Volume 11, Issue 4, April – 2026 International Journal of Innovative Science and Research Technology
ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr2488

Autonomous KGs for Large Language Model Agents in Industry: A Systematic Review

Asma Hamad Alharbi[1]; Hafiz Farooq Ahmad[2]

[1,2]Department of Computer Science, College of Computer Science and Information Technology,
King Faisal University, Alhasa, Saudi Arabia

Publication Date: 2026/05/15

Abstract: The integration of large language models (LLMs) with knowledge graphs (KGs) has emerged as a promising approach to enhance reasoning, explainability, and decision-making in intelligent systems. However, LLMs inherently lack structured knowledge representation, leading to limitations such as hallucinations and reduced reliability, particularly in industrial applications. This study presents a systematic review of autonomous knowledge graphs for LLM-based agents, focusing on their role in enabling knowledge-driven and agentic AI systems in industry. Guided by PRISMA methodology, a comprehensive search across major scientific databases resulted in 33 relevant studies, which were classified into four research dimensions: architectures, KG construction and evolution, KG-based reasoning, and industrial applications. The findings reveal that multi-agent architectures and GraphRAG-based approaches dominate current research, while LLMs enable automated and dynamic knowledge graph lifecycle management. Furthermore, applications in manufacturing, digital twins, and industrial IoT demonstrate significant improvements in efficiency and decision-making. Despite these advancements, challenges related to scalability, computational cost, domain dependency, and lack of real-world validation persist. This review concludes that autonomous knowledge graphs play a critical role in advancing LLM-based agents toward reliable and scalable industrial deployment, while highlighting key research gaps and future directions for developing robust, explainable, and industry-ready AI systems.

Keywords: *Autonomous Knowledge Graphs ; LLM-based Agents ; Multi-Agent Systems; Graph Retrieval-Augmented Generation (GraphRAG) ; Industrial Artificial Intelligence; Smart Manufacturing.*

How to Cite: Asma Hamad Alharbi; Hafiz Farooq Ahmad (2026) Autonomous KGs for Large Language Model Agents in Industry: A Systematic Review. *International Journal of Innovative Science and Research Technology*, 11(4), 4690-4702. https://doi.org/10.38124/ijisrt/26apr2488

I. INTRODUCTION

The rapid evolution of artificial intelligence has largely been fueled by developments in the area of large language models, which demonstrate impressive results in natural language processing. Large language models (LLMs) like GPT, PaLM, and others allow machines to understand natural language and produce it, enabling their use across an ever-growing range of applications in healthcare, finance, manufacturing, and software engineering (Lin et al., 2025). However, despite their impressive capabilities, LLMs still operate as statistical models that predict sequences based on probabilities learned from large-scale text corpora.

As a result, they lack structured and up-to-date representations of knowledge, leading to issues such as hallucination, lack of explainability, and limitations in complex reasoning tasks (Lin et al., 2025; Kazlaris et al., 2025). These challenges are particularly critical in industrial applications, where accuracy, transparency, and knowledge-driven decision-making are essential.

An effective approach to addressing these shortcomings is the integration of LLMs with knowledge graphs (KGs), which provide structured representations of entities and their relationships. This combination enables explicit modeling of domain knowledge, improves reasoning capabilities, and enhances interpretability, which are crucial in industrial environments (Murali et al., 2025). Recent developments have moved beyond using KGs as simple external augmentations toward more advanced frameworks that tightly integrate LLMs and KGs in a synergistic manner. In particular, the concept of autonomous knowledge graphs has emerged, where KGs are not only consumed by LLMs but also constructed, updated, and refined through LLM-driven processes (Bian, 2025; Chen & Liu, 2024; Li et al., 2025). This paradigm enables a deeper integration of symbolic knowledge and generative intelligence.

In parallel, the development of LLM-based agents has gained significant attention. These agents extend traditional LLM capabilities by incorporating components such as memory, planning, tool usage, and collaboration mechanisms. Knowledge graphs play a critical role in such systems by serving as structured long-term memory, reasoning substrates, and coordination mechanisms among multiple agents (Liu et al., 2025; Tran et al., 2025; Kostka & Chudziak, 2024; Wang et al., 2025). The convergence of autonomous knowledge graphs and LLM-based agents offers a promising pathway toward building intelligent systems capable of continuous learning, adaptive reasoning, and real-world decision-making (Zhou et al., 2024; Lumer et al., 2025).

Several paradigms characterize the integration of LLMs and knowledge graphs. First, retrieval-augmented generation (RAG) incorporates external knowledge sources to improve the factual accuracy of LLM outputs, while its extension, GraphRAG, leverages graph structures to enable more context-aware retrieval (Zhang et al., 2025; Liang et al., 2025). Second, LLMs are increasingly used for automated knowledge graph construction by extracting entities, relationships, and schemas from unstructured data (Chen & Liu, 2024; Ho et al., 2024; Aamer et al., 2025). Third, multi-agent systems facilitate collaboration among multiple LLM agents through shared knowledge representations, improving reasoning and coordination (Tan et al., 2026; Jia et al., 2025; Yu et al., 2025). Finally, neuro-symbolic and hybrid approaches aim to combine neural learning with symbolic reasoning to enhance system interpretability and performance (Pussadeniya et al., 2024; Xie et al., 2025).

Despite these advancements, several limitations persist. Many current approaches rely on static or partially updated knowledge graphs, which may not adequately reflect dynamic industrial environments. Knowledge graph construction remains challenging, particularly when dealing with noisy, incomplete, or heterogeneous data sources. Furthermore, coordinating LLM agents in multi-agent systems introduces challenges related to scalability, communication efficiency, and system stability (Wang et al., 2024; Zhou et al., 2025). Additionally, there is a lack of standardized evaluation frameworks for assessing LLM–KG systems, making it difficult to compare performance across studies.

Beyond technical challenges, there is a notable gap in the literature regarding systematic reviews that specifically address the intersection of autonomous knowledge graphs, LLM agents, and industrial applications. While existing surveys have explored individual aspects such as LLM agents, knowledge graph construction, and hallucination mitigation, they often treat these areas in isolation (Lin et al., 2025; Liu et al., 2025; Zhang et al., 2025). Moreover, most studies focus on general-purpose applications rather than industrial settings, which involve unique requirements such as real-time processing, safety constraints, scalability, and domain specificity. Recent works exploring industrial applications, such as intelligent maintenance systems, further highlight the need for integrated perspectives in this domain (Lin et al., 2026). The lack of a unified and comprehensive review addressing these intersecting aspects represents a significant limitation in the current body of knowledge.

There is a need for a systematic review of the current research on autonomous KGs used for LLM-based agents in industrial applications. Systematic review is a crucial step in synthesizing existing knowledge, categorizing approaches, and comprehending the field. By systematically analyzing the literature, it will be possible to identify the prevailing architectural patterns and assess the efficiency of different methodologies. Moreover, it will be possible to detect the major challenges that should be addressed to achieve practical implementation. Systematic review can act as a foundation for further research by identifying gaps and suggesting directions.

In this study, the authors intend to provide a systematic literature review of autonomous KGs combined with LLM agents used in industrial settings. The main goal of this paper is to answer the following research questions:

- What are the architectures and frameworks proposed for the combination of KGs and LLM agents?
- How do KGs get autonomously created, maintained, and evolve?
- How do KGs enhance the reasoning and decision-making abilities of LLM agents?
- What industrial applications have been designed using these technologies?

This study contributes to the current literature in several ways. Firstly, it proposes a unified taxonomy that classifies the existing approaches according to research themes. Secondly, it offers a mutually exclusive classification of the studies based on the principal contribution. Thirdly, it synthesizes the findings across a wide range of applications to demonstrate how autonomous KG-based LLMs agents are implemented in different industrial domains. Fourthly, it highlights the key challenges and limitations associated with these technologies, such as scalability, data quality, system integration, and evaluation. Lastly, it suggests future research directions for the advancement of the field.

II. REVIEW METHODOLOGY

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size. If you are using US letter-sized paper, please close this file and download the file “MSW_USltr_format”.

This study adopts a systematic literature review (SLR) methodology aligned with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure methodological rigor, transparency, and reproducibility. The review process follows the four standard PRISMA phases—identification, screening, eligibility, and inclusion—to systematically collect, filter, and analyze relevant studies on autonomous KGs for LLMs-based agents in industrial contexts.

The review is guided by four research questions designed to capture the key dimensions of the field: RQ1: What architectures and frameworks have been proposed for integrating KGs with LLM-based agents? RQ2: How are

KGs autonomously constructed, updated, and evolved using LLMs? RQ3: How do KGs enhance reasoning and decision-making capabilities of LLM agents? RQ4: What industrial applications have been developed using autonomous knowledge graph-based LLM agents? These research questions were carefully formulated to ensure comprehensive coverage of system design, knowledge lifecycle, reasoning capabilities, and real-world applications. Furthermore, each study was assigned to a single research question based on its primary contribution to avoid overlap and ensure analytical clarity.

A comprehensive search strategy was developed to capture relevant literature across multiple academic and technical repositories. The search was conducted using the following major databases: Scopus, Web of Science, IEEE Xplore, ACM Digital Library, ScienceDirect, arXiv, and Google Scholar. These platforms were selected to ensure coverage of both peer-reviewed publications and emerging research outputs. The inclusion of arXiv and Google Scholar was particularly important due to the rapid pace of advancements in LLM-based systems, where many influential studies are initially منتشر as preprints prior to formal peer-reviewed publication.

The search query was constructed using Boolean operators to combine key concepts related to LLMs, KGs, agents, and industrial applications:

(“large language model” OR “LLM” OR “generative AI”) AND (“knowledge graph” OR “semantic graph” OR “graph-based knowledge”) AND (“agent” OR “autonomous agent” OR “multi-agent system”) AND (“industry” OR “manufacturing” OR “enterprise” OR “industrial AI”)

Additional keywords such as GraphRAG, autonomous knowledge graph, agentic systems, digital twins, industrial intelligence, and retrieval-augmented generation were incorporated to enhance coverage. The search was restricted to studies published between 2024 and 2026, reflecting the recent emergence of LLM-agent paradigms.

The identification phase resulted in a total of 1,248 records, distributed across databases as follows: Scopus (312), Web of Science (268), IEEE Xplore (221), ACM Digital Library (147), ScienceDirect (96), arXiv (168), and Google Scholar (36). After removing duplicate records (n = 286), 962 unique studies remained for further screening. During the title and abstract screening phase, 681 studies were excluded due to irrelevance to the research scope, including studies not addressing LLMs, KGs, or agent-based systems, as well as those lacking industrial context. This resulted in 281 studies being selected for full-text assessment.

In the eligibility phase, the full texts of the remaining studies were evaluated against predefined inclusion and exclusion criteria. A total of 248 studies were excluded for the following reasons: absence of autonomous or agent-based components (102 studies), purely conceptual or survey-based contributions without implementation (61 studies), lack of industrial relevance (55 studies), and insufficient methodological detail (30 studies). Following this rigorous filtering process, 33 studies were deemed eligible and included in the final qualitative synthesis.

The inclusion and exclusion criteria used in this review are presented in Table 1.

Table 1 Inclusion and Exclusion Criteria

Criterion Type	**Inclusion Criteria**	**Exclusion Criteria**
Scope	Integration of LLMs with KGs and/or agent-based systems	Studies focusing only on LLMs or only on KGs
Autonomy	Autonomous or semi-autonomous agent-based systems	Static or non-agentic systems
Application	Industrial, enterprise, or real-world use cases	Purely theoretical or academic-only studies
Publication Type	Peer-reviewed journals, conferences, and high-quality preprints	Blogs, whitepapers, non-academic sources
Language	English	Non-English
Time Frame	2024–2026	Before 2024

To ensure the relevance and timeliness of the review, conference papers and preprints were included alongside journal publications. This decision is justified by the nature of the research domain, where cutting-edge developments in LLMs, multi-agent systems, and KGs are frequently first introduced in top-tier conferences (e.g., NeurIPS, ACL, ICML, AAAI, IEEE) or shared via preprint platforms such as arXiv. Excluding these sources would risk omitting significant recent contributions and biasing the review toward slower publication cycles. Therefore, only high-quality conference papers and preprints with clear methodologies and experimental validation were considered.

The overall study selection process is summarized using a PRISMA flow diagram in Figure 1.

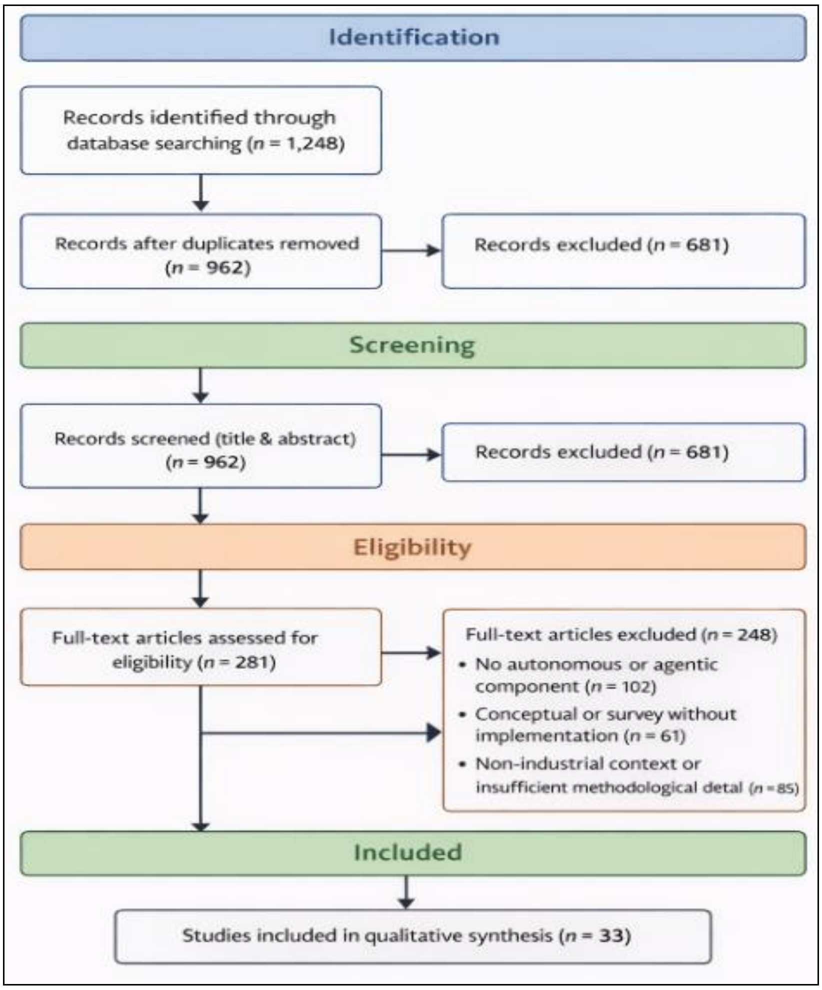

Fig 1 PRISMA Flow Diagram

To ensure the reliability and rigor of the selected studies, a structured quality assessment and risk of bias evaluation was conducted for all 33 included papers. Given the interdisciplinary and emerging nature of research on LLM-based agents and KGs, a customized quality assessment framework was adopted, inspired by established guidelines in systematic reviews in software engineering and artificial intelligence.

Each study was evaluated across five key quality dimensions: (1) clarity of research objectives, (2) methodological rigor, (3) completeness of system description, (4) evaluation and experimental validation, and (5) relevance to industrial applications. These criteria were selected to reflect both the technical depth and practical applicability of the studies. Each criterion was scored on a three-point scale: 0 (not addressed), 0.5 (partially addressed), and 1 (fully addressed), resulting in a maximum possible score of 5 per study.

To assess the risk of bias, additional factors were considered, including: dataset transparency, reproducibility of experiments, presence of comparative baselines, and potential overclaiming of results without sufficient empirical support. Studies lacking experimental validation or relying solely on conceptual frameworks were considered to have a higher risk of bias. Similarly, works that did not provide sufficient implementation details or relied on proprietary, non-reproducible datasets were flagged for moderate risk.

➢ *Based on the Aggregated Quality Scores, Studies were categorized into Three Tiers:*

- High quality (score ≥ 4.0) – well-defined methodology, strong experimental validation, and clear industrial relevance
- Moderate quality (score between 2.5 and 3.5) – partial validation or limited experimental detail
- Low quality (score < 2.5) – insufficient methodological clarity or lack of empirical support

The overall assessment indicated that the majority of included studies fell within the moderate to high-quality range, reflecting the maturity of recent research in this domain. However, a notable proportion of studies exhibited moderate risk of bias, primarily due to limited reproducibility and lack of standardized evaluation benchmarks, which is consistent with the early-stage nature of this research field. To further enhance the robustness of the review, quality scores were not used as exclusion criteria, but rather as a basis for critical analysis and interpretation of findings. This approach ensures that emerging but potentially impactful contributions are not overlooked, while still maintaining transparency regarding their limitations.

III. RESULTS

The 33 selected studies were classified into four mutually exclusive categories based on their primary contribution: RQ1 (Architecture & Framework Design), RQ2 (Knowledge Graph Construction & Evolution), RQ3 (KG-based Reasoning & Decision-Making), and RQ4 (Industrial Applications). Each study was assigned to a single category to ensure analytical clarity and avoid redundancy. The results reveal a strong trend toward multi-agent architectures, automated knowledge graph lifecycle management, and domain-specific industrial deployments, particularly in manufacturing and enterprise systems.

- *RQ1: Architecture & Framework Design*

From an examination of the research papers related to RQ1, it has been observed that multi-agent system architectures have been identified as the dominant way to combine large language models with KGs. There has been one notable architectural pattern across various studies, which includes the breakdown of complex workflows using dedicated agents responsible for performing knowledge acquisition, verification, and inference. As an instance, the KARMA architecture makes use of a hierarchical orchestration of nine agents in order to minimize errors and conflict resolution during knowledge graph enrichment (Lu et al., 2025). Additionally, CoMA-IKG is co-evolutionary multi-agent architecture used for building industrial KGs (Zhang et al., 2026). In addition to that, Agentic-KGR applies reinforcement learning techniques to allow co-evolution in multi-agent and knowledge graph systems, resulting in significant improvements in the quality of KGs (Li et al., 2025).

One key architectural pattern that has emerged from this set of studies is the design of KGs as structures of dynamic memories. The use of knowledge graph as memory substrates has been illustrated by KG-Agent (Jiang et al., 2025), and AriGraph (Anokhin et al., 2024), whereby such techniques improve multi-hop reasoning abilities and chances of successful completion of tasks. Moreover, KGs are now being designed for specific domains including MAKES-QA (Borrelli et al., 2026) as well as intelligent manufacturing systems (Zhao et al., 2026; Meng et al., 2025).

Despite the above developments, however, there are still certain limitations. Many of the proposed frameworks need simulations and data from restricted domains (Anokhin et al., 2024; Zhao et al., 2026). Further, there have been a number of architectures that rely on prior ontologies and human interventions (Zhang et al., 2026; Lu et al., 2025).Table 2 underscore the studies under RQ1 that indicates a significant shift toward distributed, knowledge-driven AI architectures, where multi-agent coordination and graph-based memory are central design principles. These architectures lay the foundation for scalable and intelligent systems but require further validation in real-world industrial settings.

Table 2 Architecture and Framework Design

Study	Methodology	Metrics	Key Findings	Limitations
Martínez Rodríguez et al. (2025)	LLM + embeddings multi-agent KG construction & verification	Construction accuracy, verification rate	Multi-agent verification improves KG quality	Domain-specific (ADAS)
Lu et al. (2025)	KARMA: 9-agent hierarchical orchestration	Correctness (83.1%), conflict reduction (18.6%)	Multi-agent improves correctness by up to 14.4%	Needs human validation
Zhang et al. (2026)	CoMA-IKG multi-agent co-evolution framework	Entity-relation F1, completeness	Enables automated industrial KG evolution	Requires initial ontology
Li et al. (2025)	Agentic-KGR with reinforcement learning	Long-term graph quality (+37%)	Co-evolution improves KG quality	Long training time
Jiang et al. (2025)	KG-Agent autonomous reasoning agent	Multi-hop accuracy (78–85%)	Outperforms larger models	Static KG limitation
Anokhin et al. (2024)	AriGraph (KG world model + memory)	Task success (+28–41%)	Hybrid memory improves agent performance	Simulated only
Borrelli et al. (2026)	MAKES-QA multi-agent KG + QA system	QA accuracy (82%)	Dynamic KG supports scientific QA	Literature-focused
Zhao et al. (2026)	Multi-agent manufacturing system	Efficiency (+42%)	End-to-end shopfloor automation	Simulation-heavy
Meng et al. (2025)	LLM + KG intelligent control framework	Decision speed (+27%)	Ontology-aligned control improves efficiency	Simulated only

Figure 2 illustrates the architecture of LLM–knowledge graph integration through a multi-agent framework. The diagram highlights how the LLM interacts with an ontology and a central knowledge graph, supported by specialized agents for extraction, reasoning, and validation. This modular design enables efficient coordination, where each agent performs a dedicated task, improving system scalability and accuracy. The knowledge graph acts as a structured memory layer, enhancing reasoning and contextual understanding. However, the figure also reflects key limitations, particularly the reliance on simulation environments and predefined structures, which may restrict adaptability in real-world industrial deployments

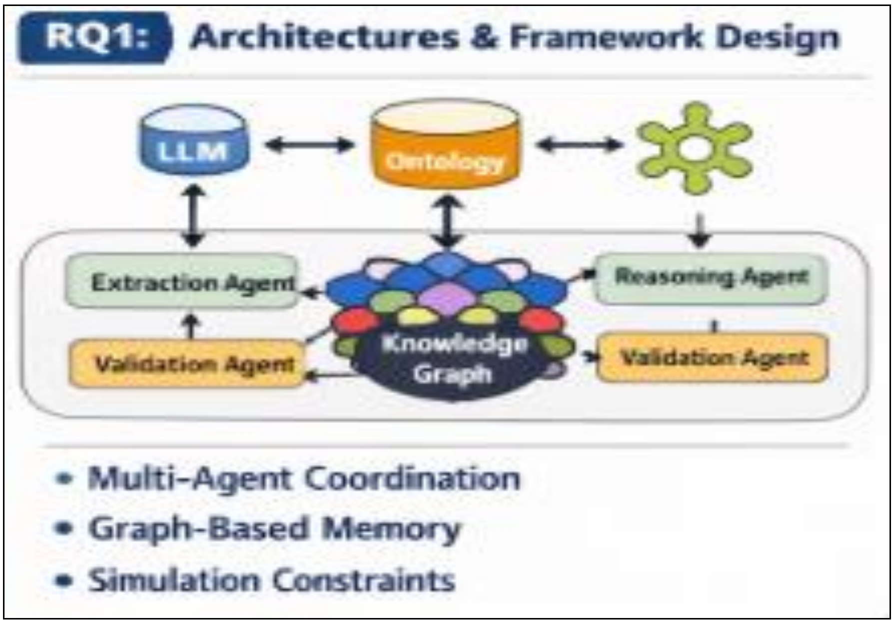

Fig 2 General Architecture and Framework

- ➢ *RQ2: KG Construction and Evolution*

The studies addressing RQ2 highlight significant progress in the automation of knowledge graph construction and lifecycle management using LLMs. A central finding is the ability of LLMs to perform end-to-end knowledge extraction and structuring tasks, including entity recognition, relation extraction, and schema induction. Bai et al. (2025) demonstrate that fully autonomous pipelines can achieve high triple precision and schema coverage while improving downstream reasoning tasks. Similarly, Tang et al. (2025) propose an integrated framework for knowledge graph construction, completion, and alignment, enabling comprehensive lifecycle management.

Another important trend is the emergence of dynamic and continuously evolving knowledge graphs, which are essential for handling real-time data. Hatem et al. (2024) show that LLM-based updating mechanisms can reduce manual effort by approximately 65%, while Wang et al. (2024) introduce temporal adaptation techniques that enhance reasoning over time-sensitive knowledge. These developments are particularly relevant in industrial contexts, where knowledge is constantly changing.

Several studies also emphasize the role of domain-specific knowledge graph construction, such as in manufacturing (Morioka et al., 2025; Yan et al., 2026), healthcare (Zhang et al., 2024), and finance (Lian et al., 2025). While these approaches demonstrate strong performance within their respective domains, they also reveal a key limitation: lack of generalizability across domains. Additionally, large-scale systems such as those proposed by Bai et al. (2025) require significant computational resources, which may hinder practical deployment.

Table 3 indicate a transition from static, manually curated knowledge graphs to autonomous, adaptive, and scalable knowledge systems. However, challenges related to computational cost, domain bias, and data quality remain critical barriers to widespread adoption.

Table 3 KG Construction and Evolution

Study	Methodology	Metrics	Key Findings	Limitations
Bai et al. (2025)	Autonomous LLM pipeline with dynamic schema induction	Triple F1 (>90%), schema coverage (92%)	Billion-scale KGs improve QA & factuality	High compute cost
Morioka et al. (2025)	LLM-based asset KG from FMEA	Cost reduction, expert quality	Reduces manual effort significantly	Depends on document quality
Tang et al. (2025)	End-to-end KG construction, completion, alignment	Alignment score	Unified lifecycle management	High compute
Hatem et al. (2024)	LLM-based KG updating	Manual effort reduction (~65%)	Enables continuous updates	Outdated triples possible
Wang etal. (2024)	Temporal KG adaptation with LLM	Temporal reasoning accuracy	Improves temporal reasoning	Not industrial-focused
Zhang et al. (2024)	LLM+graph embedding for medical KG	Completeness,fusion accuracy	Effective KG fusion	Domain-specific
Lian et al. (2025)	Multi-agent KG construction (credit domain)	Reasoning accuracy (+27%)	Cross-domain KG construction	Domain bias

Yan et al. (2026)	LLM + KG integration for machining	Recommendation accuracy (76%)	Automates machining knowledge base	Domain-specific

Figure 3 presents the automated KG pipeline highlighting the end-to-end process from data sources to temporal updating. The pipeline demonstrates how LLMs enable autonomous extraction, schema induction, and continuous evolution of knowledge graphs. This structured workflow emphasizes the transition from static to dynamic KGs, supporting real-time adaptation in data-intensive environments. The inclusion of temporal updating reflects the importance of maintaining up-to-date knowledge in industrial contexts. However, the figure also underscores key challenges, particularly the high computational cost associated with large-scale processing and the complexity of ensuring consistency during continuous updates.

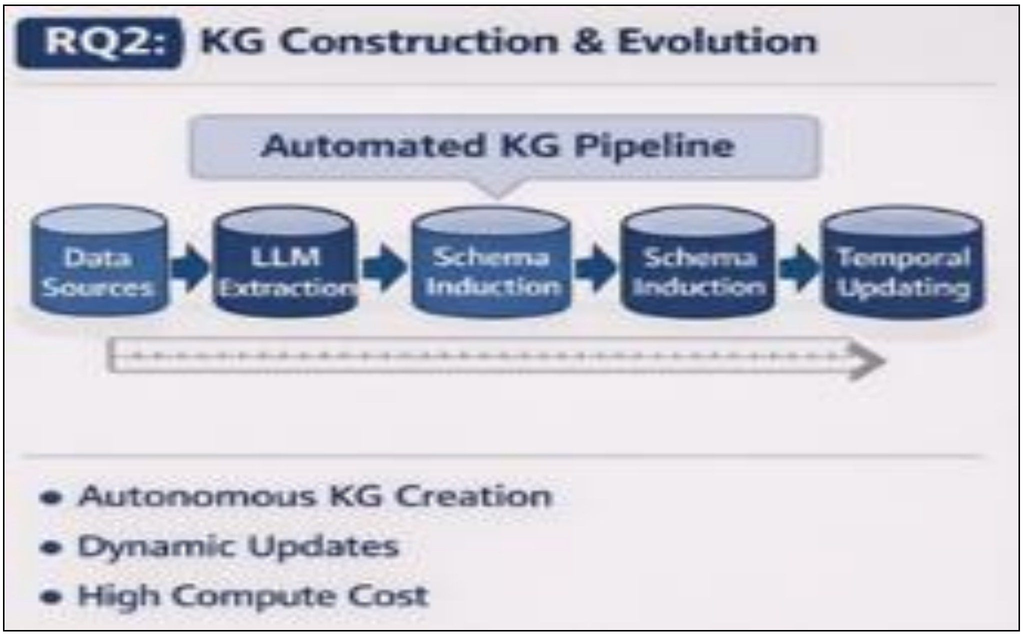

Fig 3 Automated KG Pipeline

- *RQ3: KG-Based Reasoning*

The studies under RQ3 demonstrate that knowledge graphs play a crucial role in enhancing the reasoning and decision-making capabilities of LLM-based agents. A key finding is that integrating structured knowledge significantly improves factual accuracy and reduces hallucinations. Guan et al. (2024) report a 62% reduction in hallucinations through knowledge graph-based retrofitting, highlighting the importance of post-generation validation mechanisms.

Another major trend is the adoption of GraphRAG approaches, which extend traditional retrieval-augmented generation by incorporating relational knowledge. Studies such as Knollmeyer et al. (2025) and Zang et al. (2026) show that GraphRAG significantly improves performance in domain-specific tasks, including manufacturing question answering and supply chain configuration. These approaches enable more context-aware retrieval and reasoning, particularly in complex, multi-step scenarios.

Autonomous exploration and reasoning over knowledge graphs also emerge as important capabilities. Systems such as EICopilot (Yun et al., 2025) and ARK (Polonuer et al., 2026) demonstrate improved search and exploration efficiency in large-scale enterprise knowledge graphs. Additionally, domain-specific applications, including carbon management (Wu et al., 2024) and warehouse optimization (Parekh et al., 2025), highlight the practical benefits of KG-enhanced reasoning in industrial decision-making.

However, several limitations persist. Many reasoning systems face scalability challenges, particularly when dealing with large or dynamic knowledge graphs (Yun et al., 2025; Polonuer et al., 2026). Additionally, domain-specific implementations limit general applicability, and some approaches introduce computational overhead during inference (Guan et al., 2024).

Figure 4 suggests that knowledge graphs serve as a critical reasoning backbone for LLM agents, enabling more accurate, explainable, and context-aware decision-making. However, scalability and efficiency remain key challenges.

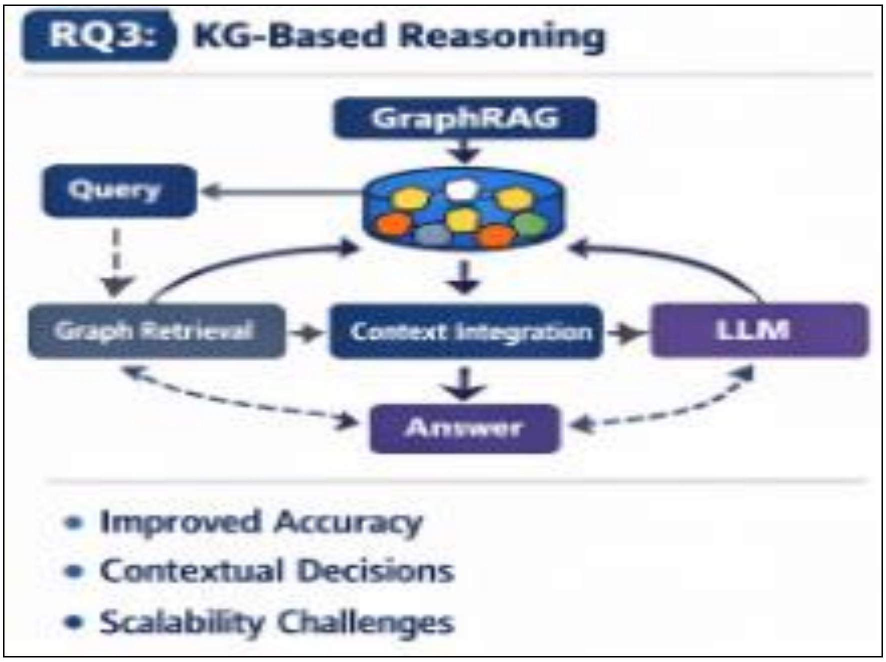

Fig 4 KG-Based Reasoning

Table 4 outlines the findings of the included studies, highlighting the recent KG-driven reasoning.

Table 4 KG-Based Reasoning Studies

Study	Methodology	Metrics	Key Findings	Limitations
Muhtadi et al. (2026)	Adaptive self-prompting agent	Predictive performance	Improves fault detection	Synthetic dataset
Guan et al. (2024)	KG-based retrofitting	Hallucination reduction (62%)	Reduces hallucinations significantly	Inference overhead
Yun et al. (2025)	KG exploration agent (EICopilot)	Search success (+33%)	Efficient enterprise retrieval	Scalability cost
Polonuer et al. (2026)	Adaptive KG exploration (ARK)	Exploration success (+29%)	Improves retrieval efficiency	Cost scaling
Knollmeyer et al. (2025)	GraphRAG for manufacturing QA	QA accuracy (+31%)	Outperforms vanilla RAG	Proprietary data
Zang et al. (2026)	Event-state KG GraphRAG	Configuration accuracy (+38%)	Improves supply chain decisions	Narrow domain
Wu et al. (2024)	ProcessCarbonAgent	Error reduction (19%)	Supports sustainability decisions	Carbon-specific
Parekh et al. (2025)	KG + LLM reasoning	Bottleneck accuracy (+33%)	Improves warehouse planning	Domain-specific

➢ *RQ4: Industrial Applications*

The studies under RQ4 demonstrate the growing adoption of LLM–KG systems in industrial applications, particularly in manufacturing, digital twins, and industrial IoT. A consistent finding is that these systems significantly improve operational efficiency, decision-making accuracy, and automation. For example, multi-agent manufacturing systems achieve up to 42% efficiency improvements (Zhao et al., 2026), while intelligent maintenance frameworks reduce downtime errors and improve diagnostic precision (Lin et al., 2026).

Digital twin technologies represent a major application area. Studies such as Gao et al. (2026) and Lou et al. (2025) show that integrating LLMs with knowledge graphs enables self-evolving and cognitive digital twins, which enhance situational awareness and human–machine collaboration. Similarly, anomaly detection systems in industrial IoT, such as SEMAS (Saleh et al., 2026) and Lemad (Ji et al., 2025), demonstrate improved performance in real-time monitoring and predictive maintenance.

Another important application domain is sustainability and optimization, where systems such as ProcessCarbonAgent (Wu et al., 2024) and CausalGID (Wu et al., 2026) support carbon management and generative design. These applications highlight the potential of LLM–KG systems to address complex industrial challenges.

Despite these promising results, several limitations are evident. Many systems are domain-specific, limiting their transferability across industries (Ji et al., 2025; Wu et al., 2024). Additionally, several studies rely on simulated environments or early-stage prototypes, raising concerns about real-world deployment (Zhao et al., 2026; Lou et al., 2025). Data quality and sensor reliability also remain critical challenges (Gao et al., 2026).

Table 5 indicate that while LLM–KG systems are beginning to demonstrate real-world impact, their adoption is still in an early and domain-constrained stage, requiring further validation and standardization.

Table 5 Industrial Applications Studies

Study	Methodology	Metrics	Key Findings	Limitations
Mao et al. (2025)	Multi-agent KG + QA system	QA accuracy	Supports industrial knowledge integration	Privacy concerns
Lin et al. (2026)	Multi-agent GraphRAG maintenance system	Precision (+35%), downtime reduction	Improves maintenance accuracy	Single equipment
Gao et al. (2026)	Self-evolving digital twin	Situational awareness (94%)	Enables adaptive manufacturing	Sensor dependency
Lou et al. (2025)	Cognitive digital twin with LLM	Collaboration (+33%)	Enhances human-machine interaction	Early prototype
Ji et al. (2025)	Multi-agent anomaly detection	Precision	Improves power grid monitoring	Domain-specific
Gong et al. (2025)	Multi-agent sensor reasoning	Reasoning accuracy	Collective intelligence improves detection	Sensor-specific
Wu et al. (2026)	Causal generative design (CausalGID)	Design quality	Improves industrial design optimization	Not KG-centric
Saleh et al. (2026)	SEMAS hierarchical multi-agent system	F1, latency, stability	Real-time edge anomaly detection	No explicit limitations

Figure 5 illustrates the diverse industrial applications of LLM–knowledge graph systems, highlighting domains such as smart manufacturing, digital twins, anomaly detection, and sustainability optimization. The figure emphasizes the strong alignment of these technologies with Industry 4.0 and Industry 5.0 paradigms, where intelligent systems support automation, monitoring, and decision-making. The presence of multiple application areas reflects the versatility of LLM–KG integration across industrial processes. However, the figure also indicates that most solutions are domain-specific, requiring tailored implementations. Additionally, challenges related to data privacy, system reliability, and integration with existing industrial infrastructures remain critical considerations for real-world deployment.

Fig 5 Types of Industrial Applications

➢ *Research Gaps and Future Directions*

• *Research Gaps*

Despite the rapid advancements in integrating autonomous knowledge graphs with LLM-based agents, several critical research gaps remain across architecture design, knowledge graph lifecycle management, reasoning capabilities, and industrial deployment.

A major gap lies in the lack of real-world validation of proposed architectures. While many studies propose sophisticated multi-agent frameworks and knowledge-driven systems, a significant portion of these approaches are evaluated in simulated environments or constrained datasets (e.g., manufacturing simulations, synthetic CPS data). This raises concerns regarding scalability, robustness, and deployment readiness in real industrial settings. Furthermore, several architectures rely on predefined ontologies or domain-specific configurations, limiting their adaptability across heterogeneous industrial domains.

Another important gap is the high computational cost associated with autonomous knowledge graph construction and maintenance. Large-scale systems, such as those leveraging web-scale corpora, require extensive computational resources, making them difficult to deploy in resource-constrained industrial environments. Additionally, continuous updating mechanisms, while promising, introduce challenges related to data consistency, latency, and outdated knowledge, particularly in rapidly changing systems.

From a reasoning perspective, although knowledge graphs significantly improve accuracy and reduce hallucinations, there is a lack of standardized evaluation benchmarks for KG-enhanced reasoning in LLM agents. Existing studies use diverse metrics such as QA accuracy, reasoning success rate, or error reduction, making cross-comparison difficult. Moreover, current approaches often face scalability limitations when applied to large and dynamic knowledge graphs, particularly in enterprise-scale environments.

In terms of industrial applications, the findings reveal that most systems are highly domain-specific, focusing on areas such as manufacturing, digital twins, or energy systems. This indicates a lack of generalizable frameworks that can be easily adapted across industries. Additionally, issues such as data quality, sensor reliability, privacy, and integration with legacy systems remain underexplored, despite being critical for real-world adoption.

Finally, there is limited attention to trustworthiness, explainability, and governance in autonomous KG–LLM systems. While knowledge graphs inherently support interpretability, few studies systematically evaluate or standardize explainability mechanisms, especially in safety-critical industrial environments.

• *Research Gaps*

Addressing these gaps opens several promising directions for future research. First, there is a need for real-world deployment and benchmarking of LLM–KG systems in industrial environments. Future studies should focus on large-scale pilot implementations, enabling the evaluation of system performance under real operational constraints, including latency, reliability, and integration complexity. Figure 6 reflects the future directions of LLM developments.

Fig 6 Future Roadmap

Second, research should aim to develop lightweight and resource-efficient approaches for KG construction and updating, particularly for edge and industrial IoT environments. Techniques such as incremental learning, distributed processing, and hybrid symbolic-neural optimization could help reduce computational overhead.

Third, the field would benefit from the development of standardized evaluation frameworks and benchmarks for KG-enhanced reasoning. Establishing common metrics for reasoning accuracy, explainability, scalability, and robustness would enable more consistent comparison across studies and accelerate progress.

Another important direction is the design of domain-agnostic and transferable frameworks, which can generalize across industries. This includes developing adaptable ontologies, modular architectures, and cross-domain knowledge integration techniques to reduce reliance on domain-specific configurations.

Furthermore, future research should focus on scalable reasoning over dynamic knowledge graphs, particularly for enterprise-scale and real-time applications. This includes optimizing GraphRAG approaches, improving graph traversal efficiency, and enabling continuous reasoning in evolving environments.

Finally, there is a critical need to address trust, explainability, and governance in LLM–KG systems. Future work should incorporate explainable AI techniques, provenance tracking, and audit mechanisms to ensure transparency and accountability. This is particularly important for safety-critical applications such as manufacturing, energy systems, and autonomous decision-making.

IV. DISCUSSIONS

This systematic review presents an overview of the current state of autonomous knowledge graphs with LLM agents, with special emphasis placed on the use cases in the industrial environment. In accordance with the research questions, the following four dimensions have been considered throughout the review: architectures, the knowledge graph life cycle, reasoning and decision making, and industrial use cases. Taken together, the findings reveal the trend toward knowledge-driven and agentic AI systems, which incorporate knowledge graphs as the building blocks for reasoning, decision making, and coordination of activities.

Multi-agent architectures are among the most significant contributions of the studies in question. These architectures provide a framework for decomposing knowledge-intensive tasks into smaller chunks, allowing for a scalable and modular design of LLM-based systems. Moreover, embedding of the knowledge graph in such architectures serves to facilitate context-based decision making, multi-hop reasoning, and explainability of decisions, which are especially important characteristics in industrial settings.Another key development is the progress toward automatic knowledge graph life cycle operations. The reviewed studies reveal that LLM-based agents are capable of performing entity recognition, relation extraction, schema induction, and graph alignment operations without extensive knowledge engineering. Besides, the recent advances in constructing dynamic knowledge graphs are especially relevant to industry, since the systems must adapt to continuous data changes in time-critical settings. Nonetheless, the high computational demands associated with the automatic life cycle operations are also worth mentioning.

When it comes to reasoning and decision-making processes, the use of knowledge graphs brings multiple improvements to the LLM systems. GraphRAG, as well as knowledge graph-based retrofitting and fine-tuning techniques, allow for mitigating the problem of LLM hallucinations. Knowledge graphs also help with causal reasoning and modeling dependencies between the entities, which is necessary for complex reasoning processes. On the other hand, scalability issues and the problem of inference latency in graph-based decision making are also worth noting.

In addition to the previously listed contributions, a significant development in the field is the successful use of knowledge-driven LLMs in industrial applications. According to the findings, the discussed systems show great promise in fields like smart manufacturing, digital twins, predictive maintenance, anomaly detection, and sustainability optimization. Examples of the corresponding systems include multi-agent manufacturing systems and intelligent maintenance frameworks, which exhibit better performance in terms of accuracy, precision, and downtime. However, most solutions are still tailored to specific industries and evaluated using simulations rather than in production settings.

Given that the field of autonomous knowledge graphs with LLM agents evolves at a rapid pace, numerous studies are not peer-reviewed but remain preprint papers or conference submissions. Although an effort has been made to include only reputable sources into the review, this limitation may lead to variations in methodological rigor and quality control criteria. In addition, the variety of the metrics used in the research, ranging from basic precision and F1 scores to various industrial performance indicators, hinders direct comparison of results and drawing generalized conclusions regarding the effectiveness of LLMs.

Besides these issues, another possible source of bias in this systematic review may lie in the selection of relevant studies and their classification. Although the review employs a methodology based on the PRISMA principles, the allocation of papers to research questions depends on the main contribution. Thus, some studies might be excluded from consideration due to being focused on one aspect of the research while paying insufficient attention to others. Another possible issue lies in focusing on the publications between 2024 and 2026.

A more fundamental limitation of this research is connected to the lack of deployment and validation of the systems in actual industrial settings. While many reviewed studies use synthetic datasets or simulate industrial settings, very few papers describe large-scale implementation and testing of LLMs in the production setting. As such, these studies do not fully cover all potential issues arising during operation, which poses a serious problem for researchers and industry practitioners alike. Based on the results obtained throughout the systematic review, there are multiple promising research directions in the field of knowledge-driven LLMs for industrial applications. First and foremost, real-world implementation and evaluation of autonomous knowledge graphs with LLM agents are crucial. These studies could assess the efficiency, scalability, latency, and integration of the systems under consideration in the real-world conditions.

A closely related topic concerns the need to develop efficient methods for building autonomous knowledge graphs, as well as for updating them. It includes the approaches like incremental learning, distribution processing, and edge computing for knowledge graphs. Another important step would involve creating robust updating procedures for knowledge graphs. Another future direction of research is concerned with the standardization of the evaluation process.

It seems reasonable to define commonly used metrics and benchmarks to facilitate cross-paper comparisons. They can include such aspects as reasonability, explainability, scalability, and industrial applicability. Furthermore, there is an urgent need for developing domain-agnostic and highly transferable knowledge graphs for industrial applications. Currently, the systems are tailored to a particular industry or dataset, which limits their generalizability.Lastly, the issue of trust, explainability, and regulation becomes more pressing in LLM–knowledge graph systems. With an ever-expanding use of these systems in industry, measures need to be taken in order to ensure transparency, traceability, and validity.

V. CONCLUSION

This systematic review has examined the emerging field of autonomous knowledge graphs for LLM-based agents, with a particular focus on industrial applications. By analyzing 33 recent studies, the review provides a structured understanding of how knowledge graphs are being integrated with LLMs to enhance system architecture, automate knowledge lifecycle processes, improve reasoning capabilities, and enable real-world industrial solutions. The findings highlight that multi-agent architectures and knowledge graph-centered designs are becoming the dominant paradigm, enabling scalable, modular, and knowledge-driven AI systems. Furthermore, the review demonstrates that LLMs play a critical role in automating knowledge graph construction and evolution, significantly reducing manual effort while enabling dynamic and adaptive knowledge systems. The integration of knowledge graphs also enhances reasoning and decision-making, particularly through approaches such as GraphRAG, which improve contextual understanding and reduce hallucinations. In industrial contexts, applications in smart manufacturing, digital twins, and anomaly detection show promising improvements in efficiency, accuracy, and automation. However, several challenges remain, including high computational costs, scalability issues, domain dependency, and limited real-world validation. These challenges indicate that while the field has made significant progress, it is still transitioning from experimental prototypes to fully deployable industrial systems. Ultimately, autonomous knowledge graphs represent a crucial advancement in the development of reliable and explainable LLM-based agents. Addressing the identified challenges and advancing toward scalable, domain-agnostic, and industry-ready solutions will be essential for realizing the full potential of knowledge-driven AI systems in complex industrial environments.

REFERENCES

[1]. N. Aamer, M. N. Asim, S. Munir, and A. Dengel, “Automating AI discovery for biomedicine through knowledge graphs and LLM agents,” bioRxiv, 2025.
[2]. P. Anokhin et al., “AriGraph: Learning knowledge graph world models with episodic memory for LLM agents,” arXiv preprint arXiv:2407.04363, 2024.
[3]. J. Bai et al., “AutoSchemaKG: Autonomous knowledge graph construction through dynamic schema induction from web-scale corpora,” arXiv preprint, 2025.
[4]. H. Bian, “LLM-empowered knowledge graph construction: A survey,” arXiv preprint, 2025.
[5]. A. Borrelli et al., “MAKES-QA: A multi-agent framework for knowledge graph construction and enrichment over scientific literature for question answering,” Expert Systems with Applications, 2026.
[6]. G. Chen and X. Liu, “A multi-agent collaborative framework for constructing knowledge graphs from text,” in Proc. IEEE Int. Conf. Knowledge Graph (ICKG), 2024.
[7]. D. Gao et al., “Self-evolving digital twins in flexible manufacturing: A large language model-powered multi-agent approach,” Digital Engineering, 2026.
[8]. W. Gong et al., “Harnessing collective intelligence of multi-agent LLM systems for sensor failure reasoning in smart manufacturing,” Journal of Industrial Information Integration, 2025.
[9]. X. Guan et al., “Mitigating large language model hallucinations via autonomous knowledge graph-based retrofitting,” in Proc. AAAI Conf. Artificial Intelligence, 2024.
[10]. S. Hatem et al., “Up to date: Automatic updating knowledge graphs using LLMs,” Procedia Computer Science, 2024.
[11]. D. H. Ho et al., “Leveraging multi-agent systems and large language models for diabetes knowledge graphs,” in Proc. IEEE Big Data, 2024.
[12]. X. Ji et al., “Lemad: LLM-empowered multi-agent system for anomaly detection in power grid services,” Electronics, 2025.
[13]. Z. Jia et al., “Enhancing multi-agent systems via reinforcement learning with LLM-based planner and graph-based policy,” in Proc. IEEE ICRA, 2025.
[14]. J. Jiang et al., “KG-Agent: An efficient autonomous agent framework for complex reasoning over knowledge graph,” in Proc. ACL, 2025.
[15]. I. Kazlaris et al., “From illusion to insight: A taxonomic survey of hallucination mitigation techniques in LLMs,” AI, 2025.
[16]. S. Knollmeyer, O. Caymazer, and D. Grossmann, “Document GraphRAG: Knowledge graph-enhanced retrieval augmented generation for document question answering within the manufacturing domain,” Electronics, 2025.
[17]. A. Kostka and J. Chudziak, “Synergizing logical reasoning, knowledge management and collaboration in multi-agent LLM system,” in Proc. PACLIC, 2024.
[18]. J. Li et al., “Agentic-KGR: Co-evolutionary knowledge graph construction through multi-agent reinforcement learning,” arXiv preprint, 2025.
[19]. Y. Li et al., “Harnessing diverse perspectives: A multi-agent framework for enhanced error detection in knowledge graphs,” in Proc. DASFAA, 2025.
[20]. G. Lian and H. Chan, “Industry-adaptive knowledge graph construction with multi-agent systems for cross-domain enterprise credit assessment,” in Proc. ICISAI, 2025.
[21]. S. Liang, Y. Zhang, and Y. Guo, “PersonaAgent with GraphRAG: Community-aware knowledge graphs for personalized LLM,” arXiv preprint, 2025.
[22]. J. Lin et al., “A multi-agent and synergistic knowledge graph retrieval-augmented generation

framework for intelligent maintenance," Journal of Manufacturing Systems, 2026.

[23]. X. Lin et al., "LLM-based agents suffer from hallucinations: A survey," arXiv preprint, 2025.

[24]. Y. Liu et al., "Graph-augmented large language model agents: Current progress and future prospects," arXiv preprint, 2025.

[25]. S. Lou et al., "A cognitive digital twin for Industry 5.0 based on a large language model agent," in Proc. IEEE ICHMS, 2025.

[26]. W. Lu, Q. Wang, and Y. Zhao, "KARMA: Leveraging multi-agent LLMs for automated knowledge graph enrichment," in Proc. NeurIPS, 2025.

[27]. E. Lumer et al., "Tool-to-agent retrieval: Bridging tools and agents for scalable LLM multi-agent systems," arXiv preprint, 2025.

[28]. T. Mao et al., "A multi-agent framework for multi-source manufacturing knowledge integration and question answering," in Proc. WWW Companion, 2025.

[29]. M. R. Martínez Rodríguez et al., "An LLM and embeddings-based multi-agentic system for knowledge graph construction and verification," in Proc. PRIMA, 2025.

[30]. Y. Meng et al., "LLM and knowledge graph-based intelligent control framework for smart manufacturing," in Proc. ASME IDETC, 2025.

[31]. T. Morioka, T. Kono, and T. Nishida, "Automatic construction of asset knowledge graph with large language model," Procedia CIRP, 2025.

[32]. M. Muhtadi et al., "Adaptive self-prompting in agentic LLM frameworks for code fault detection," Software, 2026.

[33]. R. Parekh et al., "Leveraging knowledge graphs and LLM reasoning to identify operational bottlenecks for warehouse planning assistance," arXiv preprint, 2025.

[34]. J. Polonuer et al., "Autonomous knowledge graph exploration with adaptive breadth-depth retrieval," arXiv preprint, 2026.

[35]. N. Pussadeniya et al., "CARA: A hybrid framework integrating swarm AI agents and knowledge graphs for advanced LLM reasoning," in Proc. ICAC, 2024.

[36]. R. Saleh et al., "Self-evolving multi-agent network for industrial IoT predictive maintenance," arXiv preprint, 2026.

[37]. S. Shao et al., "A medical consultation system for geriatric disease based on multi-agent architecture and knowledge graph," in Proc. HIS, 2024.

[38]. W. Tang et al., "LLM-driven framework for knowledge graph construction, completion, and alignment," in Proc. EIECS, 2025.

[39]. G. Tan et al., "LIDL: LLM integration defect localization via knowledge graph-enhanced multi-agent analysis," arXiv preprint, 2026.

[40]. K. T. Tran et al., "Multi-agent collaboration mechanisms: A survey of LLMs," arXiv preprint, 2025.

[41]. J. Wang et al., "Large language models-guided dynamic adaptation for temporal knowledge graph reasoning," in Proc. NeurIPS, 2024.

[42]. Q. Wang et al., "Rethinking the bounds of LLM reasoning: Are multi-agent discussions the key?" in Proc. ACL, 2024.

[43]. X. Wang et al., "Multi-dimensional data analysis and applications based on LLM agents and knowledge graph interactions," arXiv preprint, 2025.

[44]. T. Wu, J. Li, K. Bao, and Q. Liu, "ProcessCarbonAgent: A large language model-empowered autonomous agent for decision-making in manufacturing carbon emission management," Journal of Manufacturing Systems, 2024.

[45]. T. Wu, J. Li, Q. Liu, and J. Bao, "CausalGID: Causality-guided Pareto generative design for industrial production with large language models," Journal of Engineering Design, 2026.

[46]. Y. Xie et al., "KERAP: A knowledge-enhanced reasoning approach for accurate zero-shot diagnosis prediction using multi-agent LLMs," arXiv preprint, 2025.

[47]. F. Yan et al., "An improved LLM and knowledge graph integration method for automated machining process base construction," Journal of Manufacturing Systems, 2026.

[48]. J. Yu et al., "Dyntaskmas: A dynamic task graph-driven framework for asynchronous and parallel LLM-based multi-agent systems," in Proc. ICAPS, 2025.

[49]. Y. Yun et al., "EICopilot: Search and explore enterprise information over large-scale knowledge graphs with LLM-driven agents," in Proc. IEEE Big Data, 2025.

[50]. T. Zang et al., "Event-state knowledge graph-enhanced GraphRAG for order-driven 3D printing supply chain configuration," Expert Systems with Applications, 2026.

[51]. J. Zhang et al., "CoMA-IKG: LLM-driven multi-agent framework for automated construction of industrial knowledge graph," IEEE Transactions on Industrial Informatics, 2026.

[52]. Q. Zhang et al., "A survey of GraphRAG for customized large language models," arXiv preprint, 2025.

[53]. Y. Zhang et al., "Automatic knowledge graph construction and dynamic fusion using LLMs and graph embedding for medical informatics education," in Proc. CSEI, 2024.

[54]. Z. Zhao et al., "A large language model-based multi-agent manufacturing system for intelligent shopfloors," Advanced Engineering Informatics, 2026.

[55]. J. Zhou, L. Wang, and X. Yang, "Guardian: Safeguarding LLM multi-agent collaborations with temporal graph modeling," arXiv preprint, 2025.

[56]. Z. Zhou et al., "Synergizing LLM agents and knowledge graph for socioeconomic prediction in LBSN," arXiv preprint, 2024.