Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 277 (2026) 2055–2064

Procedia Computer Science

www.elsevier.com/locate/procedia

7th International Conference on Industry of the Future and Smart Manufacturing
(former International Conference on Industry 4.0 and Smart Manufacturing)

Evaluating GraphRAG for industrial safety: a case study on LOTO procedure failures

Sara Salvi[a], Nicolò Sabetta[a*], Francesco Costantino[a]

[a] *Department of Computer, Control and Management Engineering, Sapienza University of Rome, Via Ariosto 25, 00185 Rome, Italy*

Abstract

Large Language Models (LLMs) are transforming information access and decision support across domains, yet their application in safety-critical settings remains limited by challenges such as hallucination, lack of domain grounding, and interpretability. To address these issues, Graph Retrieval-Augmented Generation (GraphRAG) has emerged as a novel paradigm that integrates LLMs with knowledge graphs, enabling more coherent, faithful, and context-aware outputs through structured semantic retrieval.
In this context, this exploratory study explores the application of GraphRAG to the domain of industrial safety, focusing on incidents involving Lockout/Tagout (LOTO) procedure failures. By integrating a Neo4j-based knowledge graph constructed from a set of accident narratives, extracted from the OSHA database, with the generative capabilities of GPT-4o, we assess the system's ability to produce coherent, complete, and decision-relevant answers grounded in structured safety data. A total of 150 questions, categorized into six task types, were used to evaluate model performance across six metrics: Coherence, Completeness, Empowerment, Faithfulness, F1 Score, and Relevance. The results highlight GraphRAG's strengths in tasks aligned with graph structure, particularly Summarization, Classification, and Recommendation, while revealing performance limitations in more cognitively demanding tasks such as Reasoning and Comparison. The evaluation underscores the value of structured semantics in enhancing generation quality but also points to scalability and interpretability challenges.

Keywords: Graph-based RAG; LLMs; Smart Manufacturing; Occupational Hazard Prevention; Safety Management

*Nicolò Sabetta.
E-mail address: nicolo.sabetta@uniroma1.it

1. Introduction

Large Language Models (LLMs) and Generative Artificial Intelligence (GenAI) have significantly influenced various sectors by demonstrating capabilities such as coherent text generation, complex query answering, information

summarization, and natural language interaction [1]. These advancements have led to transformative applications in fields like healthcare [2], education [3], industrial engineering [4], and safety management [5], enabling new avenues for automation, decision support, and knowledge discovery.

Despite these achievements, LLMs face notable limitations when applied to specialized domains. A primary concern is their tendency to produce inaccurate or fabricated information, a phenomenon known as "hallucination". This issue often stems from the absence of domain-specific, up-to-date, or proprietary knowledge in their training data [6]. To address this, techniques such as Retrieval-Augmented Generation (RAG) have been developed [7], which enhance LLM outputs by integrating relevant external documents during the generation process. However, even if RAG enhance the capabilities in domain specific applications, issues remain. For instance, RAG systems often retrieve information which can introduce noise and affect the generation quality [8]. Moreover, traditional RAG architectures treat retrieved documents as independent, lacking mechanisms to model inter-document relationships, which can lead to fragmented or inconsistent outputs [9]. RAG remains susceptible to hallucinations as well, particularly when the retrieved content is sparse, noisy, or only weakly relevant to the query [6].

In this context, an evolution of this approach known as Graph Retrieval-Augmented Generation (GraphRAG) emerged. GraphRAG incorporates Knowledge Graphs (KGs) to further improve contextual grounding and factual consistency. Thanks to these techniques, LLMs can be applied to specialized domain, like the industrial one. In this sector, the integration of LLMs has shown promise in supporting operators [4], improving knowledge retrieval [5], and enhancing decision-making processes . A particularly compelling application is in occupational health and safety, where LLMs can contribute to incident prevention, risk management, and the reinforcement of safety practices. Moreover, the use of knowledge graphs in industrial safety has been vastly explored by researchers. KGs have emerged as a powerful tool in this context, offering structured and interconnected representations of safety data. By capturing relationships among equipment, procedures, hazards, and incidents, KGs facilitate a more nuanced understanding of complex safety information [10]. When combined with LLMs through GraphRAG, these graphs can contextualize model outputs, reduce hallucinations, and improve the reliability of safety recommendations.

In this scenario, this study evaluates the GraphRAG in the domain of industrial safety according to different evaluation metrics. More in depth, we aim to assess the system's ability to generate answers that are logically coherent, complete in addressing all relevant aspects of the question, clearly supportive of user understanding, grounded in retrieved information, and closely aligned with the intent and focus of the query. The focus is specifically on maintenance-related incidents involving Lockout/Tagout (LOTO) procedures, which are safety protocols established by the Occupational Safety and Health Administration (OSHA) in 1989 to minimize the risk of accidents during maintenance operations. The study will follow a methodological approach where, starting from a KG built from a database of 20 LOTO accidents narratives, a set of 150 questions were evaluated according to different metrics.

Finally, with this study the authors aim to answer to the following research question:

How effectively can a GraphRAG architecture support semantically grounded and decision-relevant question answering in the domain of industrial safety analysis?

The remainder of this paper is structured as follows. Section 2 provides a detailed background on GraphRAG, outlining its conceptual foundations and prior applications across diverse domains. Section 3 describes the methodology adopted in this study, including the construction of the LOTO accident knowledge graph, the integration of GraphRAG with GPT-4o, and the evaluation framework. Section 4 presents and discusses the results, with particular emphasis on the strengths and limitations of GraphRAG in safety-critical contexts. Finally, Section 5 concludes the paper by summarizing key findings, identifying limitations, and suggesting directions for future research.

2. Background

The GraphRAG is a novel extension of the RAG paradigm that incorporates graph-based representations into the retrieval-augmented pipeline [11]. Rather than treating the retrieval step as a flat list of documents, GraphRAG constructs a knowledge graph from the retrieved content, where nodes represent passages or concepts and edges encode semantic or relational connections. This graph-structured representation enables the LLM to reason over structured knowledge, fostering more coherent and context-aware outputs.

By enriching retrieval with graph-based contextualization, GraphRAG significantly enhances multi-hop reasoning, long-context understanding, and factual consistency. These capabilities render it particularly effective in complex information domains such as scientific question answering, technical support, and enterprise search. The use of knowledge graphs in this setting is especially pertinent, as they enable the representation of hierarchical and relational information, such as equipment structures, failure modes, and safety protocols, in a formal, machine-readable format.

Compared to standard RAG, GraphRAG demonstrates superior performance across several qualitative and quantitative dimensions. By leveraging structured knowledge graphs, GraphRAG enhances coherence and completeness through semantically linked context expansion, enabling more logically consistent and comprehensive answers. It also improves faithfulness by grounding responses in explicitly connected data, thereby reducing hallucinations. Furthermore, the graph structure facilitates better user understanding and interpretability, supporting empowerment.

Building on this foundation, GraphRAG has gained increasing attention for its ability to retrieve and reason over graph-structured data, including both pre-existing knowledge graphs and those dynamically constructed from unstructured text. Recent studies have shown its effectiveness in text-based applications where implicit knowledge is first transformed into graph representations, benefitting tasks such as global summarization, planning, and logical reasoning In the domain of contract analysis, researchers proposed a tuning-free GraphRAG framework that integrates LLMs with a Nested Contract Knowledge Graph to identify contractual risks more accurately and interpretably than baseline models. This method showcased notable improvements in reasoning reliability and output transparency in complex legal documents [12]. A similar hybridization of LLMs and graph-based retrieval was applied in materials science, where Microsoft's GraphRAG facilitated the extraction of structured chemical knowledge from unstructured literature. This allowed for high-fidelity prediction of catalyst performance and the discovery of unconventional catalytic behaviors, thereby exemplifying the potential of LLM-driven pipelines in experimental sciences [13].

GraphRAG has also been utilized in policy generation and logistics, where a model trained on user-generated YouTube content outperformed a standard GPT model in generating contextually appropriate policy recommendations. This illustrates how graph-enhanced context can improve generation quality in noisy, real-world datasets [14]. In the field of security export control, a knowledge graph was employed to expand prompts and enrich LLM context windows. While this improved classification of answer types, it simultaneously revealed trade-offs in regulatory category prediction, suggesting that task-specific graph construction remains a crucial factor in performance optimization [15].

To address security and data privacy concerns in sensitive environments, a hybrid QA system has been proposed that combines local open-source LLMs with GraphRAG-based retrieval. This approach enabled secure, cost-effective domain adaptation for defect management, achieving state-of-the-art results while maintaining data confidentiality [16]. The utility of GraphRAG in financial text processing was further extended through a HybridRAG framework, which combines both graph-based and vector-based retrieval. Experiments with earnings call transcripts demonstrated that HybridRAG outperforms either method alone in both retrieval and answer generation, reinforcing the complementary strengths of symbolic and distributional representations [17].

Lastly, in industrial mining, GraphRAG was integrated into a novel AI-powered cloud platform for fault diagnosis and predictive maintenance. Here, knowledge graph structures facilitated precise fault detection in mining equipment, enabling intelligent Q&A and decision-making within a highly automated system [18].

In the construction industry, knowledge graphs have been successfully employed to encode safety regulations and standards, enabling automated detection of violations and proactive identification of hazards. Similarly, in process industries, KGs have been used to model knowledge from HAZOP studies and near-miss reports, thereby improving hazard detection and safety protocols. By modeling and analyzing procedural workflows, incident reports, and hazard identifications, KGs support systematic extraction of safety-critical information, facilitating proactive risk management. Furthermore, their application in industrial control systems enhances situational awareness and supports data-driven decision-making, contributing to more effective and adaptive safety strategies.

Despite this potential, the integration of GenAI, specifically GraphRAG, into industrial safety remains an emerging yet promising frontier. Current research has largely focused on mining unstructured data sources, such as incident reports [19], maintenance logs [20], and inspection records [21], to identify patterns and predict equipment failures. While these efforts have yielded valuable insights, they often rely on static, unstructured representations and fall short of leveraging structured semantic frameworks. Only a limited number of studies have explored the synergistic

integration of LLMs with structured representations such as KGs in safety-critical settings. Thus, the systematic application of KG-augmented LLMs in generation tasks remains relatively underexplored.

A critical element of industrial safety is the implementation of LOTO procedures, which are designed to ensure that equipment is fully de-energized and cannot be inadvertently reactivated prior to the completion of maintenance or servicing activities. These protocols are essential for mitigating the risk of accidental energization, which can result in severe injuries or fatalities. LOTO-related incidents continue to pose a significant concern across industries. According to the OSHA, proper adherence to LOTO procedures has the potential to prevent approximately 50,000 injuries and 120 deaths annually. However, in 2024, OSHA recorded 2,443 violations of the LOTO standard, placing it among the ten most frequently cited workplace safety infractions [22].

The formalization of LOTO protocols within KGs enables systematic representation, automated compliance verification, and the development of simulation-based training tools. Furthermore, the integration of LLM with LOTO-enriched KGs through GraphRAG frameworks offers the potential to provide real-time, context-sensitive guidance to maintenance personnel, thereby enhancing protocol adherence and reducing the likelihood of human error. Despite existing research focused on the implementation, operational mechanisms, and documentation of LOTO procedures, the dynamic interplay between safety assurance and maintenance efficiency remains underexplored. As a result, LOTO continues to represent a partially developed process that lacks seamless integration into holistic production and maintenance management strategies.

3. Methodology

This study adopts a structured systematic approach methodology. Starting from a knowledge framework built from 20 LOTO-related accident narratives extracted from the OSHA database, a GraphRAG system was built, enabling advanced information retrieval and reasoning capabilities. As previous studies have already demonstrated the superiority of GraphRAG over standard RAG across various evaluation settings (e.g., [23], [24]), this study focuses exclusively on assessing the performance of GraphRAG in the industrial safety domain. The goal is not to replicate existing benchmarks, but to assess GraphRAG's suitability for high-risk, regulation-driven contexts like LOTO.

At the core of the KG, each Accident node functions as a central anchor point, uniquely identified by an accident_id and annotated with two temporal attributes: date and time. These central nodes are surrounded by a structured set of domain-specific entities that capture the multi-dimensional context of each event. The graph's ontology, detailing entity types and their interconnections, is defined according to the schema outlined in Table 1 and Table 2.

Table 1. Entities of the LOTO accidents' Knowledge Graph

Entities	Definition
Accident	Id, date, time of the accident
Personnel	Role of the person involved
Consequence	Injury or negative outcome
Severity	Level of harm
Location	Functional or departmental area
Type_of_procedures	Safety procedures violated or omitted
Energy	Hazardous sources
Activity	Task in progress during the incident
Place	Physical site of the event
Machine	Equipment or systems involved
Plant	Industrial sector or facility type

Table 2. Relations of the LOTO accidents' Knowledge Graph

Initial node	Relation	Final node
Accident	INVOLVES	Personnel
Accident	DISREGARDS	Type_of_procedures
Accident	TAKES_PLACE_IN	Location
Accident	WHILE	Activity
Personnel	CARRYING_OUT	Activity
Personnel	HAS_SOFFERED	Severity
Personnel	EXPOSED_TO	Consequence
Activity	WHERE	Place
Activity	ON	Machine
Machine	RELEASES	Energy
Consequence	RESULTS_IN	Severity
Location	IN	Plant

To operationalize the KG for question answering, we developed a semantic querying framework that integrates a manually curated Neo4j graph with the generative reasoning capabilities of OpenAI's GPT-4o language model. The implementation utilizes the langchain_neo4j library and the GraphCypherQAChain module from LangChain to enable structured, schema-aware question answering.

When a user submits a query in natural language, the system translates it into a Cypher query using a custom prompt that guides the model to adhere strictly to the graph's predefined schema. This prompt constrains the model to operate within the available relationships and properties of the KG, ensuring valid and interpretable queries. The resulting Cypher query is executed against the Neo4j database, and the retrieved triples are fed back into the LLM. The final response is generated by the model, combining fluent language generation with factual grounding in structured data.

To evaluate the model's robustness and sensitivity to prompt formulation, each question was tested in three variants: (1) the original formulation, (2) a synonym-based rephrasing preserving the semantic intent but altering lexical choices, and (3) a prompt-engineered version containing examples or explicit instructions to guide the model's reasoning. This yielded a total of 150 questions, enabling the analysis of response consistency and the model's ability to correctly identify and reason over different input formulations.

The questions were categorised into six high-level task types, which are Information Retrieval, Summarization, Classification, Reasoning, Comparison, and Recommendation. Examples of questions for each type of task are provided in Table 3.

Table 3. Examples of questions for each task

Task	Number of Questions	Example Questions
Information Retrieval	41	Which type of plant has the highest number of accidents?
Summarization	23	Can you provide a summary of the incident with ID 166492.015?
Classification	17	What are the three most common types of incidents?
Reasoning	29	Is there a connection between the time of the incident and its fatality?
Comparison	23	Find accidents with similar causes that occurred in different plants.
Recommendation	17	If a company wanted to reduce fatal incidents, which factors should it focus on and why?

Each response was evaluated across six complementary metrics:

1. Coherence: logical flow and internal consistency of the answer.

2. Completeness: coverage of all relevant aspects of the question.
3. Empowerment: support for user understanding and informed decision-making.
4. Faithfulness: degree to which the answer is grounded in retrieved data.
5. F1 Score: harmonic mean of precision and recall.
6. Relevance: alignment with the question's intent and topical focus.

For the evaluation process, specific success criteria were defined for each metric to ensure that human assessment of answer quality remained both objective and consistent. According to these criteria, every metric linked to a given question was scored on a three-point scale: a score of 0 indicated that the response did not satisfy the metric, 0.5 signalled partial satisfaction, and 1 represented full satisfaction based on how well the response met the predefined standards. Table 4 shows an example of this evaluation process. Following this scoring, the average value for each metric was calculated to capture central tendency, while the standard deviation (σ) was computed to measure variability in performance across responses.

Table 4. Evaluation process example

User Question	GraphRAG answer	Metric	Score	Motivation
How many fatal accidents have been caused and by what causes?	'Fatal accidents have been caused by the following: procedures disregarded or not implemented (5 accidents), notification (2 accidents), energy discharge (1 accident), and labeling (1 accident).	Empowerment	1	Identifies all the causes and specifies the exact number of associated incidents
What is the total number of fatal accidents, and what caused them?	The total number of fatal accidents is 9, caused by procedures disregarded or not implemented, notification, energy discharge, and labeling.	Empowerment	0.5	Identifies all the causes but does not specify the exact number of associated incidents

A task was deemed successful when the average score for each metric reached or exceeded 70% positive evaluations, and the standard deviation did not surpass 0.15 across semantically related metrics. This dual requirement ensured that high performance was accompanied by consistency. In situations where variability exceeded the threshold, the potential interdependence between metrics was taken into account as a compensatory factor in interpreting the results.

4. Results and discussion

The evaluation highlights GraphRAG's strong performance in tasks requiring structured understanding and factual consistency (Fig. 1, Fig. 2).

	Coherence	Completeness	Empowerment	Faithfulness	F1 Score	Relevance	Mean ± σ (per task)
Information Retrieval	1.000 ± 0.000	0.857 ± 0.350	0.524 ± 0.308	0.679 ± 0.375	0.607 ± 0.354	0.786 ± 0.364	0.742 ± 0.159
Summarization	1.000 ± 0.000	1.000 ± 0.000	0.875 ± 0.217	0.979 ± 0.100	0.958 ± 0.200	0.979 ± 0.100	0.965 ± 0.043
Reasoning	1.000 ± 0.000	0.667 ± 0.471	0.550 ± 0.435	0.583 ± 0.430	0.617 ± 0.460	0.617 ± 0.460	0.672 ± 0.151
Classification	1.000 ± 0.000	0.833 ± 0.236	0.667 ± 0.236	1.000 ± 0.000	0.833 ± 0.236	0.833 ± 0.236	0.861 ± 0.115
Comparison	1.000 ± 0.000	0.750 ± 0.433	0.500 ± 0.354	0.625 ± 0.415	0.500 ± 0.354	0.750 ± 0.433	0.688 ± 0.173
Recommendation	1.000 ± 0.000	0.944 ± 0.229	0.778 ± 0.342	0.889 ± 0.314	0.778 ± 0.342	0.889 ± 0.314	0.880 ± 0.081
Mean ± σ (per metric)	1.000 ± 0.000	0.833 ± 0.360	0.625 ± 0.356	0.760 ± 0.373	0.692 ± 0.377	0.795 ± 0.371	

mean ≥ 0.7 AND σ ≤ 0.15
mean ≥ 0.7 XOR σ ≤ 0.15
mean ≤0.7 AND σ ≥ 0.15

Fig. 1. Performance metrics heatmap per task

Coherence achieved perfect scores (1.000 ± 0.000) across all tasks, confirming the model's ability to generate syntactically and logically well-formed responses. This consistency likely stems from the synergy between the graph's structure and the language model's generative capacity.

Metrics like Completeness, Relevance, and Faithfulness also exceeded the success threshold on average, though with higher variance. These fluctuations were especially pronounced in tasks such as Reasoning and Comparison, where the model often struggled to fully integrate information across multiple nodes, leading to incomplete or partially grounded answers. In contrast, simpler tasks like Summarization and Classification showed near-perfect scores across these dimensions.

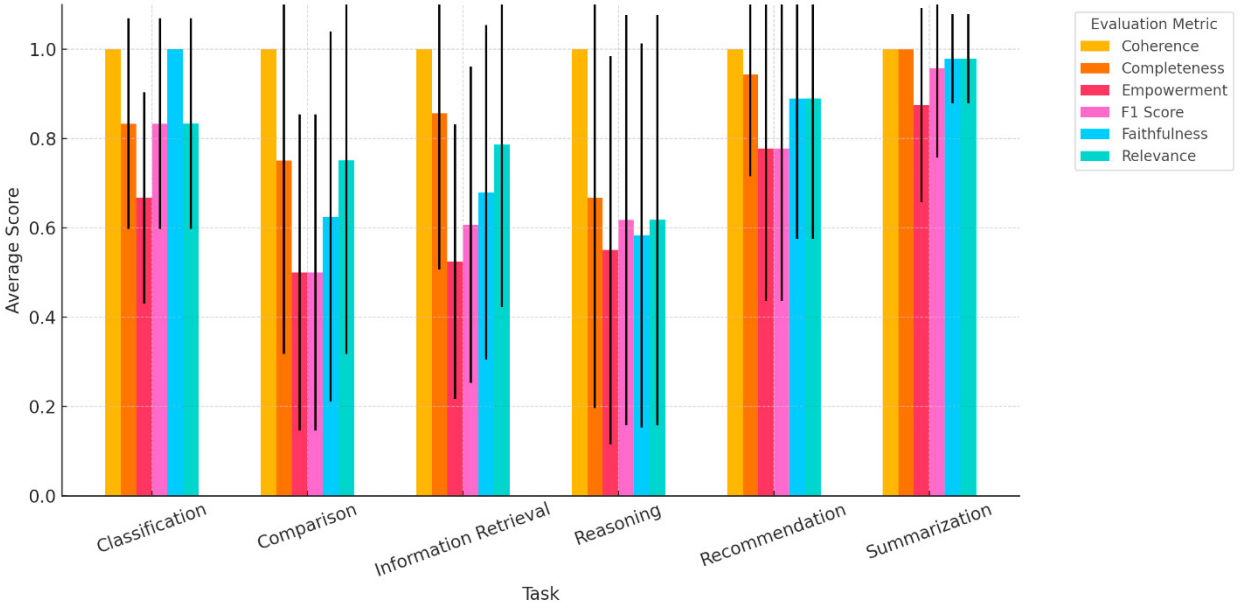

Fig. 2. GraphRAG Performance Across Tasks And Metrics

Empowerment, however, remains a weakness. With a mean of 0.625 and high variability, the system often failed to produce actionable or user-enabling responses, especially in tasks requiring decision support. Even when the KG contained relevant data, the model frequently defaulted to vague answers, exposing limitations in navigating semantically distant information.

The robustness of the system was further tested through variations in question formulation: original, synonym-based, and prompt-engineered. While lexical rephrasings had minimal impact, prompt-engineered queries introduced notable performance shifts. In some cases, these helped recover relevant information; in others, they misdirected the model, especially in complex reasoning tasks. This underscores how strategic prompting can be both a powerful asset and a potential risk in real-world applications.

Analyzing performance by task type, Summarization stands out with the highest overall results (mean = 0.965, σ = 0.043), followed by Recommendation and Classification, which benefited from clear graph structures and simpler logic. Conversely, Reasoning and Comparison remain challenging due to the system's limited ability to aggregate and interpret information across larger graph segments. Some improvements were achieved by restructuring the graph, but further developments, such as community-based traversal, may be needed to support more advanced inference.

These findings also open relevant perspectives for industrial safety applications, particularly in the context of accident prevention. The system's ability to generate coherent, accurate, and contextually grounded answers can support the identification of recurring risks and procedural gaps, especially those related to the absence of LOTO protocols. Integrated within safety assessment tools or training platforms, GraphRAG could assist operators and risk managers in anticipating critical scenarios, proposing preventive measures, and enhancing overall awareness. Its potential extends beyond passive information retrieval. In fact, by enabling dynamic exploration of incident patterns

and causal relations, the system could foster a more proactive and data-driven safety culture. Moreover, its traceability features make it suitable for compliance auditing and post-incident analysis, offering transparent justifications for generated outputs. In safety-critical environments, where interpretability and trust are essential, GraphRAG emerges not only as a knowledge retrieval tool but as a strategic component for risk-informed decision making and organizational learning.

5. Conclusion

This study explored the application of the GraphRAG architecture in the high-stakes domain of industrial accidents involving missing LOTO procedures, by integrating structured knowledge from a Neo4j-based knowledge graph with the generative capabilities of GPT-4o. The system was evaluated on a manually curated dataset, using a formal ontology and 150 questions across six task types, assessed through six performance metrics.

The findings indicate that GraphRAG performs well in tasks that align closely with the structure of the knowledge graph, particularly in Summarization, Classification, and Recommendation. In these areas, the integration of semantic structure and language generation yielded responses that were coherent, relevant, and faithful, reinforcing the promise of hybrid architectures in schema-driven domains. However, notable limitations emerged in more complex cognitive tasks such as Reasoning and Comparison. The system demonstrated reduced faithfulness and completeness when required to synthesize information across semantically distant or weakly connected nodes. Empowerment, which measures the user's ability to act effectively on system outputs, also remained limited, suggesting that factual correctness alone is insufficient to ensure usable insights. These challenges reflect a lack of global semantic integration and difficulties in navigating complex graph relationships. Additionally, the manual construction of the knowledge graph limits scalability and coverage, while prompt engineering alone was inadequate to overcome deeper inference constraints.

Importantly, this paper presents an exploratory investigation that lays the groundwork for future research. Subsequent work will include a systematic comparison with alternative RAG techniques, including traditional vector-based RAG, to better assess the relative advantages and trade-offs of graph-based approaches. Future directions will also examine advanced variants such as Community-based GraphRAG [25] and PathRAG [26] to enhance semantic propagation, alongside improvements in task-aware prompting, dynamic subgraph selection, and tighter coupling between graph reasoning and language generation.

In conclusion, while GraphRAG demonstrates strong potential for producing structured and accurate outputs in domain-specific contexts, realizing its full value for expert reasoning and decision-making, particularly in industrial safety, will require broader contextual understanding and comparative validation against established retrieval architectures.

Acknowledgements

This research was financed by INAIL (Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro), BRIC2022-ID63.

References

[1] H. Benbya, Deakin University, F. Strich, Deakin University, T. Tamm, and Deakin University, 'Navigating Generative Artificial Intelligence Promises and Perils for Knowledge and Creative Work', *JAIS*, vol. 25, no. 1, pp. 23–36, 2024, doi: 10.17705/1jais.00861.

[2] J. Vrdoljak, Z. Boban, M. Vilović, M. Kumrić, and J. Božić, 'A Review of Large Language Models in Medical Education, Clinical Decision Support, and Healthcare Administration', *Healthcare*, vol. 13, no. 6, Art. no. 6, Jan. 2025, doi: 10.3390/healthcare13060603.

[3] J. Swacha and M. Gracel, 'Retrieval-Augmented Generation (RAG) Chatbots for Education: A Survey of Applications', *Applied Sciences*, vol. 15, no. 8, Art. no. 8, Jan. 2025, doi: 10.3390/app15084234.

[4] S. Colabianchi, F. Costantino, and N. Sabetta, 'Assessment of a large language model based digital intelligent assistant in assembly manufacturing', *Computers in Industry*, vol. 162, 2024, doi: 10.1016/j.compind.2024.104129.
[5] N. Sabetta, F. Costantino, and S. Stabile, 'A comparative analysis for automated information extraction from OSHA Lockout/Tagout accident narratives with Large Language Model', *Procedia Computer Science*, vol. 253, pp. 1362–1372, Jan. 2025, doi: 10.1016/j.procs.2025.01.198.
[6] T. Hwang, S. Jeong, S. Cho, S. Han, and J. C. Park, 'DSLR: Document Refinement with Sentence-Level Re-ranking and Reconstruction to Enhance Retrieval-Augmented Generation', presented at the KnowledgeNLP 2024 - 3rd Workshop on Knowledge Augmented Methods for NLP, Proceedings of the Workshop, 2024, pp. 73–92.
[7] W. Fan *et al.*, 'A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models', presented at the Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2024, pp. 6491–6501. doi: 10.1145/3637528.3671470.
[8] M. Li, H. Kilicoglu, H. Xu, and R. Zhang, 'BiomedRAG: A retrieval augmented large language model for biomedicine', *Journal of Biomedical Informatics*, vol. 162, 2025, doi: 10.1016/j.jbi.2024.104769.
[9] M. H. Heydari, A. Hemmat, E. Naman, and A. Fatemi, 'Context Awareness Gate For Retrieval Augmented Generation', presented at the 15th International Conference on Information and Knowledge Technology, IKT 2024, 2024, pp. 260–264. doi: 10.1109/IKT65497.2024.10892659.
[10] F. Simone, S. M. Ansaldi, P. Agnello, and R. Patriarca, 'Industrial safety management in the digital era: Constructing a knowledge graph from near misses', *Computers in Industry*, vol. 146, p. 103849, Apr. 2023, doi: 10.1016/j.compind.2022.103849.
[11] V. Kamra, L. Gupta, D. Arora, and A. K. Yadav, 'Enhancing Document Retrieval Using AI and Graph-Based RAG Techniques', presented at the Proceedings - IEEE 5th International Conference on Communication, Computing and Industry 6.0 2024, C2I6 2024, 2024. doi: 10.1109/C2I663243.2024.10895931.
[12] C. Zheng, S. Wong, X. Su, Y. Tang, A. Nawaz, and M. Kassem, 'Automating construction contract review using knowledge graph-enhanced large language models', *Automation in Construction*, vol. 175, 2025, doi: 10.1016/j.autcon.2025.106179.
[13] C. Wei *et al.*, 'Large Language Models Assisted Materials Development: Case of Predictive Analytics for Oxygen Evolution Reaction Catalysts of (Oxy)hydroxides', *ACS Sustainable Chemistry and Engineering*, vol. 13, no. 14, pp. 5368–5380, 2025, doi: 10.1021/acssuschemeng.5c00798.
[14] H. Naganawa and E. Hirata, 'Enhancing Policy Generation with GraphRAG and YouTube Data: A Logistics Case Study', *Electronics (Switzerland)*, vol. 14, no. 7, 2025, doi: 10.3390/electronics14071241.
[15] R. Rzepka and A. Obayashi, 'Effectiveness of Security Export Control Ontology for Predicting Answer Type and Regulation Categories', presented at the ICAAI 2024 - Conference Proceedings of the 2024 8th International Conference on Advances in Artificial Intelligence, 2025, pp. 156–161. doi: 10.1145/3704137.3704180.
[16] K. Jeon and G. Lee, 'Hybrid large language model approach for prompt and sensitive defect management: A comparative analysis of hybrid, non-hybrid, and GraphRAG approaches', *Advanced Engineering Informatics*, vol. 64, 2025, doi: 10.1016/j.aei.2024.103076.
[17] B. Sarmah, D. Mehta, B. Hall, R. Rao, S. Patel, and S. Pasquali, 'HybridRAG: Integrating Knowledge Graphs and Vector Retrieval Augmented Generation for Efficient Information Extraction', presented at the ICAIF 2024 - 5th ACM International Conference on AI in Finance, 2024, pp. 608–616. doi: 10.1145/3677052.3698671.
[18] Y. Wang, Y. Feng, C. Xi, B. Wang, B. Tang, and Y. Geng, 'Development of an Intelligent Coal Production and Operation Platform Based on a Real-Time Data Warehouse and AI Model', *Energies*, vol. 17, no. 20, 2024, doi: 10.3390/en17205205.
[19] A. J. Nakhal A, R. Patriarca, G. Di Gravio, G. Antonioni, and N. Paltrinieri, 'Investigating occupational and operational industrial safety data through Business Intelligence and Machine Learning', *Journal of Loss Prevention in the Process Industries*, vol. 73, p. 104608, Nov. 2021, doi: 10.1016/j.jlp.2021.104608.
[20] R. K. Gangadhari, M. Rabiee, V. Khanzode, S. Murthy, and P. Kumar Tarei, 'From unstructured accident reports to a hybrid decision support system for occupational risk management: The consensus converging approach', *Journal of Safety Research*, vol. 89, pp. 91–104, Jun. 2024, doi: 10.1016/j.jsr.2024.02.006.
[21] H. Samarasinghe and S. Heenatigala, 'Insights from the Field: A Comprehensive Analysis of Industrial

Accidents in Plants and Strategies for Enhanced Workplace Safety', Feb. 02, 2024, *arXiv*: arXiv:2403.05539. doi: 10.48550/arXiv.2403.05539.
[22] OSHA, 'Commonly Used Statistics | Occupational Safety and Health Administration'. [Online]. Available: https://www.osha.gov/data/commonstats
[23] Y. Zhou *et al.*, 'In-depth Analysis of Graph-based RAG in a Unified Framework', Mar. 06, 2025, *arXiv*: arXiv:2503.04338. doi: 10.48550/arXiv.2503.04338.
[24] H. Han *et al.*, 'RAG vs. GraphRAG: A Systematic Evaluation and Key Insights', Feb. 17, 2025, *arXiv*: arXiv:2502.11371. doi: 10.48550/arXiv.2502.11371.
[25] D. Edge *et al.*, 'From Local to Global: A Graph RAG Approach to Query-Focused Summarization', Feb. 19, 2025, *arXiv*: arXiv:2404.16130. doi: 10.48550/arXiv.2404.16130.
[26] B. Chen *et al.*, 'PathRAG: Pruning Graph-based Retrieval Augmented Generation with Relational Paths', Feb. 18, 2025, *arXiv*: arXiv:2502.14902. doi: 10.48550/arXiv.2502.14902.