Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P9506-cd

Knowledge Graph Construction of Large Language Model Retrieval Augmented Generation for Oil and Gas HSE Professional Q&A System

Yiyue Chen

Research Institute of Safety and Environmental Technology, CNPC, China. E-mail: chenyiyue@cnpc.com.cn

Jiyu Zhai

Research Institute of Safety and Environmental Technology, CNPC, China. E-mail: zhaijiyu@cnpc.com.cn

Shuncheng Wu

Research Institute of Safety and Environmental Technology, CNPC, China. E-mail: wushcheng@cnpc.com.cn

Kun Tian

Research Institute of Safety and Environmental Technology, CNPC, China. E-mail: tiankuntk@cnpc.com.cn

Xu Song

Research Institute of Safety and Environmental Technology, CNPC, China. E-mail: songx_dod@petrochina.com.cn

Abstracts: Among Oil and gas HSE (Health, Safety, and Environment) management, supervision and inspection are essential to stable production. During inspections, dealing with complex professional knowledge are challenged by heavy manual review and low efficiency. Currently, while Large Language Models unfold powerful text comprehension and generation capabilities, their performance is limited in the HSE field due to lack of specialized knowledge and data resources. There exists a contradiction between the professional answers demand and the non-targeted, lacking reference basis responses. Therefore, knowledge graph technology is introduced to enrich knowledge resources and enhance question-answering effectiveness. This study focused on constructing a knowledge graph in the oil and gas HSE field, to enhance the accuracy, recall and pertinence of LLM question-answering. Firstly, collect HSE related policy files, design ontology model concepts, including production process, applicable object, business domain, theory and policy, data sources, etc. Secondly, design ontology model relations, including part-of, kind-of, instance-of, attribute-of, meet-above, etc. Thirdly, identify entities of policy file and extract relationships to construct domain knowledge graph. Finally, store the vectorized content of sliced policy files in the vector knowledge base, to provide rich and accurate knowledge external sources for the large model. Question-answering tests show that the constructed knowledge graph significantly enhances the performance of Large Language Model answering. In small scene query, it can directly answer the standard specification requirements; In standard query, the most applicable files are matched according to multi-dimensional information; In reasoning question, it supports the analysis of the thought chain, and provides the legal basis step by step. Overall, the construction of the knowledge graph effectively tackles the challenge of evidence-based question-answering by large models within the professional field, offering valuable knowledge support to technicians and managers during production activities and HSE inspections.

Keywords: Knowledge graph, Domain ontology model, Large Language Model(LLM), Retrieval Augmented Generation(RAG), Oil and gas production, HSE management system.

1. Introduction

On-site supervision and inspection is one of the important means of safety management in oil and gas production site. Through timely detection of hidden risks, put forward targeted rectification measures, cut off the accident evolution chain (Fabiano et al., 2022). Currently, the supervision and inspection work are mainly carried out manually, and in the inspection work, the professional safety knowledge of personnel is highly dependent. When faced with uncertainty, or knowledge outside the memory reserve, tools

are needed to supplement HSE expertise on whether the current scenario is in violation and which regulations are violated.

In the traditional HSE field knowledge question and answer, some inspectors voluntarily sorted out the commonly used standard specification file base or sorted out the common risk determination rule base (Huang et al., 2022). Knowledge query can be realized by keyword retrieval. Due to the huge specifications of the knowledge system, organizing manual sorting and updating is a time-consuming and labor-intensive work, and the frequency of knowledge updating is limited, with an update once every 1-2 years considered good. But even if updated every year, it is difficult to guarantee full coverage of knowledge. At the same time, if the keywords entered are not accurate, there may be an issue that the desired results are not returned (P and H N, 2023). Knowledge retrieval, which relies on semantic retrieval technology, provides more accurate retrieval results by understanding the semantics of user queries and matching relevant knowledge. It can analyze and understand the document content and user questions at the semantic level, so as to extract the concept information and category information to match the query needs of users (Zhang et al., 2013). The knowledge queried through the network database covers a wide range, but the depth of the field needs to be further analyzed and processed by manual. Knowledge graph extracts the knowledge in the text, represents entities or concepts with nodes, and relations represent the connections between entities. It builds a relational network, thus providing users with a relational perspective to analyze problems. In addition, knowledge management systems usually adopt structured storage methods, such as hierarchical structure, classification system, labeling and keyword annotation, which are helpful for knowledge organization and retrieval (Siddharth and Luo, 2024). Rule-based reasoning is also used in some advanced knowledge management systems. The efficiency of inference machine can be improved and the consistency of knowledge base can be solved through knowledge relation construction, granularity decomposition and implicative inference (Du et al., 2022).

At present, large model technology is developing rapidly. Through large-scale text data training, large language model has learned rich language knowledge and patterns, has the ability of emergence, and shows a strong ability of understanding and text generation. It performs well in question-and-answer tasks, can quickly handle complex and diverse user queries, and has absolute advantages in terms of speed in question understanding and answer production direction (OpenAI, 2023, Tom B. Brown, 2020). In the energy and chemical industry, China National Petroleum Corporation (CNPC) has used more than half a century of industry data accumulation to build more than 280TB of high-quality data sets and launched the Kunlun Large Model. It is capable of processing industry-specific complex data and business logic to provide specialized solutions. However, due to the lack of integration of data resources and professional knowledge of HSE management system, the performance of large language model in question answering tasks in HSE field is poor, and the problem of illusion still exists (Huang et al., 2024, Hongbin Ye, 2023). In order to solve the large model nonsense, the mainstream solution is to use the combination of large model and knowledge graph. By synthesizing the understanding generation of large models and the advantages of strong correlation in knowledge graph, the quality assurance of answers is realized (Shirui Pan, 2023, Patrick Lewis, 2021). In the medical field, knowledge retrieval (KR) and reasoning answer (RA) are proposed, which use entity linking algorithms and subgraph construction and fusion strategies to retrieve question-related knowledge, and alleviate the hallucination problem (Zeng et al., 2025). In the law field , legal knowledge graphs are proposed to provide legal citations of interest and provide practical evidence for large model answers (Hannah et al., 2025).

Knowledge graphs are widely applied in enhance knowledge richness and accuracy, as well as facilitate knowledge reasoning and generation. It can provide a logical basis for the reasoning and analysis of the selection of relevant theory files. At present, some relevant knowledge graphs foundation of oil and gas field has been established, including oil and gas emergency response, oil and gas station, chemical equipment, etc. (Chen et al., 2022a, Chen et al., 2022b). The establishment of knowledge organization construction method in oil and gas HSE field as an external link can make up for the high accuracy, high recall rate and high pertinence requirements

of large models in oil and gas HSE field.

Therefore, the main challenge presented is to create clear and convenient knowledge in a form of storage that facilitates data retrieval and provides knowledge support. The key is the design of ontology model of knowledge graph and the filling of graph database based on entity relation extraction. This paper mainly focuses on the knowledge structure required for oil and gas production safety management.

2. Preliminary

The theory of knowledge augmented large model technology mainly compensates for the illusion of large model through strongly related knowledge relationship. General technical route mainly includes knowledge ontology model design, knowledge graph filling and retrieve augmented generation 3 steps, as shown in Fig. 1. The scientific design of the knowledge graph data structure and the high-quality triplet filling directly affect the reference knowledge of the large model answer.

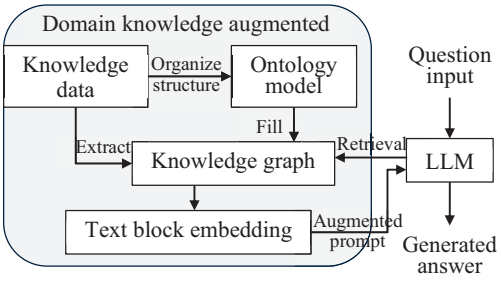

Fig. 1. General technical route of domain knowledge augmented generation of large models

2.1. *Ontology model design*

The ontology model is formalized to make it accurately reflect the internal logic and correlation of knowledge.

- Extensively collect data sources of domain knowledge, sort out data structures, and classify and organize knowledge according to fields or topics based on expert experience
- Clear concepts, attributes, relations and other elements, define the hierarchical structure and semantic relations of the ontology, and use the ontology language (such as RDF, OWL, etc.) to formalize the ontology model

2.2. *Knowledge graph fill*

The entities and relationships, extracted based on ontology model, are represented as triples and imported into graph database to construct knowledge graph.

- Entity/relationship extraction. Utilizing natural language processing technology, such as named entity and relationship extraction, to extract entities and relationships from text for preprocessing. These extracted elements are then organized and classified according to predefined ontology models.
- Knowledge graph storage. A graph database, such as Neo4j, is selected to store the knowledge graph. The extracted information, including entities, relationships, and attributes, is imported into the knowledge graph for structured representation.
- Text block vectorization. Text blocks within the knowledge graph, such as entity descriptions and relationship descriptions, are represented in a vectorized form. Word embedding techniques, like BERT, are employed to map these text blocks into high-dimensional vector spaces, preserving their semantic information. The vectorized text blocks are then indexed and stored, facilitating subsequent similarity retrieval and knowledge search.

2.3. *Retrieval augmented generation*

The large language model comprehends the user's question input, extracts key information, and retrieves relevant data from knowledge graph. It obtains corresponding text block vector, which is utilized to improve the quality of the answer.

- Similarity search. Based on the intent and content of the user's query, similar entities, relationships, or text blocks are retrieved from the knowledge graph. A vector similarity algorithm is employed to rank and filter the search results, returning those best match the query requirements.
- Build query interface. The search results are provided as prompts to the large language model, enhancing its capability to tackle knowledge-intensive tasks.

3. Methods

As the foundational framework for the knowledge graph, the ontology model of oil and gas HSE domain requires a scientific and concise design, free from redundancy and omissions.

3.1. *Oil and gas HSE field ontology concepts*

The concepts involved in the HSE domain ontology model are divided into the following 5 categories. Integrate safety hazard records, industry standards and regulations, etc., and refine sub-concepts under various concepts in turn.

3.1.1. *Process of production*

The Process of oil and gas production and operation is primarily structured into various production stages and specific operational steps that are pertinent to HSE knowledge in the oil and gas industry. It encompasses 17 primary stages of oil and gas production, which are detailed in Table 1. For brevity, the subdivision of secondary links within each stage has been omitted.

Table 1. First process level of oil and gas production and operation.

ID	First process level
P1	Explore
P2	Oil field production capacity building
P3	Ocean engineering
P4	Oil and gas production management
P5	transport
P6	Engineering construction
P7	Oil refining treatment
P8	Chemical engineering
P9	Oil and gas store
P10	Refined oil sales
P11	Natural gas sales
P12	Machining
P13	Plant inspection and maintenance
P14	Dangerous operation
P15	Safety production management
P16	Supervision and inspection
P17	Emergency

3.1.2. O*bject of applicable*

The applicable oil and gas Objects are primarily classified based on the themes of oil and gas HSE knowledge. It comprises 8 first-level objects related to oil and gas entities, as presented in Table 2. For conciseness, the segmentation lists of second and third-level objects under each first-level object have been omitted.

Table 2. First applicable oil and gas object.

ID	First applicable oil and gas object
O1	HSE management object
O2	Emergency resource
O3	Personnel
O4	Station, project, and pipeline
O5	Equipment and facilities
O6	Medium
O7	Environment
O8	Technology
O9	Information resource
O10	Other object

3.1.3. *Domain of business*

The business Domain of oil and gas HSE knowledge is primarily categorized based on the specific scope of business it covers. It may be associated with the organizational structure of the unit to which the HSE Q&A user belongs, functioning as a criterion for establishing the priority of subsequent hierarchical file categorization rules.

3.1.4. *Theory and policy*

The classification of theory and policy documents is organized into primary categories, encompassing laws, regulations, rules, and standards promulgated by countries, regions, and organizations at various levels, as well as published intellectual property outcomes. There are ten first-level classifications, as presented in Table 3. Within these, certain documents pertaining to user organizations within group enterprises can be further subdivided into group company documents, professional company documents, enterprise documents, documents of enterprise secondary units, production site documents, and other internal documents of the organization, among other secondary classifications. Similar subdivisions are not listed under the other first-level categories for brevity.

Table 3. First theory and policy classification level.

ID	First theory and policy classification level
T1	Law
T2	Regulation
T3	Rule
T4	Normative document
T5	Internal organizational file
T6	Standard
T7	Important indicative spirit
T8	Thesis
T9	Patent
T10	Monograph

3.1.5. *Source of data*

Source of data is primarily categorized based on the various data modes of the information collected at the production site, including text records, video surveillance, sensor parameters, among others. This data is utilized to document abnormal monitoring and early warning information pertaining to personnel status, processes, equipment, and facilities, as well as to store daily risk identification and analysis reports related to various objects.

3.2. *Oil and gas HSE field ontology relations*

Integrate hidden danger inspection record, industry standards and regulations, and other text data, and refine the relations between entities such as kind-of, part-of, instance-of, attribute-of, and meet-above, et al.

- Kind-of signifies a classification relationship where one element is a specific type or subclass of another. For instance, “hot work” is a kind of “hazardous operation”.
- Part-of represents signifies that one element constitutes a component of a larger element or whole, such as a “gasoline hydrogenation unit” that is part-of a “petrochemical project”.
- Instance-of a class signifies that a specific instance belongs to a broader class or concept. For example, “Emergency response regulations for production safety accidents” is an instance-of “HSE policy file”.
- Attribute-of signifies that an element possesses specific attributes, which can be objects (has-Object), processes (has-Process), or domains (has-Domain), among others. For instance, the “Emergency Response Regulations for Production Safety Accidents” has the process attribute of “emergency response”, and the object attribute of “production safety accidents”.
- Meet-above relation signifies that a certain theory or policy, meets or exceeds the requirements of another theory or policy. For example, the requirements of “a company's safety management system” need to meet or exceed the relevant requirements of the “national safety production standards”.

3.3. *Extraction and filling of entity and relation*

Taking policy file entities as an example, perform entity and relation extraction.

Define the attributes of the HSE policy file name instance, including name, update time, execution status, etc. The name attribute, which represents the official Chinese full name of the policy file, is unique and serves as the primary key for indexing and retrieving HSE policy file nodes within the subsequent knowledge graph. It also functions as the unique identifier for storing the corresponding vector representations in the knowledge base associated with the HSE policy file. The update time includes the date of promulgation, entry into force and latest update, as the basis for verification of the applicable version of the file; Execution status includes pre-implementation, implementation, and repealed after publication, which is used to verify whether the file is applicable to the supplementary knowledge conditions asked by users.

Entity identification and relation extraction of policy files are carried out using corresponding NLP models. Extract key entities corresponding to ontology model design concepts, and determine correlation relations between entities form several triplets of HSE policy files. Note that a single HSE policy document PF_i has attributes and related entity information as:

$$PF_i = \{PF_i^{\mathrm{N}}, pfC, pfO, pfP, pfD\} \quad (1)$$

where，PF_i^{N} is an instance information set of policy file names. *pfC* is policy file has a set of classification concept directed by has-Category. *pfO* is policy file has a set of the concept of an object to which has-Object refers. *pfP* is a set of has-Process oriented link concept in policy file. *pfD* is a set of instances of domain concepts that the policy file has a business domain relationship pointing to.

The entities and relations of the extracted policy files are entered into the graph database successively to form the knowledge graph.

3.4. *Knowledge retrieval*

When tackling a new question-answering task, the large model first separates the user's known information from the question-asking information to identify the set of key target attributes for knowledge retrieval. Subsequently, policy search rules are formulated, and knowledge graph search queries are generated to compile a candidate set of theoretical policy documents. To ensure the accuracy and relevance of the provided knowledge, the document update date and valid status are verified, confirming that the knowledge

furnished to the large model for answering is from the latest version and within the validity period

Examine each policy file within the candidate set, and the vector representations of the text blocks contained therein are computed. These vector representations are compared to the user's query vector to assess similarity. Based on the computed similarity scores, the policy document exhibiting the highest similarity to the user's question is selected as the primary reference. Lastly, specific terms, suggestions, or guidance pertinent to the user's query are extracted from the selected policy document, serving as a crucial foundation for the large model to generate a comprehensive and accurate response.

4. Case study

As an example, personnel from the supervisory center of Oil Field A encountered multiple situations requiring HSE expertise during their inspection of operations at a high-altitude production well site. Compare the results of knowledge query with large model query with the existing keyword search tools in the market.

4.1. *Scenario 1: Subdivision domain knowledge query*

The inspector inspects the high work ticket issued on the day of the oil production well site, and judges the high work classification according to the operation situation. Query the applicable work classification standards for heights Compare the question input and answer output of the two tools respectively, as shown in Table 4.

Table 4. Input and output of the two tool questions in Scenario 1.

Tool	Question input and answer output
Key-word search	**Search**: High level operations classification Standard A field **Results**: AQ 3025-2008 Safety Code for high altitude operation of chemical production units 4: High-altitude operation classification: high altitude operation is divided into one, two, three and extra high altitude operation, in line with the provisions of GB/T 3608. 4.1.1 When the operating height is 2m≤h < 5m, it is called the first-level high operation.
Know-ledge enhan-cement model	**Question**: The classification standard of high work in A oilfield **Answer**: According to "Q/SY AXXX Safety Management Standard for high altitude Operation", the high-altitude operation was classified into level four (IV), three (III), two (II) and one (I). a) Level I height operation: the base height of the operation is 2 meters to 5 meters (including 2 meters and 5 meters), and the possible fall radius is 3 meters; b) Level II high operation: the base height of the operation is 5 meters to 15 meters (including 15 meters), and the possible fall radius is 4 meters; c) Level III high operation: the base height of the operation is 15 meters to 30 meters (including 30 meters), and the possible fall radius is 5 meters; d) Level IV high operation: the base height of the operation is more than 30 meters, and the possible fall radius is 6 meters.

It can be seen that the keyword retrieval tool can not cover the specific standards of the production enterprise. According to the type of design label, the large model finds the keywords: business field (subclass of Domain) and production and operation link (subclass of Process) as the index, and locates the standard Safety management standard of high-altitude operation in the knowledge graph of policy documents. The answer not only includes the division of the height of the operation base, but also adds the radius of the possible fall range, and accurately answers the classification standard applicable to the scene.

4.2. *Scenario 2: Specified contextual knowledge query*

During the inspection at the production well site, inspectors noted a shortage of crane pipe length during crude oil tanker loading. Query the use of crane pipes in this scenario. Compare the question input and answer output of the two tools respectively, as shown in Table 5.

Table 5. Input and output of the two tool questions in Scenario 2.

Tool	Question input and answer output
Key-word search	**Search**: Tanker crane pipe crude oil **Results**: GA1131-2014 General Rules for Fire Safety Management of Storage Places 12.8 The fire safety management of oil handling areas should meet the following requirements: a) The oil unloading system of railway tank cars from the bottom and the oil unloading system of automobile tank cars from horizontal containers of I, II, and III Class A-

	level products shall adopt closed pipeline system as required; b) The crane pipe is of sufficient length to reach the tank bottom when loading armored, Class B, and Class C A-type oil into railway and automobile tank cars from the top.
Know-ledge enhan-cement model	**Question**: The height requirement of crane pipe when loading crude oil tanker **Answer**: No more than 200mm. According to SY/T 7385-2017 Anti-static Safety Technical Specification 3.4.1.5: When using top loading, the crane tube should be placed at the bottom of the tank. The distance between the crane pipe outlet and the bottom of the tank car should not be greater than 200mm. And GB50350-2015 Design Code for Oil and Gas Gathering and Transportation in Oilfields 2.10: When filling crude oil to railway tankers and automobile tankers with crane pipes, the crane pipes that can be inserted into the bottom of the tank cars should be used. The liquid flow rate in the crane tube shall not be greater than 1m/s before the crane tube is immersed in the crude oil and shall not be greater than 4.5m/s after the crane tube is immersed in the crude oil.

It can be seen that compared with the restriction of the number of keyword search terms, there is no small scene similar to the filling of specified crude oil tank trucks, and after returning "I, II, and III Class A-level oil" in the answer, it is also necessary to manually judge that the crude oil belongs to Class A oil; The large model can directly search the specified scene, and directly answer "Use the oil loading crane tube that can be plugged into the bottom of the tank truck", which is more powerful.

4.3. *Scenario 3: Reasoning-based professional knowledge Q&A*

In the process of operation supervision, the inspectors found that the operators were standing on the roof of the tank car to load oil. The safety belt worn by the operator and the safety rope are connected by a buffer device. To check whether the buffer device is used correctly in this scenario, compare the question input and answer output of the two tools respectively, as shown in Table 6.

Table 6. Input and output of the two tool questions in Scenario 3.

Tool	Question input and answer output
Key-word search	**Search**: High working buffer **Results**: "QSY1236-2009 Operation Safety Management Code" 5.1.1 fall protection should be achieved by taking measures such as eliminating fall hazards, fall prevention and fall control. The order of preference for fall protection measures is as follows: - Try to work on the ground, avoid high work; - Fixed staircases, guardrails, barriers and restraining systems; - The use of work platforms, such as scaffolding or work platforms with lifts; - Use edge limit safety rope to avoid the operator's body near the edge of high work; - Use fall protection equipment, such as full-body harnesses and safety ropes with buffers. If the above protective measures cannot be implemented, do not carry out high work.
Know-ledge enhan-cement model	**Question**: Do I need to wear a buffer when working high up on the tank roof? **Answer**: According to the Safety Management Standard for Work at Heights (Q/SY-AXX), the buffer device is capable of transferring energy or reducing the impact force borne by the worker during the fall prevention process, and its breaking strength should reach 2268kg. Such as fall stoppers, stitched lanyards, specially woven lanyards, ripped or deformed lanyards, elastic life lines, etc. The buffer device is not allowed to operate at heights below 5m. So, the answer to your question is: when the height of the tank roof is estimated to be less than 5 meters, it is generally not necessary to use a buffer device. However, the specific need to use, but also need to judge according to the working environment and risk assessment results.

It can be seen that the correlation between the return answer and the question is small compared with the direct full-text search keywords. After the large model understands the problem logic, it splits the thought chain according to the prompt word template. In order: buffer device use conditions (above 5 meters from the ground) - tank roof height estimation (less than 5 meters from the ground) - judge whether the use is correct or not (should not be used) order, directly give the answer to the question, the buffer device should not be used in the operation high on the tank roof.

5. Conclusion

1) Combined with the oil and gas HSE management system logic, this paper proposes an oil and gas HSE knowledge structure that can efficiently organize production process,

applicable object, business domain, theory and policy, data sources. Provides professional questions and answers for HSE management personnel of different enterprises, levels and positions in a large language model. Provide the required domain knowledge supplement and rigorous reference policy document sources to achieve high accuracy, high recall and highly targeted professional questions and answers in the oil and gas HSE field.

2) Application examples show that after adding the knowledge graph of oil and gas HSE field, it can directly answer the standard specification requirements of the specified scene in the query question of small scenes, and the retrieval function is more powerful; In the use of standard query, according to the business field, link, user unit and other information keywords, match the most suitable for the questioner's standard file; In the reasoning question and answer, the large model can split the thinking chain and analyze successively, and directly answer the question with legal basis for the problem scenario.

3) At present, it only involves the slices of theoretical policy documents, and there are many process documents in the process of accident, risk analysis and hidden danger investigation, which also contain entity attributes. In the next step, we can further improve the relationship of knowledge graph after processing, so as to enrich the basis of answering knowledge.

References

CHEN, Y., ZHANG, L., HU, J., CHEN, C., FAN, X. & LI, X. 2022a. An emergency task recommendation model of long-distance oil and gas pipeline based on knowledge graph convolution network. Process Safety and Environmental Protection, 167, 651-661.

CHEN, Y., ZHANG, L., HU, J., LIU, Z. & XU, K. 2022b. Emergency response recommendation for long-distance oil and gas pipeline based on an accident case representation model. Journal of Loss Prevention in the Process Industries, 77.

DU, L., DING, X., XIONG, K., LIU, T. & QIN, B. 2022. Enhancing pretrained language models with structured commonsense knowledge for textual inference. Knowledge-Based Systems, 254.

FABIANO, B., PETTINATO, M., CURRò, F. & REVERBERI, A. P. 2022. A field study on human factor and safety performances in a downstream oil industry. Safety Science, 153.

HANNAH, G., SOUSA, R. T., DASOULAS, I. & D'AMATO, C. 2025. On the legal implications of Large Language Model answers: A prompt engineering approach and a view beyond by exploiting Knowledge Graphs. Journal of Web Semantics, 84.

HONGBIN YE, T. L., AIJIA ZHANG, WEI HUA, WEIQIANG JIA 2023. Cognitive Mirage: A Review of Hallucinations in Large Language Models. arXiv preprint arXiv, 2309.06794.

HUANG, L., YU, W., MA, W., ZHONG, W., FENG, Z., WANG, H., CHEN, Q., PENG, W., FENG, X., QIN, B. & LIU, T. 2024. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. ACM Transactions on Information Systems.

HUANG, Y., ZHANG, Z., TAO, Y. & HU, H. 2022. Quantitative risk assessment of railway intrusions with text mining and fuzzy Rule-Based Bow-Tie model. Advanced Engineering Informatics, 54.

OPENAI, J. A., STEVEN ADLER, et al. 2023. GPT-4 Technical Report. OpenAI.

P, S. & H N, C. 2023. Context based diversification on keyword search by conceptualization of typed terms of the query. International Journal of Information Management Data Insights, 3.

PATRICK LEWIS, E. P., ALEKSANDRA PIKTUS, et al, DOUWE KIELA 2021. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS 2020.

SHIRUI PAN, L. L., YUFEI WANG, CHEN CHEN, JIAPU WANG, XINDONG WU 2023. Unifying Large Language Models andKnowledge Graphs: A Roadmap. IEEE Transactions on Knowledge and Data Engineering (TKDE) 2024.

SIDDHARTH, L. & LUO, J. 2024. Retrieval augmented generation using engineering design knowledge. Knowledge-Based Systems, 303.

TOM B. BROWN, B. M., NICK RYDER, et al, DARIO AMODEI 2020. Language Models are Few-Shot Learners. Advances in neural information processing systems, 33, 1877-1901.

ZENG, Z., CHENG, Q., HU, X., ZHUANG, Y., LIU, X., HE, K. & LIU, Z. 2025. KoSEL: Knowledge subgraph enhanced large language model for medical question answering. Knowledge-Based Systems, 309.

ZHANG, X., HOU, X., CHEN, X. & ZHUANG, T. 2013. Ontology-based semantic retrieval for engineering domain knowledge. Neurocomputing, 116, 382-391.