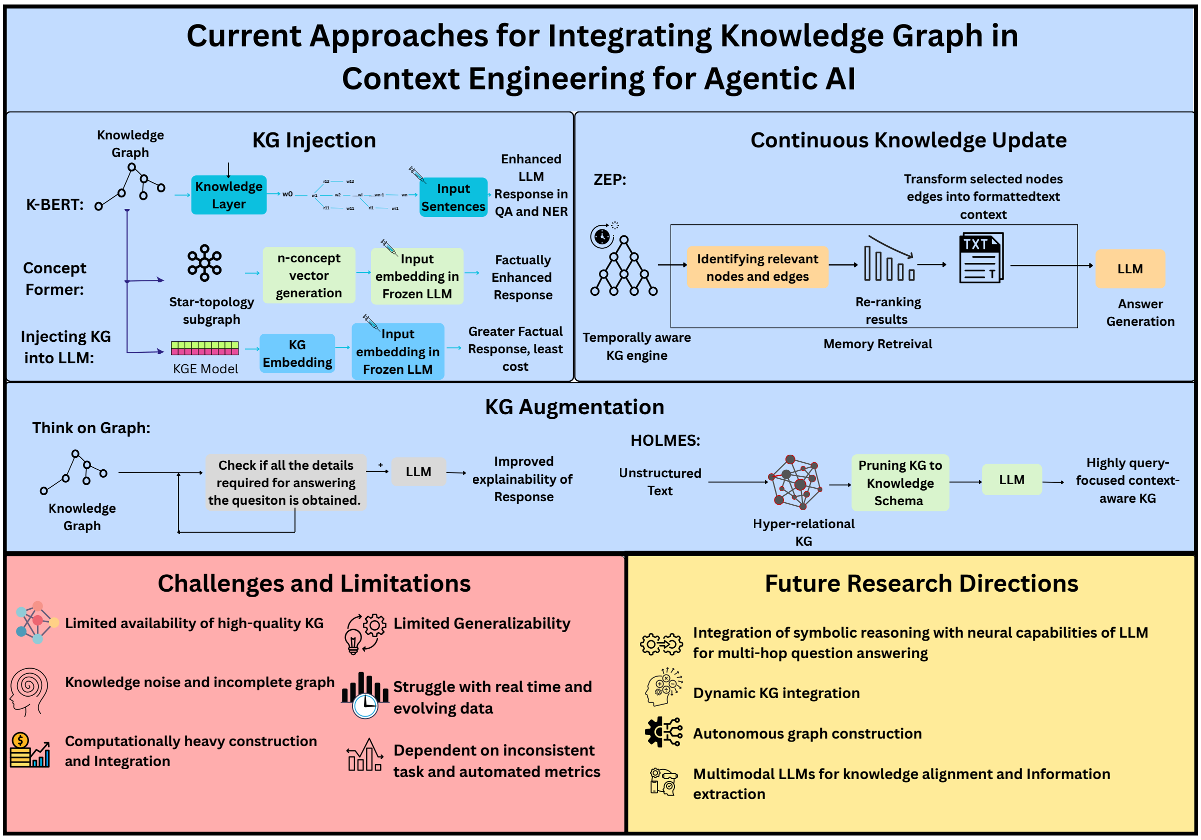


Figure 1: Key aspects in KG-based CE approaches in Agentic AI

Therefore, we aim to explore the current landscape of KG-augmented LLM methods from a CE perspective, addressing the following three research questions:

- **RQ1**: What are the recent approaches for integrating KGs in CE for Agentic AI?
- **RQ2**: What are the experienced challenges and limitations in KG-enhanced CE?
- **RQ3**: What are the current gaps to advance the interdisciplinary field?

Figure 1 illustrates the workflow of the paper, highlighting different approaches for KG integration in LLMs, challenges faced, and concluding with future research perspectives: Section 2 presents the scope and methodology of the literature study, while Section 3 provides the background, including the basics of CE, Agentic AI, and KG. Section 4 details integration methodologies with their limitations and challenges. Section 5 outlines future research directions.

2. Methodology of the Literature Study

For RQ 1/2, we conducted the following steps for recent articles published between 2020 and 2025:

Initial Search: We use 3 research databases, IEEE Xplore (accessed on 10 August 2025), ACM Digital Library (accessed on 10 August 2025), and arXiv (accessed on 12 August 2025) to access relevant articles. Afterwards, the search engine Google Scholar is also used for article search. As CE and KG-LLM integration are rapidly evolving fields, we include methodologies that are currently available as pre-prints but have not completed the peer-review cycle yet. We use particular keywords to probe journals and proceedings that focus on CE, ontology, and the use of KGs to improve the quality, performance, and efficiency of LLMs and agentic systems. Specifically, we probe articles whose titles and abstracts (i) matches “Context Engineering”, “In-Context Learning”, “Chain-of-thought” or “Retrieval-Augmented

Generation (RAG)"; and (ii) matches either "Knowledge Graph", "Structured Knowledge", "Knowledge Base" or "Ontology"; and (iii) matches either "Agentic AI", "Agentic Systems", "LLM" or "Generative AI". From the initial search, we collect 436 articles that are potentially related to the topics of our study.

Selection and Inclusion: In the initial selection step, we exclude 340 and select 96 articles based on the title and abstract that are not related to CE in AI systems using KGs. In the 2nd selection step, we preliminarily read the remaining articles and filter out the articles based on methods and reported quality-related outcomes - narrowing down the number of articles to 52. In the last selection step, we fully read each article and exclude theoretical articles that purely discuss KG with no LLM components. Finally, we select 35 articles that discuss CE for Agentic AI, propose novel KG-LLM integration methods, and contribute to enhancing LLM performance as well as output quality through better context.

Categorization and Taxonomy: We classify the KG-LLM integration approaches into categories according to the stage at which the KG is integrated into the model, and discuss it in section 4. The specific taxonomy is chosen because it systematically covers knowledge integration into the model lifecycle and its impact on model performance and reliability.We identify five categories of KG integration: Pre-training, Post-training, KG-based Augmentation, Inference-time Integration, and Continuous Update.

Analysis and Synthesis: We extract and analyze model name, core methodology, key contributions, integration type, datasets used, relative performance, and limitations of each article (see table 1), which is the basis for the identification of common challenges (section 4) and future research directions (section 5).

3. Background

Although LLMs have revolutionized AI applications, their effectiveness remains dependent on the quality and structure of contextual information provided to them [13]. Traditional prompt engineering approaches are often insufficient, so CE focuses on managing contextual information, addressing the hallucinations and factual inaccuracies [14]. KGs can add structured, well-organized information with well-defined semantics to LLM, reduce hallucinations, and increase the factual precision of responses [15].

3.1. Prompt Engineering

A prompt in GenAI is a textual input that enhances the model output, ranging from simple text to specific information [16]. Different prompting techniques like descriptive prompts in image generation models like DALL-E 3 and simple queries to complex problem statements in GPT-5 [17] are used in practice. They may guide LLMs for logical reasoning with simple queries and advanced techniques like chain-of-thought prompting [14]. So, prompt engineering crafts the optimal prompt to achieve a specific goal and get desired domain-specific output [18]. It requires a blend of domain knowledge, an understanding of the underlying behavior of the model, and careful adaptation to the specificities of the chosen LLM, such as its instruction-tuning regime, context window limitations, and response calibration mechanisms. [16].

3.2. Retrieval-augmented generation (RAG)

Retrieval Augmented Generation (RAG) improves the capability of LLM in knowledge-intensive tasks, continuous knowledge updates, and provides domain-specific information [19]. RAG addresses LLM limitations such as hallucination and short context window by providing important contextual information, including knowledge from external databases [20, 21]. Naive RAG focused on basic chunk similarity and an incomplete understanding of queries [21] whereas advanced RAG introduced hierarchical indexing and reranking. Modular RAG introduced a task-specific, flexible, and modular architecture [21]. Recently, GraphRAG [22] excels in capturing relational knowledge for more accurate and context-aware retrieval by relying on KGs to support multi-hop traversal and entity-relation matching, thereby integrating various degrees of reasoning over structured data.

Table 1
Comparative summary of Knowledge Graph (KG)–augmented Language Models in Context Engineering (CE)

Model / System	Integration Type & Key Contributions	Core Methodology	Dataset(s)	Performance / Findings	Limitations
K-BERT [23]	KG injection for contextual representation and factual grounding	Soft-position embeddings, visible matrix to prevent noise and Mask Transformer to preserve original semantics	CN-DBpedia, HowNet, MedicalKG	+1–2% gains in precision/recall/F1 over BERT	Relies on high-quality KGs; single-hop reasoning
KG-FiD [24]	KG-based passage graphs improve retrieval grounding for QA	Fusion-in-Decoder with Graph Attention Network reranking	Natural Questions, TriviaQA	+1.5% EM, 1.1% TriviaQA improvement, 60% lower computational cost	Static KG; no multi-hop inference
ConceptFormer [25]	KG-derived concept vectors enhance factual recall with compact embedding	Two-stage KG embedding injection into GPT-2	Synthetic, Wikipedia	+348% Hit@10 synthetic, +272% real text	One-hop only; small LLM scale
Injecting KGs into LLMs [26]	KGE embeddings fused with frozen LLM for graph-aware reasoning	GraphToken framework (TransE, DistMult, RotatE)	Synthetic, AIDS, MUTAG, AQSOL	0.96 (0-hop), 0.85–0.9 (1-hop), 0.75 (2-hop)	Node-only queries; performance drop on complex tasks
KGAT [27] (Bias Mitigation)	KG-integrated fairness-aware training for bias reduction	KG-Augmented GNN with multi-head attention	Wikipedia, IMDb, COMPAS	+15% demographic parity, +5% accuracy, −7% bias	High computational cost; static KGs
KnowPrompt [28]	KG-aware prompt tuning for relation extraction	Virtual Type & Answer Words with joint optimization	SemEval, DialogRE, TACRED	+5.4 F1 over baselines; +22.4% few-shot F1	Relation-only; multi-label RE difficult
KERE [29]	Ontology-based RE using multi-level KG integration	ReOnto, DocRE-CLip, RGCN, DistMult	DocRED, BioRel	F10.9; outperforming baselines	KB noise; scalability limits
KARMA [30]	Multi-agent LLM pipeline for automatic KG enrichment	Nine LLM agents	PubMed	83.1 % correctness; 38K new entities	No human validation; domain variation
LLMs+Hallucination [	KG grounding reduces hallucination and improves factuality	KG injection during pretrain/inference/post-gen	Medical datasets	Improved factual accuracy and multilingual gains	Static KG; lacks active detection
THINK-ON-GRAPH [32]	Joint KG–LLM reasoning via beam search exploration	Dynamic KG traversal with beam search (ToG)	9 KG QA datasets including WebQSP, GrailQA, QALD10-en	SOTA on 6/9 QA datasets;	depends on KG completeness and beam-search hyperparameters
TrumorGPT (2024) [33]	GraphRAG-based fact-checking with semantic health KG	Semantic retrieval + reasoning	PolitiFact (600 claims)	85.5% accuracy; 85% TP, 92% TN	Weak on long text; binary-only
TrumorGPT (2025) [34]	Graph-based RAG for political fact-checking	GraphRAG + semantic similarity ranking	PolitiFact (100 claims)	88% TP; 93% TN	Narrow domain; binary classification
Semantic Verification in RAG [35]	KG-based fact verification and personalization in RAG	RAG + FactReRanker; iterative design	Civic QA datasets	Higher factual accuracy and relevance	Costly KG upkeep; domain limits
EFSUM [36]	KG-based factual summarization for zero-shot QA	Retrieval of existing KG facts followed by Summarization via Distillation & DPO	Mintika, WebQSP	Superior helpfulness and faithfulness; improved QA accuracy (density & clarity)	Retriever errors (e.g., hop separation) & Metric limitations
Interactive-KBQA [37]	Multi-turn KBQA using LLM-guided SPARQL generation	Few-shot semantic parsing with LLMs	WebQSP, MetaQA, KQA Pro	+29.8% comparative QA gain	High compute; API limitations
ZEP [38]	Temporal KG memory for agentic reasoning	Dynamic bi-temporal KG (Graphiti) + LLM	DMR, LongMemEval	94.8% DMR; +18.5% LME; −90% latency	Weak single-session recall; lacks ontology
SURGE [39]	KG-augmented dialogue generation for consistency	GNN subgraph retriever + invariant graph encoder	KQA Dialogue	Outperforms baselines on F1/ROUGE	Small model; entity linking limits
FOLK [40]	Logical claim verification with KG grounding	FOL predicates; LLM decomposition	FEVEROUS, HOVER, SciFact	11.30% average improvement (Macro F-1) over CoT/Self-Ask baselines	Synthetic claims lack real-world complexity; High computational cost
Question-Guided KG Re-scoring	KG re-ranking for question-guided reasoning	Question-attended GNN encoding	OBQA, ARC, Riddle, PIQA	Improved factual reasoning accuracy	Symbolic reasoning limits
VisDoM [41]	Multimodal retrieval-based reasoning for visual-text QA	Parallel text/visual RAG with CoT reasoning	VisDoMBench	+12–20% over baselines; stable scaling	High compute; visual bias
HOLMES [42]	Hyper-relational KG reasoning for multi-hop QA	Level-order (BFS) over entity–document graph, schema-based pruning, hyper-triple construction + verbalization	HotpotQA, MuSiQue	+20% EM on HotpotQA, +26% EM on MuSiQue;	May omit relevant facts due to extraction/schema pruning; graph construction and pruning overhead;
Learning to Plan from KGs [43]	KG-based logical plan generation for retrieval-augmented QA	Pattern grounding + NL verbalization	HotPotQA, MuSiQue, Bamboogle	Higher QA accuracy; better planning	Limited question-type coverage

3.3. Agentic AI

Agentic AI is presented as an evolution of GenAI applications, enhancing systems to operate independently, perform broader aspects rather than isolated tasks, and execute complex activities [44]. Historically proposed as foundational part of the Semantic Web ecosystem [45], modern AI Agents extended the capabilities of LLMs by leveraging external tools, function calling, and workflows, enabling them to perform more complex processes through planning, tool selection, and feedback loops. The Agentic AI paradigm further extends this autonomy by designing systems that consist of multiple agents that coordinate and communicate with each other as well as perform tasks collaboratively, adapting to dynamic conditions to achieve a broader goal [46, 47].

3.4. Context Engineering

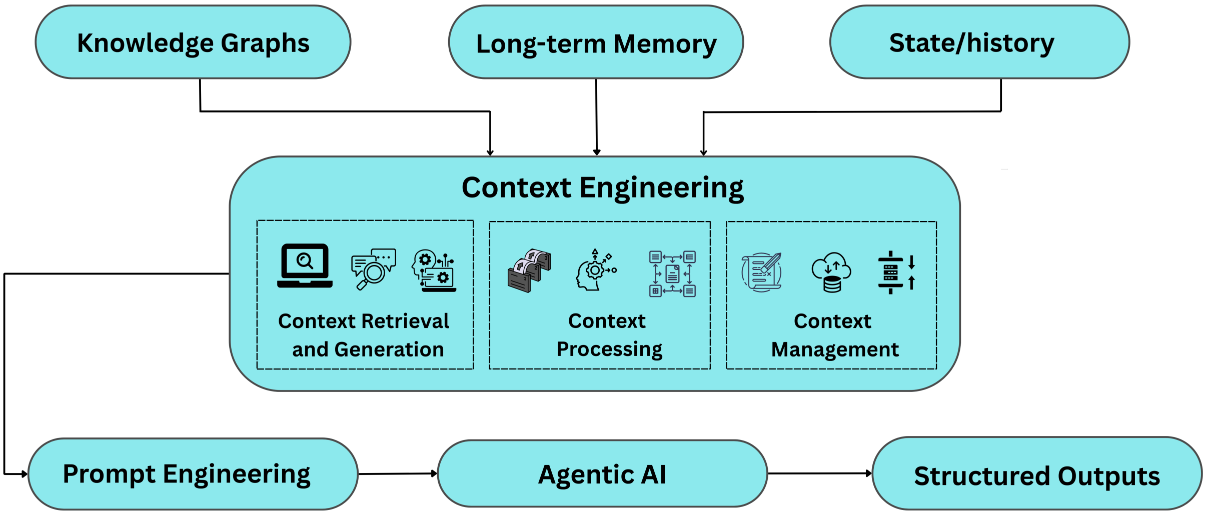


Figure 2: An overview of Context Engineering.

Addressing the shortcomings of traditional prompt engineering, CE enhances LLM capabilities by systematically designing, filtering, and structuring input information. At the core of this framework lies the information architecture, which structures context in a hierarchical order and groups related concepts to reduce processing load [48]. Furthermore, contextual relevance and filtering optimise limited context windows by applying query-context alignment, ranking details by importance, and reducing redundancy to ensure more accurate and useful responses [49, 50]. Multimodal integration further boosts these capabilities by incorporating diverse modalities (text, visual, and temporal information), which enable complex cross-modal reasoning.

Figure 2 shows key components of modern AI systems, highlighting the role of CE to enable more capable and goal-directed Agentic AI systems: Different sources of information, incl. KGs, long-term memory, and past state or history, are used together to create, process, and manage useful context. This context helps to guide AI systems to act more intelligently and produce better structured results.

3.5. Knowledge Graphs: Representation and Reasoning

KGs provide structured, machine-interpretable representations of knowledge through subject-predicate-object triples, enabling semantic understanding and automated reasoning across complex, interconnected datasets [51, 52]. They are widely used in domains such as search, recommendation systems, information retrieval, and data integration [53]. Their graph-based structure, comprising nodes (entities, literals) and edges (relations), presents a rich semantic representation [52]. Furthermore, bidirectional integration of KGs with LLMs has created new opportunities for contextual engineering, where KGs enhance context ingestion and query enrichment while LLMs contribute to KG construction and relationship prediction [54]. Modern KGs capture factual information and deeper relationships between concepts [55] such as hierarchies and causal links that are beyond surface level association [51, 56]. In addition to semantic richness, KGs support multi-hop and type-based reasoning, which supports tasks such as classification, generalization and logical reasoning [57]. Moreover, the dynamic nature of KGs allows updates that add new entities without requiring reconstruction of full graph [58].

4. Approaches for Integrating KGs into Context Engineering and Their Limitations

KG and LLM have recently gained increased attention, as both technologies are highly complementary in their capabilities [59]. LLMs excel at natural language understanding and generation, while KGs offer structured, semantically rich information that enhances LLM effectiveness and interpretability. In this section, we explore several methodologies to integrate KGs into CE for Agentic AI and the limitations they suffer. We also explore different datasets used to advance KGs, CE, and Agentic AI.

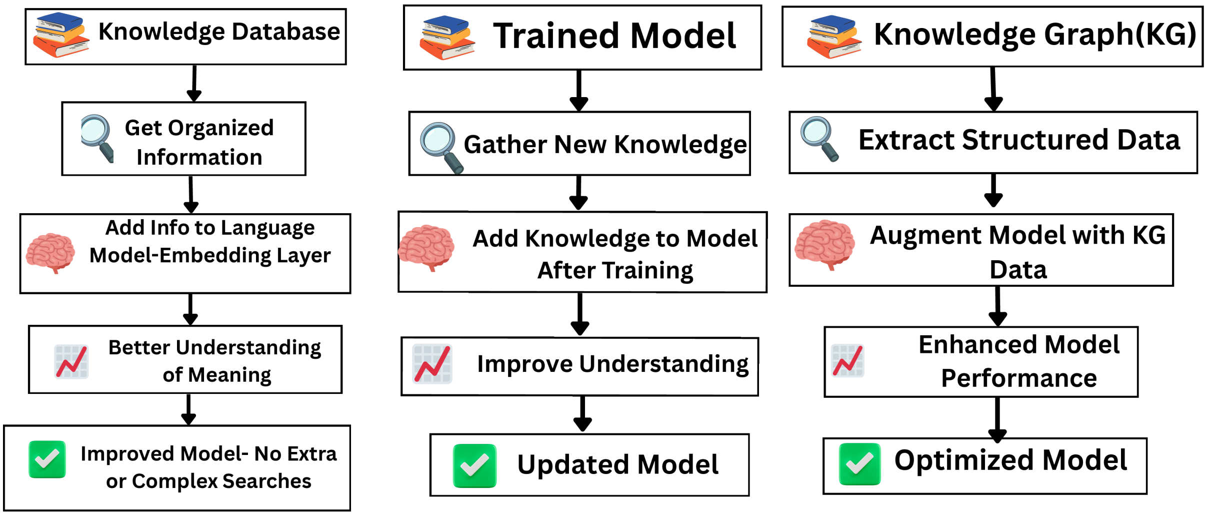
4.1. Knowledge Integration Methodologies

We categorize KG–LLM integration methods into pre-training, post-training, KG-based augmentation, inference time integration and continuous knowledge updates based on the type of integration of KG into the LLM pipeline. Figure 3 illustrates the interaction of each method with model training or inference, while Figure 4 contains the taxonomy of KG integration methodologies in the literature.

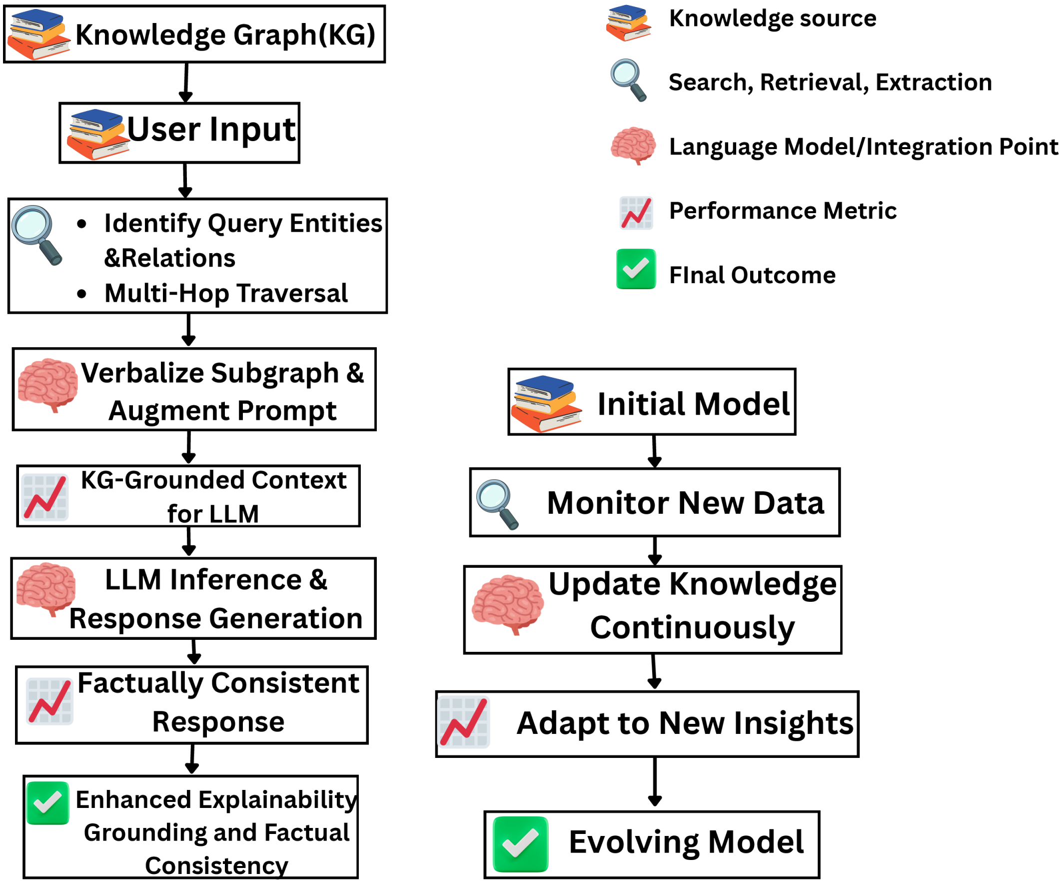
Pre-training integration involves using structured knowledge into the model's embedding or representation layer even before training. As indicated in figure 3a, KG is used to shape embeddings before the main training in pre-training integration approaches. Early injection-based works such as K-BERT [23] inject KG triples directly into input sentences through a knowledge layer as structured "sentence trees," controlled by soft-position embeddings and visible matrices. It is used for expert reasoning in tasks like question answering (Q&A) and Named Entity Recognition (NER). A similar approach is employed in ConceptFormer [25], which injects KG-derived concept vectors into the LLM embedding space without retraining. Both methods effectively ground local context and improve domain precision but remain limited by the quality and completeness of curated graphs. Building on previous injection-based strategies, Graph-Token uses embeddings to integrate KG representations directly into a frozen LLM that enables knowledge-aware reasoning without additional fine-tuning [26]. Although these approaches demonstrate significant improvement in generating grounded responses, current evaluations focus mainly on node-level reasoning tasks such as existence, counting, and identification. More complex tasks, for example, edge-centric and graph-level reasoning, remain underexplored.

On the contrary, *post-training* integration adds new knowledge to an already trained model (see figure 3a). Building on this direction, Lavrinovics et al. [31] propose a KG-based hallucination mitigation framework, where knowledge from KGs is integrated at multiple stages of the LLM pipeline. Their method extends conventional factual injection approaches by enabling autonomous fact, self-correction, and reasoning via KG-based memory and hybrid mitigation strategies. However, the approach remains limited by static graph updates and modular complexity.

In *KG-based augmentation*, we augment the model with data from KGs for enhanced model performance in terms of accuracy and depth. *Inference time* integration (figure 3b) does not require LLMs to store KGs; instead, they dynamically query and use KGs at the moment of answering a user query. Hence, the aim of this approach is to create a grounding mechanism that supports LLMs in generating responses that are explainable, factually grounded, and consistent. Furthermore, *continuous update* involves repetitively monitoring new data and continuously updating knowledge; hence, the model is ever-evolving and can adapt to new insights (figure 3a). Techniques use KGs during inference to boost the capabilities of LLMs using semantic and structural relationships without requiring embedding during pre-training. For example, THINK-ON-GRAPH [32] enables KGs and LLMs to work



(a) Pre-training, post-training, and KG-based augmentation



(b) Inference-time integration and continuous knowledge update

Figure 3: A classification of knowledge integration.

together through a beam search algorithm to dynamically explore multiple reasoning paths within a KG, improving decision-making and explainability. HOLMES [42] enhances interpretability through hyper-relational schemas and controlled multi-hop BFS expansion to identify missing facts and retrieve supporting evidence. Their performance depends on the graph quality and completeness, where missing or noisy triples can disrupt reasoning chains. Large-scale traversal of graphs such as Wikidata remains computationally demanding.

Recent research uses KGs as structural scaffolds to enhance retrieval and context building in retrieval-

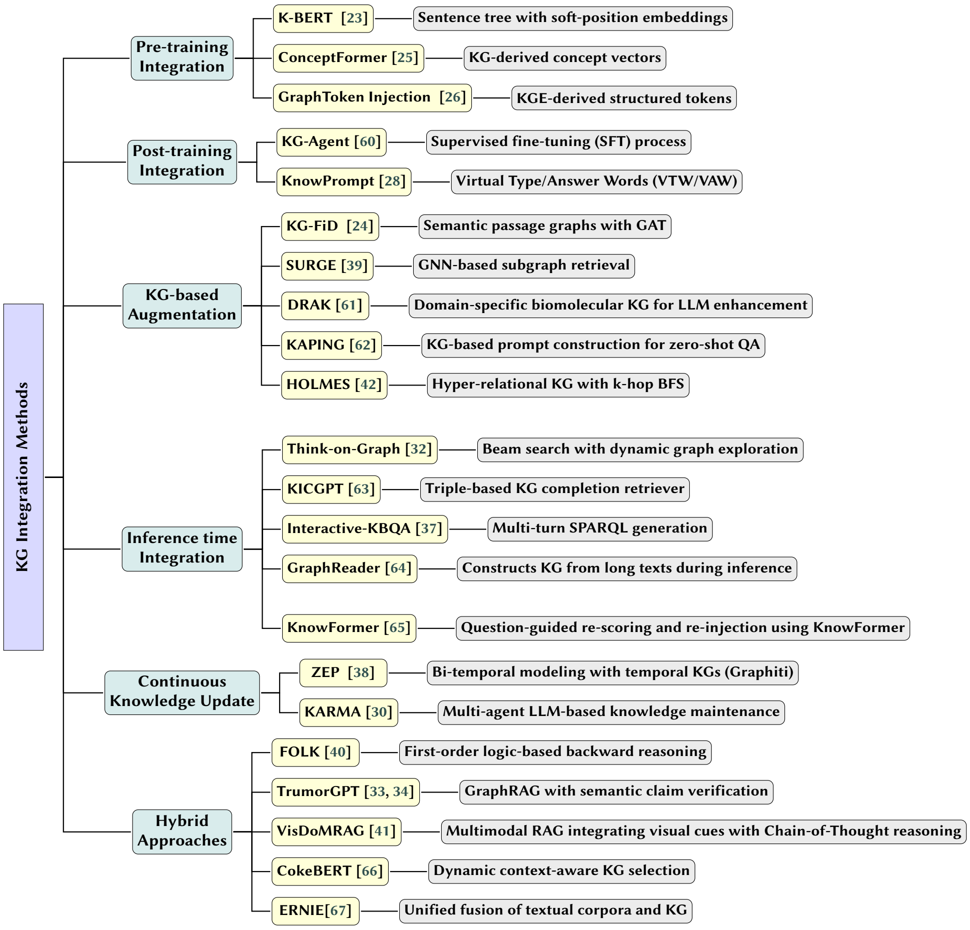


Figure 4: Taxonomy of Knowledge Graph Integration Methodologies

augmented generation (RAG) systems. For example, KG-FiD [24] integrates semantic passage graphs into Fusion-in-Decoder architectures that enable more precise reranking and improve the extraction of answer-relevant text. This yields notable gains in exact-match accuracy and reduces computational cost. Similarly, frameworks such as Evidence-Focused Fact Summarization (EFSUM) [36] and TrumorGPT [33] employ KG-based subgraph construction, OpenIE-derived triples, or graph-guided document filtering to enhance faithfulness and mitigate hallucinations. These retrieval methods are vulnerable to entity-linking and subgraph errors, often yielding confident but wrong answers in fast-changing domains.

4.2. Emerging Paradigms: Dynamic Memory, Hybrid Systems, and Open Challenges

The field is rapidly evolving toward incorporating KGs into complex systems, giving rise to dynamic agent memory architectures and hybrid symbolic-neural models. ZEP introduces a temporally aware KG that unifies episodic, semantic, and community subgraphs for agent memory, enabling accurate, low-latency long-term memory for real applications. KARMA [30] applies a multi-agent LLM system to independently ingest, segment, align, and validate new knowledge for expanding KG coverage. However, they also have issues concerning quality assurance, validation, and reliability, as automatically generated triples can introduce factual inconsistencies, requiring human oversight to maintain reliability.

Other approaches adopt hybrid architectures that combine symbolic structure with LLM generation. Systems such as FOLK, SURGE [39], Interactive-KBQA [37], and KERE [29] leverage logical representations, subgraph retrieval, and ontological constraints to enable explainable, entity-level reasoning across tasks including dialogue and relation extraction. While these methods can enhance faithfulness, consistency, and interpretability, they also introduce system-level complexities, as errors in retrieval, linking, or component interaction can propagate and undermine symbolic guarantees. More broadly, recurring challenges emerge around balancing structural precision with neural adaptability: graph-based methods improve factual consistency but depend on graph quality and face scalability limits, and KGs are evolving from static background resources to dynamic, agent-driven memory systems for multi-turn, multimodal, temporally extended reasoning [38, 30, 37]. Current methods span from structural KG injection (e.g., K-BERT [23]) to dynamic, agent-based architectures (e.g., ZEP, KARMA [30]) (see Table 1), reflecting a shift toward more context-aware and explainable reasoning systems.

These studies highlight the rapid advancement of LLM-KG integration with the need of establishing a consistent evaluation framework and addressing the challenges of large-scale, dynamically evolving graphs. With operational use, balancing adaptability with reliability will remain a major challenge.

4.3. Datasets and different evaluation metrices

This section highlights datasets from papers advancing **Agentic AI**, **CE**, and **KGs**, focusing on those most relevant to our analysis. From the survey, HotpotQA is the most frequently used dataset due to its robust framework for multi-hop question answering. WebQSP, CWQ, DocRED, Synthetic, PolitiFact, and DBpedia are also frequently used for question answering, relation extraction, and fact checking.

Table 2 summarizes the datasets and their usages in literature in three research areas: Knowledge Graphs (KG), Context Engineering (CE), and Agentic AI (AA).

5. Future Research Direction

We identified several research directions based on our literature study to address fundamental limitations in knowledge correctness and reliability while opening new directions for future innovation.

Neuro Symbolic Integration: An emerging direction for advancing CE is the integration of symbolic reasoning from KGs with the neural capabilities of LLMs [69]. Approaches like ConceptFormer [25] and HOLMES [42] inject KG-derived knowledge into LLMs, while THINK-ON-GRAPH [32] demonstrates how symbolic reasoning and neural generation can work together in multi-hop question answering. However, most methods rely on static alignments or task-specific pipelines, limiting scalability and dynamic updates. Future research should focus on seamlessly aligning structured KGs and unstructured LLM outputs, integrating neural and symbolic reasoning to improve interpretability [**?**]. This will enhance the development of trustworthy, real-time, context-aware agentic AI systems [**?**].

Autonomous and Self-Updating KGs: Most current systems like KG-FiD [24], KGAT [27], and EFSUM [36] rely on static KGs, which struggle with evolving data contexts. Self-updating KGs address this by enabling real-time integration of facts and relationships without system retraining. It still faces challenges with static and outdated information due to manual validation and query accuracy [70]. Systems like ZEP [38] and KARMA [30], demonstrate the potential for autonomous construction. Future systems should update their KGs autonomously by monitoring external data, identifying missing information, and making necessary adjustments to maintain a consistent knowledge structure.

Multimodal Context Grounding for Real-World Understanding: KGs should integrate text, images, audio, and video to better model the real world and improve CE. Systems like VisDoMRAG [41] show 12–20% gains on visually rich datasets but still face visual bias and cross-modal alignment issues. Current multimodal methods (e.g., image labeling and symbol grounding with datasets like MSCOCO [71]) reach only 43% accuracy in detecting misclassified objects [71] and fail to capture abstract or emotional concepts. Utilizing multimodal LLMs to extract information from sources like medical scans and verify data consistency, along with attention mechanisms for live context updates, has great potential. This will enhance healthcare, robotics, and chat systems [72].

Quality: With the rise of Agentic AI and hybrid LLM-KG systems, ensuring quality becomes more complex. While it's important to assess individual component quality to improve the overall system, it's also crucial to explore how small quality issues in components may accumulate into larger problems. Investigating the addition of monitoring components to mitigate such issues seems promising. We encourage research into quality-by-design software architectures to address potential quality challenges in future hybrid Agentic AI systems.

Table 2
Datasets that are related to Knowledge Graph (KG), Context Engineering (CE), and Agentic AI (AA)

Dataset	Description	KG	CE	AA
Synthetic [26]	Random graphs used for node-related reasoning tasks with KG embeddings in LLMs	[26]	[25]	
AIDS [26]	Molecular graphs for anti-HIV activity, used for KG reasoning and context embedding	[26]		
MUTAG [26]	Nitroaromatic compound graphs, used for KG reasoning and context embedding	[26]		
AQSOL [26]	Molecular graphs for solubility, used for KG reasoning and context embedding	[26]		
PubMed data [30]	Biomedical texts used for automated KG enrichment with multi-agent LLMs	[30]		[30]
WebQSP [36]	KG-based QA dataset with Freebase/Wikidata, used for reasoning and context construction	[32, 36]	[32, 36, 37]	[32, 37]
GrailQA [68]	KG-based QA dataset with Freebase/Wikidata, used for reasoning and context construction	[32]	[32]	[32]
CWQ [32]	Complex KG-based QA dataset with Wikidata, used for reasoning and context construction	[32, 37]	[32, 37]	[32, 37]
BioRel [29]	For sentence-level relation extraction	[29]	[29]	
DocRED [29]	Document-level relation extraction dataset with Wikidata, used for KG-based extraction	[29]	[29]	
HotpotQA	Multi-hop QA dataset with Wikipedia-based KGs, used for reasoning and context construction	[42]	[42, 43]	[43]
MuSiQue [42]	Multi-hop QA dataset with complex reasoning, used for KG construction and reasoning	[42]	[42, 43]	[43]
2WikiMultiHopQA	Multi-hop QA dataset with Wikidataand complex		[43]	[43]
MetaQA [37]	KG-based QA dataset with multi-hop queries, used for reasoning and interactions	[37]	[37]	[37]
Natural Questions (NQ) [24]	Open-domain QA dataset with Wikipedia-based KGs, used for fact summarization		[24]	
TriviaQA [24]	Trivia QA dataset with Wikipedia-based KGs, used for fact summarization		[24]	
Mintaka [36]	Multilingual QA dataset with Wikipedia-based KGs, used for fact summarization	[36]	[36]	
Chnsenticorp [23]	A hotel review dataset for single-sentence sentiment classification	[23]		
MedicalKG [23]	A self-developed Chinese medical concept KG	[23]	[23]	
CN-DBpedia [23]	A large open-domain encyclopedic Chinese KG	[23]	[23]	
HowNet [23]	Knowledge base that links words to semantic units	[23]	[23]	
DialogRE [28]	A dataset for relation extraction in dialogue contexts	[28]	[28]	
TACRED [28]	A large-scale relation extraction dataset with diverse relation labels	[28]	[28]	
VisDoM [41]	Multimodal QA dataset with visually rich documents, used for context curation			[41]
HOVER [40]	Multi-hop claim verification dataset and verify complex claims against multiple information sources		[40]	
FEVEROUS [40]	Benchmark dataset for complex claim verification over un- and structured data		[40]	

Legend: KG: Dataset has been used in the citation as a structured graph (nodes, edges, or triples) to support KG reasoning or embedding.
CE: Dataset for shaping an LLM's context, such as through demonstrations, prompt construction, or embedding-based contextual signals.
AA: Dataset has been used within an agentic framework involving multi-step reasoning, tool interaction, or other interactive LLM behaviors.
Empty cells imply that the dataset is not used for that dimension in the referenced literature.

6. Conclusion

In conclusion, this study provides a thorough analysis of KG-based CE in Agentic AI systems, highlighting key research questions and integration strategies. We identified limitations and proposed future directions for improving context-aware models, with a focus on enhancing the quality and reliability of contextual knowledge. Beyond conventional integration methods, we also explored areas like continuous knowledge updates, neurosymbolic integration, self-updating KGs, and multimodal CE, all of which must address critical challenges of knowledge quality, consistency, and trustworthiness.

7. Declaration on Generative AI

During the preparation of this work, the authors used Grammarly to improve grammar, check spelling, and reword. After using these tool(s)/service(s), the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] D. Gao, K. Chen, B. Chen, H. Dai, L. Jin, W. Jiang, W. Ning, S. Yu, Q. Xuan, X. Cai, et al., Llms-based machine translation for e-commerce, Expert Systems with Applications 258 (2024) 125087.

[2] Y. Zhuang, Y. Yu, K. Wang, H. Sun, C. Zhang, Toolqa: A dataset for llm question answering with external tools, in: A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, S. Levine (Eds.), Advances in Neural Information Processing Systems, volume 36, Curran Associates, Inc., 2023, pp. 50117–50143. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/9cb2a7495900f8b602cb10159246a016-Paper-Datasets_and_Benchmarks.pdf.

[3] H. Zhang, P. S. Yu, J. Zhang, A systematic survey of text summarization: From statistical methods to large language models, ACM Comput. Surv. 57 (2025). URL: https://doi.org/10.1145/3731445. doi:`10.1145/3731445`.

[4] C. Liu, Z. Xie, S. Zhao, J. Zhou, T. Xu, M. Li, E. Chen, Speak from heart: An emotion-guided llm-based multimodal method for emotional dialogue generation, in: Proceedings of the 2024 International Conference on Multimedia Retrieval, ICMR '24, Association for Computing Machinery, New York, NY, USA, 2024, p. 533–542. URL: https://doi.org/10.1145/3652583.3658104. doi:`10.1145/3652583.3658104`.

[5] H. F. Atlam, Llms in cyber security: Bridging practice and education, Big Data and Cognitive Computing 9 (2025) 184.

[6] D. Dobriy, F. Bauer, A. Azzam, D. Banerjee, A. Polleres, Agentic SPARQL: Evaluating SPARQL-MCP-powered Intelligent Agents on the Federated KGQA Benchmark, Technical Report, Vienna Unversity of Economics and Business, 2026. doi:`10.57938/83c86964-2d48-46f1-b655-5bef78c1a837`.

[7] Z. Zhang, C. Wang, Y. Wang, E. Shi, Y. Ma, W. Zhong, J. Chen, M. Mao, Z. Zheng, Llm hallucinations in practical code generation: Phenomena, mechanism, and mitigation, Proceedings of the ACM on Software Engineering 2 (2025) 481–503.

[8] R. Shankar, Context is king: From prompt engineering to context engineering in healthcare ai, Available at SSRN 5365971 (2025). URL: http://dx.doi.org/10.2139/ssrn.5365971.

[9] H. Abu-Rasheed, C. Weber, M. Fathi, Knowledge graphs as context sources for llm-based explanations of learning recommendations, in: 2024 IEEE Global Engineering Education Conference (EDUCON), IEEE, 2024, p. 1–5. URL: http://dx.doi.org/10.1109/EDUCON60312.2024.10578654. doi:`10.1109/educon60312.2024.10578654`.

[10] W. Li, G. Qi, Q. Ji, Hybrid reasoning in knowledge graphs: Combing symbolic reasoning and statistical reasoning, Semantic Web 11 (2020) 53–62.

[11] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, X. Wu, Unifying large language models and knowledge graphs: A roadmap, IEEE Transactions on Knowledge and Data Engineering 36 (2024) 3580–3599. URL: http://dx.doi.org/10.1109/TKDE.2024.3352100. doi:`10.1109/tkde.2024.3352100`.

[12] H. Ren, H. Dai, B. Dai, X. Chen, D. Zhou, J. Leskovec, D. Schuurmans, Smore: Knowledge graph completion and multi-hop reasoning in massive knowledge graphs, in: Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining, 2022, pp. 1472–1482.

[13] S. An, Z. Ma, Z. Lin, N. Zheng, J.-G. Lou, W. Chen, Make your llm fully utilize the context, Advances in Neural Information Processing Systems 37 (2024) 62160–62188.

[14] Y. Xia, R. Wang, X. Liu, M. Li, T. Yu, X. Chen, J. McAuley, S. Li, Beyond chain-of-thought: A survey of chain-of-X paradigms for LLMs, in: O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, S. Schockaert (Eds.), Proceedings of the 31st International Conference on Computational Linguistics, Association for Computational Linguistics, Abu Dhabi, UAE, 2025, pp. 10795–10809. URL: https://aclanthology.org/2025.coling-main.719/.

[15] L. Cai, C. Yu, Y. Kang, Y. Fu, H. Zhang, Y. Zhao, Practices, opportunities and challenges in the fusion of knowledge graphs and large language models, Frontiers in Computer Science 7 (2025) 1590632.

[16] T. F. Heston, C. Khun, Prompt engineering in medical education, International Medical Education 2 (2023) 198–205.

[17] L. Giray, Prompt engineering with chatgpt: a guide for academic writers, Annals of biomedical engineering 51 (2023) 2629–2633.

[18] Q. Ye, M. Ahmed, R. Pryzant, F. Khani, Prompt engineering a prompt engineer, in: Findings of the Association for Computational Linguistics: ACL 2024, 2024, pp. 355–385.

[19] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, M. Wang, H. Wang, Retrieval-augmented generation for large language models: A survey, 2024. URL: https://arxiv.org/abs/2312.10997. `arXiv:2312.10997`.

[20] L. Mei, J. Yao, Y. Ge, Y. Wang, B. Bi, Y. Cai, J. Liu, M. Li, Z.-Z. Li, D. Zhang, C. Zhou, J. Mao, T. Xia, J. Guo, S. Liu, A survey of context engineering for large language models, 2025. URL: https://arxiv.org/abs/2507.13334. `arXiv:2507.13334`.

[21] Y. Gao, Y. Xiong, M. Wang, H. Wang, Modular rag: Transforming rag systems into lego-like reconfigurable frameworks, 2024. URL: https://arxiv.org/abs/2407.21059. `arXiv:2407.21059`.

[22] B. Peng, Y. Zhu, Y. Liu, X. Bo, H. Shi, C. Hong, Y. Zhang, S. Tang, Graph retrieval-augmented generation: A survey, 2025.

[23] W. Liu, P. Zhou, Z. Zhao, Z. Wang, Q. Ju, H. Deng, P. Wang, K-bert: Enabling language representation with knowledge graph, in: Proceedings of the AAAI conference on artificial intelligence, volume 34, 2020, pp. 2901–2908.

[24] D. Yu, C. Zhu, Y. Fang, W. Yu, S. Wang, Y. Xu, X. Ren, Y. Yang, M. Zeng, Kg-fid: Infusing knowledge graph in fusion-in-decoder for open-domain question answering, in: Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: long papers), 2022, pp. 4961–4974.

[25] J. Barmettler, A. Bernstein, L. Rossetto, Conceptformer: Towards efficient use of knowledge-graph embeddings in large language models, 2025. URL: https://arxiv.org/abs/2504.07624. `arXiv:2504.07624`.

[26] E. Coppolillo, Injecting knowledge graphs into large language models, arXiv preprint arXiv:2505.07554 (2025).

[27] R. Kumar, H. Kumar, K. Shalini, Detecting and mitigating bias in llms through knowledge graph-augmented training, in: 2025 International Conference on Artificial Intelligence and Data Engineering (AIDE), IEEE, 2025, pp. 608–613.

[28] X. Chen, N. Zhang, X. Xie, S. Deng, Y. Yao, C. Tan, F. Huang, L. Si, H. Chen, Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction, in: Proceedings of the ACM Web Conference 2022, WWW '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 2778–2788. URL: https://doi.org/10.1145/3485447.3511998. doi:`10.1145/3485447.3511998`.

[29] M. Jain, Knowledge enabled relation extraction, in: Companion Proceedings of the ACM Web Conference 2024, 2024, pp. 1210–1213.

[30] Y. Lu, J. Wang, Karma: Leveraging multi-agent llms for automated knowledge graph enrichment, 2025. URL: https://arxiv.org/abs/2502.06472. `arXiv:2502.06472`.

[31] E. Lavrinovics, R. Biswas, J. Bjerva, K. Hose, Knowledge graphs, large language models, and hallucinations: An nlp perspective, Journal of Web Semantics 85 (2025) 100844.

[32] J. Sun, C. Xu, L. Tang, S. Wang, C. Lin, Y. Gong, L. M. Ni, H.-Y. Shum, J. Guo, Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph, arXiv preprint arXiv:2307.07697 (2023).

[33] C. N. Hang, P.-D. Yu, C. W. Tan, Trumorgpt: Query optimization and semantic reasoning over networks for automated fact-checking, in: 2024 58th Annual Conference on Information Sciences and Systems (CISS), 2024, pp. 1–6. doi:`10.1109/CISS59072.2024.10480162`.

[34] C. N. Hang, P.-D. Yu, C. W. Tan, Trumorgpt: Graph-based retrieval-augmented large language model for fact-checking, IEEE Transactions on Artificial Intelligence (2025) 1–15. URL: http://dx.doi.org/10.1109/TAI.2025.3567369. doi:`10.1109/tai.2025.3567369`.

[35] A. Martin, H. F. Witschel, M. Mandl, M. Stockhecke, Semantic verification in large language model-based retrieval augmented generation, in: Proceedings of the AAAI Symposium Series, volume 3, 2024, pp. 188–192.

[36] S. Ko, H. Cho, H. Chae, J. Yeo, D. Lee, Evidence-focused fact summarization for knowledge-augmented zero-shot question answering, in: Y. Al-Onaizan, M. Bansal, Y.-N. Chen (Eds.), Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Miami, Florida, USA, 2024, pp. 10636–10651. URL: https://aclanthology.org/2024.emnlp-main.594/. doi:`10.18653/v1/2024.emnlp-main.594`.

[37] G. Xiong, J. Bao, W. Zhao, Interactive-kbqa: Multi-turn interactions for knowledge base question answering with large language models, arXiv preprint arXiv:2402.15131 (2024).

[38] P. Rasmussen, P. Paliychuk, T. Beauvais, J. Ryan, D. Chalef, Zep: A temporal knowledge graph architecture for agent memory, 2025. URL: https://arxiv.org/abs/2501.13956. `arXiv:2501.13956`.

[39] M. Kang, J. M. Kwak, J. Baek, S. J. Hwang, Knowledge graph-augmented language models for knowledge-grounded dialogue generation, 2023. URL: https://arxiv.org/abs/2305.18846. `arXiv:2305.18846`.

[40] H. Wang, K. Shu, Explainable claim verification via knowledge-grounded reasoning with large language models, in: H. Bouamor, J. Pino, K. Bali (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2023, Association for Computational Linguistics, Singapore, 2023, pp. 6288–6304. URL: https://aclanthology.org/2023.findings-emnlp.416/. doi:`10.18653/v1/2023.findings-emnlp.416`.

[41] M. Suri, P. Mathur, F. Dernoncourt, K. Goswami, R. A. Rossi, D. Manocha, Visdom: Multi-document qa with visually rich elements using multimodal retrieval-augmented generation, in: Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), 2025, pp. 6088–6109.

[42] P. Panda, A. Agarwal, C. Devaguptapu, M. Kaul, P. Ap, Holmes: Hyper-relational knowledge graphs for multi-hop question answering using llms, in: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2024, pp. 13263–13282.

[43] J. Wang, M. Chen, B. Hu, D. Yang, Z. Liu, Y. Shen, P. Wei, Z. Zhang, J. Gu, J. Zhou, J. Z. Pan, W. Zhang, H. Chen, Learning to plan for retrieval-augmented large language models from knowledge graphs, in: Y. Al-Onaizan, M. Bansal, Y.-N. Chen (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2024, Association for Computational Linguistics, Miami, Florida, USA, 2024, pp. 7813–7835. URL: https://aclanthology.org/2024.findings-emnlp.459/. doi:`10.18653/v1/2024.findings-emnlp.459`.

[44] J. Schneider, Generative to agentic ai: Survey, conceptualization, and challenges, arXiv preprint arXiv:2504.18875 (2025).

[45] A. Polleres, F. Bauer, D. Dobriy, T. Käfer, L. Kubelka, A. Harth, T. Wehr, On the historic roots of Agentic AI in Semantic Web Services, in: The ACM Web Conference 2026, 2026. URL: http://polleres.net/publications/poll-etal2026WebConfHoW.pdf, to appear, extended abstract (invited, History of the Web track).

[46] R. Sapkota, K. I. Roumeliotis, M. Karkee, Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges, Information Fusion (2025) 103599.

[47] D. B. Acharya, K. Kuppan, B. Divya, Agentic ai: Autonomous intelligence for complex goals—a comprehensive survey, IEEe Access 13 (2025) 18912–18936.

[48] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, et al., Retrieval-augmented generation for knowledge-intensive nlp tasks, Advances in neural information processing systems 33 (2020) 9459–9474.

[49] L. Brehme, T. Ströhle, R. Breu, Can llms be trusted for evaluating rag systems? a survey of methods and datasets (2025) 16–23.

[50] V. Lapov, N. Laurent, L. Araya, G. Ortiz, S. Albrecht, Dynamic context integration in large language models using a novel progressive layering framework, Research Square Preprints (2024). URL: https://doi.org/10.21203/rs.3.rs-5357232/v1.

[51] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al., Knowledge graphs, ACM Computing Surveys (Csur) 54 (2021) 1–37.

[52] S. Ji, S. Pan, E. Cambria, P. Marttinen, P. S. Yu, A survey on knowledge graphs: Representation, acquisition, and applications, IEEE transactions on neural networks and learning systems 33 (2021)

494–514.

[53] J. Jiaxin, X. Huang, B. Choi, J. Xu, S. S Bhowmick, L. Xu, Ppkws: An efficient framework for keyword search on public-private networks, 2020, pp. 457–468. doi:10.1109/ICDE48307.2020.00046.

[54] A. Kau, X. He, A. Nambissan, A. Astudillo, H. Yin, A. Aryani, Combining knowledge graphs and large language models, 2024. URL: https://arxiv.org/abs/2407.06564. arXiv:2407.06564.

[55] M. Yahya, J. G. Breslin, M. I. Ali, Semantic web and knowledge graphs for industry 4.0, Applied Sciences 11 (2021) 5110.

[56] F. N. Al-Aswadi, H. Y. Chan, K. H. Gan, From ontology to knowledge graph trend: ontology as foundation layer for knowledge graph, in: Iberoamerican Knowledge Graphs and Semantic Web Conference, Springer, 2022, pp. 330–340.

[57] K. Liang, L. Meng, M. Liu, Y. Liu, W. Tu, S. Wang, S. Zhou, X. Liu, F. Sun, K. He, A survey of knowledge graph reasoning on graph types: Static, dynamic, and multi-modal, IEEE Transactions on Pattern Analysis and Machine Intelligence 46 (2024) 9456–9478.

[58] S. M. Mohamed, S. Farah, A. M. Lotfy, K. A. Rizk, A. Y. Saeed, S. H. Mohamed, G. Khoriba, T. Arafa, Knowledge graphs: The future of data integration and insightful discovery, in: Advanced Research Trends in Sustainable Solutions, Data Analytics, and Security, IGI Global Scientific Publishing, 2025, pp. 99–146.

[59] H. Khorashadizadeh, F. Z. Amara, M. Ezzabady, F. Ieng, S. Tiwari, N. Mihindukulasooriya, J. Groppe, S. Sahri, F. Benamara, S. Groppe, Research trends for the interplay between large language models and knowledge graphs, Proceedings of the VLDB Endowment. ISSN 2150 (2024) 8097.

[60] J. Jiang, K. Zhou, W. X. Zhao, Y. Song, C. Zhu, H. Zhu, J.-R. Wen, Kg-agent: An efficient autonomous agent framework for complex reasoning over knowledge graph, in: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2025, pp. 9505–9523.

[61] J. Liu, X. Huang, Z. Chen, Y. Fang, Drak: Unlocking molecular insights withnbsp;domain-specific retrieval-augmented knowledge innbsp;llms, in: Natural Language Processing and Chinese Computing: 13th National CCF Conference, NLPCC 2024, Hangzhou, China, November 1–3, 2024, Proceedings, Part II, Springer-Verlag, Berlin, Heidelberg, 2024, p. 255–267. URL: https://doi.org/10.1007/978-981-97-9434-8_20. doi:10.1007/978-981-97-9434-8_20.

[62] J. Baek, A. F. Aji, A. Saffari, Knowledge-augmented language model prompting for zero-shot knowledge graph question answering, Proceedings of the First Workshop on Matching From Unstructured and Structured Data (MATCHING 2023) (2023). URL: https://api.semanticscholar.org/CorpusID:260063238.

[63] Y. Wei, Q. Huang, Y. Zhang, J. Kwok, KICGPT: Large language model with knowledge in context for knowledge graph completion, in: H. Bouamor, J. Pino, K. Bali (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2023, Association for Computational Linguistics, Singapore, 2023, pp. 8667–8683. URL: https://aclanthology.org/2023.findings-emnlp.580/. doi:10.18653/v1/2023.findings-emnlp.580.

[64] S. Li, Y. He, H. Guo, X. Bu, G. Bai, J. Liu, J. Liu, X. Qu, Y. Li, W. Ouyang, W. Su, B. Zheng, GraphReader: Building graph-based agent to enhance long-context abilities of large language models, in: Y. Al-Onaizan, M. Bansal, Y.-N. Chen (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2024, Association for Computational Linguistics, Miami, Florida, USA, 2024, pp. 12758–12786. doi:10.18653/v1/2024.findings-emnlp.746.

[65] Y. Zhang, K. Chen, X. Bai, Z. Kang, Q. Guo, M. Zhang, Question-guided knowledge graph re-scoring and injection for knowledge graph question answering, in: Findings of the Association for Computational Linguistics: EMNLP 2024, 2024, pp. 8972–8985.

[66] Y. Su, X. Han, Z. Zhang, Y. Lin, P. Li, Z. Liu, J. Zhou, M. Sun, Cokebert: Contextual knowledge selection and embedding towards enhanced pre-trained language models, AI Open 2 (2021) 127–134.

[67] Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, Q. Liu, ERNIE: Enhanced language representation with informative entities, in: A. Korhonen, D. Traum, L. Màrquez (Eds.), Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational

Linguistics, Florence, Italy, 2019, pp. 1441–1451. URL: https://aclanthology.org/P19-1139/. doi:`10.18653/v1/P19-1139`.

[68] Y. Gu, S. Kase, M. Vanni, B. M. Sadler, P. Liang, X. Yan, Y. Su, Beyond I.I.D.: three levels of generalization for question answering on knowledge bases, in: J. Leskovec, M. Grobelnik, M. Najork, J. Tang, L. Zia (Eds.), WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021, ACM / IW3C2, 2021, pp. 3477–3488. URL: https://doi.org/10.1145/3442381.3449992. doi:`10.1145/3442381.3449992`.

[69] B. P. Bhuyan, A. Ramdane-Cherif, R. Tomar, T. Singh, Neuro-symbolic artificial intelligence: a survey, Neural Computing and Applications 36 (2024) 12809–12844.

[70] S. Hatem, G. Khoriba, M. H. Gad-Elrab, M. ElHelw, Up to date: Automatic updating knowledge graphs using llms, Procedia Computer Science 244 (2024) 327–334. URL: https://www.sciencedirect.com/science/article/pii/S1877050924030072. doi:`https://doi.org/10.1016/j.procs.2024.10.206`, 6th International Conference on AI in Computational Linguistics.

[71] X. Zhu, Z. Li, X. Wang, X. Jiang, P. Sun, X. Wang, Y. Xiao, N. J. Yuan, Multi-modal knowledge graph construction and application: A survey, IEEE Transactions on Knowledge and Data Engineering 36 (2024) 715–735. URL: http://dx.doi.org/10.1109/TKDE.2022.3224228. doi:`10.1109/tkde.2022.3224228`.

[72] X. Wang, B. Meng, H. Chen, Y. Meng, K. Lv, W. Zhu, Tiva-kg: A multimodal knowledge graph with text, image, video and audio, in: Proceedings of the 31st ACM International Conference on Multimedia, MM '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 2391–2399. URL: https://doi.org/10.1145/3581783.3612266. doi:`10.1145/3581783.3612266`.