processes

MDPI

Article

Process-Oriented Dual-Layer Knowledge GraphRAG for Reservoir Engineering Decision Support

Bin Jiang [1,*], Zhaonian Liu [1,2], Ning Wang [1], Zhuoyang Li [1], Yinliang Shi [1] and Botao Lin [2]

[1] CNOOC Research Institute Co., Ltd., Beijing 100028, China
[2] College of Artificial Intelligence, China University of Petroleum (Beijing), Beijing 102249, China
* Correspondence: jiangbin@cnooc.com.cn

Abstract

This study presents a dual-layer GraphRAG framework for petroleum engineering question answering, in which instance-level facts and domain-level concepts are explicitly separated and integrated into retrieval-augmented generation. To evaluate the framework, a benchmark of 40 expert-constructed Q&A pairs was developed, covering factual, definitional, and explanatory queries derived from a real offshore oilfield dataset. Results show that the dual-layer graph consistently outperforms a single-layer baseline. Answer accuracy improves from 0.65 to 0.70, faithfulness from 0.54 to 0.61, and context relevance from 0.69 to 0.72, confirming that the system retrieves factual parameters more reliably and provides conceptually grounded explanations. Gains in evidence recall and coverage are more modest, highlighting areas for further optimization. A case study illustrates the framework's ability to expand petroleum terminology (e.g., "sandstone → clastic rock"), producing responses that are not only quantitatively more reliable but also qualitatively more informative. The dual-layer design effectively addresses the semantic consistency gap in petroleum QA, offering practical value for reservoir evaluation, lithology interpretation, and technical decision support. These findings demonstrate the potential of GraphRAG to enhance knowledge management and intelligent services in petroleum engineering.

Keywords: reservoir engineering; knowledge graph; Retrieval-Augmented Generation (RAG); GraphRAG; process-oriented knowledge modeling; intelligent decision support

check for updates

Academic Editor: Dicho Stratiev

Received: 10 September 2025
Revised: 4 October 2025
Accepted: 6 October 2025
Published: 10 October 2025

Citation: Jiang, B.; Liu, Z.; Wang, N.; Li, Z.; Shi, Y.; Lin, B. Process-Oriented Dual-Layer Knowledge GraphRAG for Reservoir Engineering Decision Support. *Processes* **2025**, *13*, 3230. https://doi.org/10.3390/pr13103230

1. Introduction

With the oil and gas industry entering a phase of refined management and intelligent transformation, enterprises face growing demands for organizing and reasoning over complex domain knowledge. Over decades, petroleum companies have accumulated vast quantities of unstructured textual assets—exploration/development reports, well histories, construction plans, and incident summaries—whose heterogeneous formats impede automated identification, structured storage, and efficient reuse [1–4]. Meanwhile, industrial initiatives are piloting large-model-assisted knowledge services and governance in exploration & production (E&P), yet practical deployment still contends with data silos, terminology inconsistency, multi-source data integration, and traceability requirements [5–8].

Recent progress in Large Language Models (LLMs) has revitalized knowledge-intensive NLP. However, parametric knowledge alone cannot ensure timeliness or verifiability. Retrieval-Augmented Generation (RAG) addresses this by coupling retrieval with generation to ground outputs in external evidence [9–12]. Beyond flat text retrieval, advances in link-aware pretraining and retrieval (e.g., LinkBERT, RETRO, BEIR) improve document linkage and benchmarked generalization [13–15]. Still, many engineering questions require relational reasoning across entities, hierarchies, and causal/geo constraints. Surveys on Graph Retrieval-Augmented Generation (GraphRAG) summarize how graph structures and graph retrieval fuse with LLMs to extend semantic coverage and interpretability [16–18].

Technically, integrating knowledge graphs (KGs) with LLMs strengthens multi-hop and structure-aware QA. Representative lines include KG-aware reasoning with language models [19], ranking-guided graph augmentation for domain QA [20], graph chain-of-thought prompting [21], and link-prediction signals that steer generation [22]. Engineering practice further highlights failure modes and hardening strategies across the RAG pipeline (indexing, retrieval, fusion, and evaluation) [23]. For document-centric settings, document GraphRAG constructs knowledge graphs from intra-/inter-document structures to support interpretable retrieval paths and grounded answers [24]. Joint modeling of dense retrievers with graph readers has also been explored to tighten retriever–reader coupling [25].

In process engineering and the petroleum domain, KGs have been used to standardize safety reports and risk knowledge [26], to couple graph structure with predictive models such as pipeline-corrosion estimation [27], and to build domain ontologies and large-scale E&P graphs that unify multi-source data and support production analytics [28,29]. Industrial safety and petrochemical applications demonstrate KG-driven HAZOP/incident knowledge extraction and reasoning [30,31]. For pipeline operations, graph neural networks have been used to assign calorific values across network topologies, illustrating the benefit of graph-structured modeling for energy transport systems [32].

At the method level, hybrid retrieval across textual corpora and relational knowledge bases improves coverage for "hybrid" questions [33]; LLM-assisted graph prompting and traversal agents enhance multi-document reasoning over KG-structured contexts [34]; and conflict-aware decoding mitigates knowledge conflicts between retrieved evidence and parametric memory [35]. Graph-infused fusion-in-decoder architectures [36], trustworthiness frameworks tailored to RAG [37], adaptive self-aware retrieval policies [38], and chain-guided multi-hop retriever–reader designs [39] further improve faithfulness, controllability, and efficiency.

This paper targets reservoir geology Q&A and proposes a dual-layer knowledge GraphRAG system tailored to petroleum engineering scenarios. Our contributions are threefold: (1) a dual-layer KG integrating authoritative textbook concept graphs with enterprise document instance graphs to normalize terminology and align definitions; (2) an entity–relation–alignment retrieval module that selects subgraphs by relevance and interpretable relation paths; (3) a generation-control module enforcing source attribution and domain style to deliver professional, traceable answers. Case studies on typical reservoir tasks show improvements over strong RAG baselines in professional consistency, knowledge hit rate, and answer accuracy.

2. Materials

2.1. Case Background and Data Sources

This study uses an internal geological and reservoir engineering report (hereafter referred to as the "Field A dataset") from a newly discovered offshore depression-zone oilfield in the southern sea area. For confidentiality reasons, the actual oilfield name, well identifiers, and formation names have been anonymized.

The dataset consists of multiple wells with complete drilling and coring records. In total, three wells were drilled with a cumulative footage exceeding 8000 m. Side-wall coring yielded nearly one hundred cores, with high recovery rates and most samples indicating oil-bearing characteristics.

Comprehensive logging data were acquired, including natural gamma spectroscopy, resistivity, sonic, density, neutron, NMR, and logging-while-drilling measurements. Formation testing covered dozens of pressure and fluid sampling points, with valid oil, gas, and water samples collected for PVT and laboratory analyses. One drill-stem test in a target formation confirmed stable commercial oil and gas flow. Laboratory work included more than a dozen analysis programs and several hundred samples, covering routine core analysis, microscopic and special core tests, and fluid property evaluation.

Seismic interpretation was based on a 3D survey acquired in recent years, with high-fold coverage and fine binning. The processed volume provides vertical resolution on the order of tens of meters, sufficient to resolve reservoir-scale stratigraphic features. Across the study area, over a dozen seismic horizons were mapped at a uniform picking density, forming the structural framework for subsequent reservoir modeling.

These integrated data sources jointly form a comprehensive and representative dataset, exhibiting the "unstructured + semi-structured" characteristics of petroleum industry corpora. The dataset is suitable for supporting tasks such as entity recognition, relation extraction, and context modeling, and also provides a realistic foundation for evaluating semantic retrieval, knowledge reasoning, and professional response capabilities of the proposed intelligent question–answer (Q&A) framework.

2.2. Domain-Specific Question–Answer Dataset

To quantitatively evaluate the proposed framework, a benchmark set of 40 domain-specific question–answer (Q&A) pairs was constructed by petroleum engineering experts based on the Field A dataset. The Q&A pairs were designed to reflect realistic information needs encountered in reservoir engineering decision-making processes. The dataset covers two categories of questions:

Query-type (30 pairs): factual and parameter-focused questions directly concerning the case study (e.g., main reservoir lithology, average porosity values, permeability ranges, oil saturation levels, and thickness variations).

Analysis-type (10 pairs): reasoning-oriented questions requiring integration of geological and engineering knowledge (e.g., implications of porosity–permeability distribution for reservoir evaluation, impact of effective thickness variation on production).

This Q&A dataset serves as the ground truth benchmark for assessing the framework's performance in semantic retrieval, knowledge reasoning, and professional response capabilities, ensuring both domain accuracy and practical relevance to petroleum engineering workflows.

3. Methods

3.1. Framework Overview

To enhance the accuracy and authority of large language models in petroleum exploration and development, this study proposes a dual-layer knowledge graph retrieval-augmented generation (GraphRAG) framework tailored for specialized engineering scenarios (Figure 1). Our method builds upon the open-source framework nano-graphrag [40] through deep customization and optimization. In particular, we integrate domain-specific knowledge graph construction with structured prompt design strategies, explicitly embedding semantic labels and graph paths into the prompts. This ensures that Q&A outputs achieve higher logical interpretability, improved terminology citation standardization, and better alignment with engineering discourse [36–38], which is particularly valuable for addressing multi-coupling problems in reservoir geology and engineering [41].

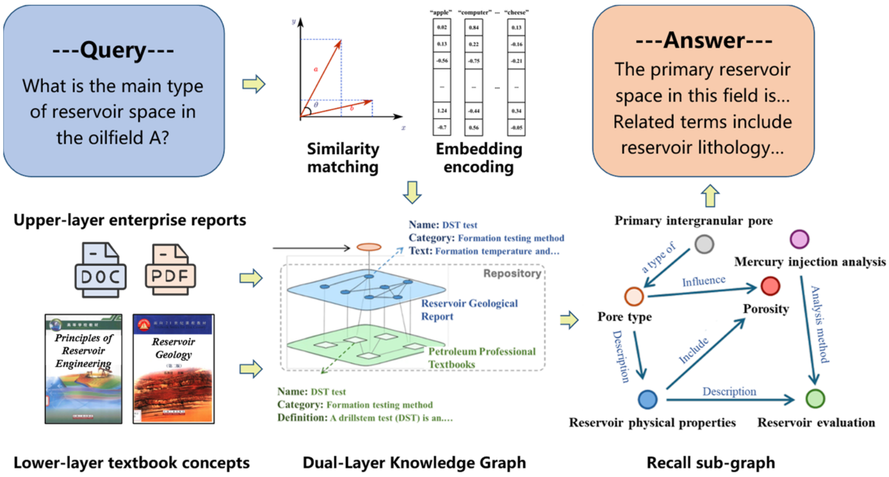

Figure 1. The Retrieval-Augmented Generation Framework Based on a Dual-Layer Knowledge Graph for Reservoir Geology.

The system architecture consists of three key modules:

Dual-layer knowledge graph construction. The conceptual layer is derived from authoritative petroleum engineering textbooks, structuring essential concepts, hierarchical dependencies, and causal/spatial relationships; the instance layer integrates enterprise document knowledge, including well locations, reservoir parameters, and operational records, thereby enabling precise semantic retrieval for specific engineering problems.

Terminology mapping mechanism. A cross-layer semantic alignment strategy ensures that instance-level terms (e.g., project-specific abbreviations or field parameters) can trace back to their standardized conceptual definitions. This mechanism enhances both interpretability and credibility, avoiding ambiguity that often arises in domain-specific language use.

Generation control module. We design differentiated prompt templates that dynamically switch between "definition-tracing" and "engineering Q&A" modes based on user intent. This ensures outputs remain professional, verifiable, and contextually rich while preventing hallucination or terminological drift.

By customizing the nano-graphrag architecture with dual-layer graph structures, semantic mapping, and controlled prompting, our approach not only improves contextual relevance and terminology consistency in professional Q&A, but also strengthens traceability and structural interpretability. It is particularly suited for multi-source knowledge fusion and intelligent decision-making systems in the oil and gas industry.

3.2. Dual-Layer Knowledge Graph Construction

To achieve professional semantic understanding and upstream-downstream knowledge integration, this paper designs and constructs a dual-layer knowledge graph architecture characterized by hierarchical clarity and semantic complementarity. The foundational layer, termed the conceptual graph, systematically organizes core terminology and ontological relationships based on authoritative textbooks. The upper layer, defined as the instance graph, focuses on operational data and conclusion expressions from enterprise-level documents, emphasizing application semantics and pragmatic structures of domain terminology in specific scenarios.

The bottom-layer conceptual graph is built using a typical entity-relationship model, with primary data sources derived from the textbook Reservoir Geology. Graph nodes include categories such as "Terminology," "Definition," "Classification," and "Related Terms," while edge types encompass semantic relationships like "Defined As," "Classified As," "Related Concept," and "Associated With." For instance, the term "Porosity" connects its definition node to the category node "Reservoir Physical Property Parameters" via the "Classified As" edge, and establishes "Related Concept" edges with terms like "Permeability" and "Effective Porosity" to construct its ontological structure within the conceptual network. The specific example is shown in Table 1. The final graph is stored in .graphml format, facilitating standardized access and integration in subsequent processing workflows.

Table 1. Examples of node types in the upper-layer knowledge graph.

Node Type	Example
Well Location	A7 Well
Parameter	Porosity = 18.2%, Permeability = 45.6 mD
Conclusion	Good reservoir connectivity

The upper instance graph focuses on enterprise documents such as reservoir development reports and feasibility studies, emphasizing the extraction of concrete entities, numerical parameters, and operational conclusions from texts. By employing lightweight terminology recognition, part-of-speech combination analysis, and dependency parsing techniques, structural information within documents is identified. This establishes node types such as "Well Location," "Geological Parameters," "Structural Features," "Development Measures," and "Production Capacity Conclusions," alongside edge types like "Occurred In," "Evaluated As," "Located In Structure," and "Involves Terminology." Unlike the foundational graph, the instance graph emphasizes dynamic associations and contextual dependencies of domain terminology in practical workflows, providing context-sensitive knowledge support for generative tasks. To facilitate subsequent graph-structure retrieval and path generation, this graph also adopts standardized .graphml format storage, ensuring cross-graph structural compatibility.

Through the dual-layer graph architecture, the system achieves bidirectional linkage between terminology definitions and instance contexts, while enabling semantic logic traceability and terminology standardization. This establishes a dual knowledge foundation of authority and structure for subsequent generative tasks.

3.3. Terminology Mapping Mechanism

To achieve semantic connectivity and terminology consistency between the conceptual and instance layers of the dual-layer graph, we introduce a terminology mapping mechanism grounded in semantic embedding and similarity computation. The objective is to align natural-language expressions in the upper instance graph with standardized authoritative terms in the lower conceptual graph. This alignment provides traceable definitions, normalized terminology usage, and authoritative references during question–answering processes.

In practice, we employ the nomic-embed-text model to encode all terminology nodes across both layers. This high-performance embedding model transforms terms of varying granularity and context into fixed-dimensional semantic vectors (dimension = 768). On this basis, cosine similarity is computed between instance-level terms and conceptual-level terminology nodes. Pairs with similarity scores above 0.9 are selected as candidate mappings, representing strong semantic consistency.

To minimize false matches and semantic drift, a rule-based filtering mechanism is applied as a second stage. The filtering process imposes additional linguistic and structural constraints, such as stem consistency (e.g., "porosity" vs. "effective porosity") and classification alignment (e.g., ensuring reservoir parameters map only to reservoir property categories). After validation, confirmed mappings are linked by "semantic mapping" edges between the conceptual and instance graphs, thereby establishing an explicit alignment channel (Figure 2).

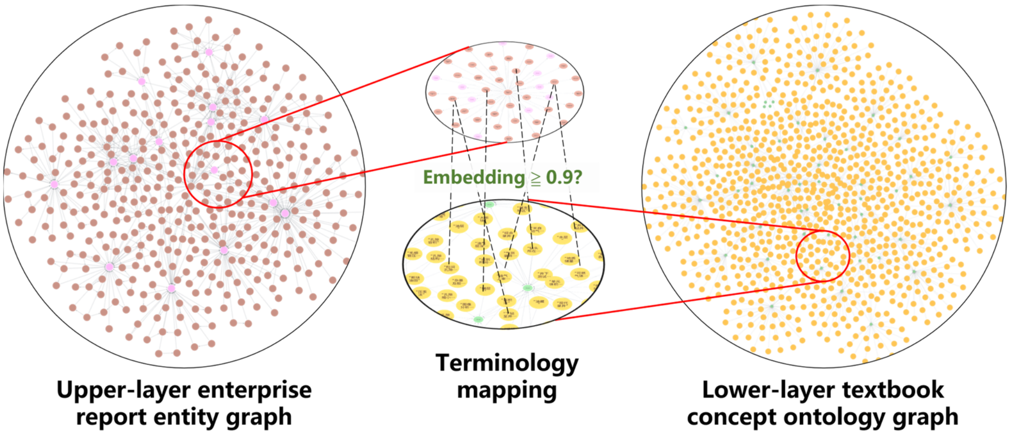

Figure 2. Terminology Mapping Mechanism of the Dual-layer Knowledge Graph.

This mapping mechanism enhances both standardization and interpretability. For example, when a user queries "How tight is the reservoir in this well?", the system maps the colloquial expression "tight reservoir" in the instance graph to its standardized definition and classification criteria in the conceptual graph. Consequently, the retrieval-augmented generation process references authoritative definitions while maintaining logical clarity and professional terminology. This ensures that responses remain both domain-accurate and traceable.

3.4. Controlled Generation Module

To further enhance the professional and practical capabilities of the question-answering system, this paper introduces a retrieval-augmented generation workflow tailored for petroleum exploration and development fields. Based on the classical GraphRAG framework, the proposed approach integrates a "dual-layer knowledge graph structure" and a "prompt-guided mechanism" to build a retrieval-augmented generation process suitable for petroleum exploration and development domains. This workflow fully combines instantiated enterprise data (upper-layer graph) with authoritative professional knowledge (lower-layer graph), achieving accurate citation of domain expertise and controlled content generation through dual-layer graph construction and terminology mapping mechanisms.

The specific workflow includes: after the user submits a natural language question, the system first performs semantic-related community retrieval in the upper-layer in-stance graph to extract information fragments matching the query while identifying in-volved core terminology; subsequently, through the pre-constructed terminology semantic mapping mechanism, these terms are aligned with standard terminology definitions in the lower knowledge graph to extract authoritative definition content; the system then fuses these heterogeneous information sources into a unified context and constructs a structured prompt template, which is input into the large language model for content generation.

To adapt to different types of question scenarios, the system designs two generation control templates, respectively, targeting information-query tasks and summary-analysis tasks:

Information-query tasks typically focus on explicit knowledge such as terminology definitions and parameter descriptions. For this type of question, the prompt template is built upon the semantic mapping results between "terms-standard terms," directly embedding authoritative definitions from the lower graph. It explicitly indicates the cosine similarity matching scores and source paths between terms to enhance traceability. For example, when a user asks "What is effective thickness?" the system can automatically identify the term "effective thickness," judge its semantic consistency with the lower graph's definition node through cosine similarity, extract the standard terminology definition, and return the definition text and knowledge graph source as response basis.

Summary-analysis tasks focus more on explaining complex relationships and reasoning causal mechanisms, such as "The impact of effective thickness variation on pro-duction in a certain well." These questions require the system to comprehensively sched-ule multi-hop entity paths and contextual statements from the upper graph while integrating multiple terminology definitions from the lower graph to construct a complete reasoning chain. To address this, the system adopts a structured prompt template guiding the model to first establish an analytical path, then combine terminology explanations for causal inference. The prompt explicitly labels logical relationships between entities, in-volved terminology definition texts, and semantic matching scores, enabling the model to maintain language generation capabilities while possessing clear logical structure and professional support.

Additionally, this generation control process supports multiple prompt control strategies: it can either adopt a strong constraint approach by directly injecting terminology definitions into the model context to achieve "hard embedding" of professional terminology, or employ a soft guidance strategy by designing prompt intentions to emphasize specific definitions or logical clues, thereby improving the pertinence and diversity of model responses. This module not only enhances the professionalism and credibility of generated content but also provides an expansion foundation for future customized question-answering strategies in industrial application scenarios.

3.5. Experimental Setup

Experiments were conducted on a high-performance server equipped with two Intel Xeon Platinum 8358P processors (Intel Corporation, Santa Clara, CA, USA), 2 TB DDR4 ECC memory, and eight NVIDIA Tesla A800 GPUs (NVIDIA Corporation, Santa Clara, CA, USA). The large language model service was deployed through Ollama (https://ollama.com, accessed on 5 October 2025), using Qwen2.5-72B (Alibaba Cloud, Hangzhou, China) as the generator. For semantic representation, the nomic-embed-text model was applied with an embedding dimension of 768, supporting semantic representation and similarity computation throughout the retrieval-augmented QA process.

For comparison, two system configurations were evaluated under identical prompts and datasets:

Original nano-graphrag framework—lightweight open-source version without domain-specific customization.

Proposed dual-layer GraphRAG—instance + conceptual layers aligned by terminology mapping.

This setup ensures reproducibility and enables a fair comparison across different levels of GraphRAG systems.

3.6. Evaluation Metrics

To ensure comparable experimental results, both graph structures are tested under the same question set, prompt strategy, and generation model configuration. To systematically evaluate the impact of graph architecture on QA performance, six quantitative metrics [42] are introduced across two dimensions: retrieval performance and generation accuracy. These metrics cover core aspects such as semantic alignment, entity hit rate, and evidence completeness, specifically including: Context Relevance, Evidence Recall, Lexical Overlap, Answer Accuracy, Faithfulness, and Evidence Coverage.

This generation control process supports multiple prompt strategies: it can either directly inject terminology definitions into the model context for "hard embedding" of professional terminology or use soft guidance strategies by designing prompt intentions to emphasize specific definitions or logical clues, thereby improving the pertinence and diversity of model responses. This module not only enhances the professionalism and credibility of generated content but also provides an expansion foundation for future customized QA strategies in industrial application scenarios.

The hybrid evaluation mechanism combines manual annotation and LLM embedding similarity. Context Relevance measures the semantic consistency between retrieved graph content and query intent, reflecting whether graph nodes are closely related to the question. Higher scores indicate more focused retrieval results with fewer distractions. Context Relevance is scored manually, with final scores calculated as the average of ratings from two reviewers.

$$Context\ Relevance = \frac{r_1 + r_2}{2},\ r_1, r_2 \in \{0, 0.5, 1\} \tag{1}$$

where r_1, r_2 denote the relevance scores given by two reviewers for the question-answer pair.

Evidence Recall: Measures whether the system has fully retrieved all supporting graph evidence required by the answer, focusing on the coverage of the system's retrieval. If any key claim in the reference answer is not reflected in the context, the score is reduced:

$$Evidence\ Recall = \frac{1}{|\mathcal{R}|}\sum_{c \in \mathcal{R}} 1(S(c, C)) \tag{2}$$

where $\mathcal{R}$ denotes the set of claims to be supported in the reference answer, C denotes the retrieved context, $S(c, C)$ indicates whether the retrieved content C supports claim c, and $1(\cdot)$ is the indicator function.

Lexical Overlap: Assesses the lexical-level similarity between the generated answer and the reference answer, usually computed via Longest Common Subsequence (LCS) or n-gram overlap, serving as a basic metric for surface linguistic similarity.

Answer Accuracy: This metric jointly evaluates the semantic plausibility and factual correctness of the answer, combining semantic embedding similarity with statement-level factual alignment precision. Its composite form is:

$$Answer\ Accuracy = \alpha \cdot FC + (1 - \alpha) \cdot SS \tag{3}$$

where FC denotes Factual Correctness, SS denotes Semantic Similarity, α is the weighting coefficient, set to 0.5 in this paper. The two items are defined as follows:

$$FC = \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + \mathrm{FP} + \mathrm{FN}},\ SS = \cos(\mathbf{f}_i, \mathbf{c}_j) \tag{4}$$

where TP, FP, FN denote the judgment results for fact units, $\mathbf{f}_i, \mathbf{c}_j$ are the embedding vectors of the reference answer and the generated answer.

Faithfulness: Measures whether the generated answer strictly relies on the retrieved graph results, preventing the model from producing "hallucinated" content inconsistent with the context. The faithfulness score is the proportion of statements in the answer that are explicitly supported by the graph:

$$Faithfulness = \frac{|\ \{c \in A |\ S(c, C)\}|}{|\ A|} \tag{5}$$

where A is the set of all statements in the generated answer, C is the graph context, $S(c, C)$ determines whether the statement is supported by the context.

Evidence Coverage: This metric complements recall and focuses on whether the system adequately cites the existing key knowledge points in the graph, reflecting the completeness of knowledge utilization in the generated answer:

$$Evidence\ Coverage = \frac{|\{e \in E|\ M(e, G)\}|}{|E|} \quad (6)$$

where E is the set of knowledge points required to appear in the reference answer, G is the generated answer, $M(e, G)$ indicates whether knowledge point e is manifested in G.

4. Results

4.1. Quantitative Evaluation

To systematically evaluate the performance gains introduced by the proposed du-al-layer graph mechanism, we compared it against a conventional single-layer graph across six evaluation dimensions under identical model and prompt configurations. The radar chart in Figure 3 provides a quantitative view of these results.

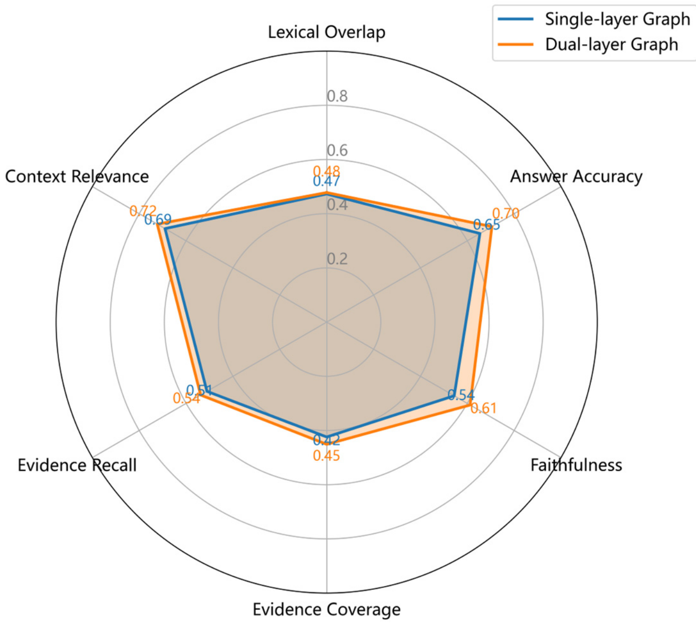

Figure 3. Radar Chart of Evaluation Metric Comparisons.

Overall, the dual-layer graph outperforms the single-layer baseline across all six metrics, though the magnitude of improvement varies. In terms of answer accuracy, the dual-layer graph reaches 0.70 compared to 0.65 for the single-layer baseline, reflecting the benefit of semantic grounding through terminology mapping. Faithfulness improves from 0.54 to 0.61, confirming that enhanced conceptual support reduces hallucination and ensures responses remain aligned with authoritative definitions. Context relevance shows one of the most notable gains, increasing from 0.69 to 0.72, indicating that the integration of instance-level data with conceptual knowledge helps the system better situate answers within broader geological and reservoir contexts.

By contrast, improvements in evidence recall (from 0.51 to 0.54) and evidence coverage (from 0.42 to 0.45) are modest. This limitation stems from the current retrieval process, which primarily operates on instance nodes without explicitly leveraging cross-layer paths. As a result, while the terminology mapping improves semantic precision, the retrieval backbone still requires optimization to fully exploit structural advantages. Lexical overlap remains nearly unchanged (0.47 vs. 0.48), which is expected, as literal token overlap is largely unaffected by deeper semantic modeling.

4.2. Case Study

Beyond quantitative metrics, case studies provide further insight into the advantages of the proposed dual-layer graph framework. Two representative examples are presented below.

Case 1: Reservoir Rock Type Identification

For the query "Q12: What is the reservoir rock type of Field A?", the single-layer graph retrieves only the instance node "lithology: sandstone", generating the response "The reservoir consists mainly of sandstone." By contrast, the dual-layer graph supplements this with textbook-level knowledge that "sandstone is a clastic rock and one of the typical reservoir rock types." Through terminology mapping and structural association, the final response becomes: "The reservoir rock of Field A is of clastic type, mainly composed of sandstone." This example demonstrates improvements in accuracy and faithfulness, while also highlighting the dual-layer system's ability to expand terminology, construct logical chains, and provide contextual explanations.

Case 2: Reservoir Type Characterization

Another representative query asked "Q25: What is the reservoir type of Layer Z, and what are its characteristics?" The single-layer retrieval extracts only direct attributes of the layer—"fine sandstone," "porosity: 22%," "permeability: 171 mD"—resulting in the limited response: "Layer Z is a fine sandstone reservoir with 22% porosity and 171 mD permeability." The dual-layer graph, however, incorporates additional semantic relations: "fine sandstone is a type of reservoir rock," "reservoir type is defined by lithological and petrophysical properties," and "controlling factors include sedimentation, tectonics, and diagenesis." With these expansions, the generated response is enriched as follows: "Layer Z is classified as a fine-sandstone reservoir type developed in delta-front distributary-channel facies. It corresponds to gray fine sandstone with measured porosity of 22% and permeability of 171 mD. The reservoir type is jointly defined by lithology and petrophysical properties. Its features include good sorting, visible sedimentary structures such as trough cross-bedding, relatively thick sand bodies, and clear lateral boundaries, consistent with a delta-front environment."

Together, these cases illustrate how the dual-layer system goes beyond attribute-level retrieval to deliver professional, context-aware answers. By integrating domain definitions and explanatory factors, it provides reservoir engineers with richer and more actionable information, bridging the gap between factual recall and practical decision support.

4.3. Discussion

In the petroleum domain, knowledge graph–based applications have already demonstrated measurable benefits, but their improvements remain modest compared with the dual-layer GraphRAG framework. Prior efforts have explored ontology-based conceptual modeling [28], engineering-oriented KG construction for project management [43], and structuring unstructured text such as safety reports [26], with surveys emphasizing persistent challenges of semantic heterogeneity and multi-source integration [4]. At the system level, PetroKG improved log-interpretation accuracy through heterogeneous upstream data, yet its reasoning ability was limited to factual retrieval without explanatory capacity [29]. Petro KGraph reported gains in recall and relevance over keyword search by extracting entities and relations from technical documents, but still struggled with domain-level interpretation [44]. By contrast, our dual-layer design introduces explicit logical chains that contextualize petroleum-specific terminology (e.g., "sandstone → clastic rock"), thereby addressing the semantic consistency gap and producing responses that are both quantitatively more reliable and qualitatively more informative.

The observed improvements have direct practical significance. Higher factual precision reduces the risk of misinterpretation in critical parameters such as porosity and saturation, while improved faithfulness minimizes unsupported or hallucinated statements. Contextual relevance ensures that results are expressed in petroleum engineering terminology, facilitating seamless integration into professional workflows. These qualities collectively enhance the reliability of intelligent QA systems as decision-support tools in reservoir evaluation and field development.

Nevertheless, several weaknesses should be acknowledged. The system is sensitive to terminology variation: users employing synonyms, abbreviations, or non-standard expressions may encounter mismatches in retrieval. Ambiguous or underspecified queries also remain challenging, as the model may generate plausible but unverified explanations. Furthermore, while the current case study demonstrates feasibility, scalability to enterprise-level datasets—including thousands of wells, millions of log curves, and decades of seismic surveys—poses substantial technical challenges. Addressing these issues will require advances in distributed graph storage, multimodal integration, and robust terminology normalization to ensure applicability across large-scale petroleum data environments.

5. Conclusions

This study introduced a dual-layer GraphRAG framework for petroleum QA that explicitly separates instance-level facts from domain-level concepts. Evaluated on 40 expert-designed Q&A pairs, the framework consistently outperformed a single-layer baseline, demonstrating measurable gains in accuracy, faithfulness, and contextual relevance. The primary contribution lies in its ability to construct explicit logical chains that bridge factual retrieval with conceptual reasoning, enabling responses that are both technically precise and professionally meaningful.

The practical value is clear: the framework provides reservoir engineers with accurate, faithful, and terminology-aligned answers that directly support tasks such as lithology interpretation, reservoir property estimation, and field evaluation. At the same time, limitations remain. The current validation is restricted to a single anonymized dataset; the system is sensitive to terminology variation and ambiguous prompts; and scalability to enterprise-level multimodal databases has not yet been tested.

Future research should expand testing across multiple oilfields, incorporate more diverse data sources (e.g., seismic attributes, production history), and develop advanced reasoning strategies such as path-aware retrieval, graph neural encoders, and multi-hop inference. These directions will extend the applicability of the framework and advance the role of GraphRAG in petroleum engineering decision-support systems.

Author Contributions: Conceptualization, B.J. and Z.L. (Zhaonian Liu); methodology, B.J.; validation, B.J., Z.L. (Zhuoyang Li); investigation, N.W.; resources, Z.L. (Zhuoyang Li); data curation, Y.S.; writing—original draft preparation, B.J.; writing—review and editing, B.L.; supervision, B.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data supporting the findings of this study are not publicly available due to confidentiality agreements related to internal oilfield documents and proprietary engineering data. Requests for access to the data can be directed to the corresponding author and may be granted under reasonable conditions and non-disclosure agreements.

Acknowledgments: The authors would like to thank CNOOC Research Institute Co., Ltd. for providing engineering documentation, technical support, and internal validation resources. During the preparation of this manuscript, the authors used ChatGPT-4o for the purpose of language polishing. The authors have reviewed and edited all generated content and take full responsibility for the final version of the manuscript.

Conflicts of Interest: Authors Bin Jiang, Zhaonian Liu, Ning Wang, Zhuoyang Li and Yinliang Shi were employed by CNOOC Research Institute Co., Ltd. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RAG	Retrieval-Augmented Generation
KG	Knowledge Graph
LLM	Large Language Model
QA	Question Answering
Neo4j	A graph database management system
BLEU	Bilingual Evaluation Understudy (metric for text similarity)
ROUGE	Recall-Oriented Understudy for Gisting Evaluation (metric)

References

1. Sun, F. Exploration and Practice of Exploration and Development Data Governance in CNOOC. *China Offshore Oil Gas* **2024**, *36*, 169–176. [CrossRef]
2. Xiong, H.; Zhao, C.; Liu, W. Development Directions and Key Implementation Paths of Large Models in the Oil and Gas Industry. *Daqing Pet. Geol. Dev.* **2024**, *43*, 214–224. [CrossRef]
3. Li, X. Design and Implementation of a Crowdsourcing Knowledge Management System for Oil Drilling Based on Blockchain Technology. Master's Thesis, Xi'an Shiyou University, Xi'an, China, 2024. [CrossRef]
4. He, T.; Zhang, Q. Current Applications and Development Trends of Knowledge Graphs in Oil and Gas Exploration and Development. *Nat. Gas Ind.* **2024**, *44*, 55–67. [CrossRef]
5. Pan, H.; Liu, J.; Gong, B.; Zhu, Y.; Bai, J.; Huang, H.; Fang, Z.; Jing, H.; Liu, C.; Kuang, T.; et al. Construction and Preliminary Application of Large Language Model for Reservoir Performance Analysis. *Pet. Explor. Dev.* **2024**, *51*, 1175–1182. [CrossRef]
6. Liu, H.; Ren, Y.; Li, X.; Deng, Y.; Wang, Y.; Cao, Q.; Du, J.; Lin, Z.; Wang, W. Research Status and Application of Artificial Intelligence Large Models in the Oil and Gas Industry. *Pet. Explor. Dev.* **2024**, *51*, 910–923. [CrossRef]
7. Liu, H.; Yin, H.; Luo, Z.; Wang, X. Integrating Chemistry Knowledge in Large Language Models via Prompt Engineering. *Synth. Syst. Biotechnol.* **2025**, *10*, 23–38. [CrossRef]
8. Li, Q. Reservoir Science: A Multi-Coupling Communication Platform to Promote Energy Transformation, Climate Change and Environmental Protection. *Reserv. Sci.* **2025**, *1*, 1–2. [CrossRef]
9. Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.T.; Rocktäschel, T.; et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In Proceedings of the International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 6–12 December 2020.
10. Karpukhin, V.; Oguz, B.; Min, S.; Lewis, P.S.; Wu, L.; Edunov, S.; Chen, D.; Yih, W.T. Dense Passage Retrieval for Open-Domain Question Answering. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online Event, 16–20 November 2020; pp. 6769–6781.
11. Guu, K.; Lee, K.; Tung, Z.; Pasupat, P.; Chang, M. REALM: Retrieval-Augmented Language Model Pre-Training. In Proceedings of the International Conference on Machine Learning, Virtually, 13–18 July 2020; pp. 3929–3938.
12. Asai, A.; Min, S.; Zhong, Z.; Chen, D. Retrieval-Based Language Models and Applications. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 6: Tutorial Abstracts), Toronto, ON, Canada, 9–14 July 2023*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2023; pp. 41–46.
13. Yasunaga, M.; Leskovec, J.; Liang, P. LinkBERT: Pretraining Language Models with Document Links. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Dublin, Ireland, 22–27 May 2022. [CrossRef]
14. Borgeaud, S.; Mensch, A.; Hoffmann, J.; Cai, T.; Rutherford, E.; Millican, K.; Van Den Driessche, G.B.; Lespiau, J.B.; Damoc, B.; Clark, A. Improving Language Models by Retrieving from Trillions of Tokens (RETRO). In Proceedings of the International Conference on Machine Learning and Computing, Guangzhou, China, 18–21 February 2022.

15. Thakur, N.; Reimers, N.; Rücklé, A.; Srivastava, A.; Gurevych, I. BEIR: A Heterogeneous Benchmark for Zero-Shot Evaluation of Information Retrieval Models. *Inf. Retr. J.* **2021**, *25*, 517–569.
16. Peng, B.; Zhu, Y.; Liu, Y.; Bo, X.; Shi, H.; Hong, C.; Zhang, Y.; Tang, S. Graph Retrieval-Augmented Generation: A Survey. *arXiv* **2024**, arXiv:2408.08921. [CrossRef]
17. Han, H.; Wang, Y.; Shomer, H.; Guo, K.; Ding, J.; Lei, Y.; Halappanavar, M.; Rossi, R.A.; Mukherjee, S.; Tang, X.; et al. Retrieval-Augmented Generation with Graphs (GraphRAG): A Survey. *arXiv* **2024**, arXiv:2404.05868.
18. Zhang, Q.; Chen, S.; Bei, Y.; Yuan, Z.; Zhou, H.; Hong, Z.; Chen, H.; Xiao, Y.; Zhou, C.; Chang, Y.; et al. A Survey of Graph Retrieval-Augmented Generation for Customized Large Language Models. *arXiv* **2025**, arXiv:2501.13958.
19. Yasunaga, M.; Ren, H.; Bosselut, A.; Liang, P.; Leskovec, J. QA-GNN: Reasoning with Language Models and Knowledge Graphs for Question Answering. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 6–11 June 2021.
20. Yang, R.; Liu, H.; Marrese-Taylor, E.; Zeng, Q.; Ke, Y.H.; Li, W.; Cheng, L.; Chen, Q.; Caverlee, J.; Matsuo, Y.; et al. KG-Rank: Enhancing LLMs for Medical QA with Knowledge Graphs and Ranking Techniques. *arXiv* **2024**, arXiv:2403.05881.
21. Jin, B.; Xie, C.; Zhang, J.; Roy, K.K.; Zhang, Y.; Li, Z.; Li, R.; Tang, X.; Wang, S.; Meng, Y.; et al. Graph Chain-of-Thought: Augmenting LLMs by Reasoning on Graphs. In *Findings of the Association for Computational Linguistics, Proceedings of the ACL 2024, Bangkok, Thailand, 11–16 August 2024*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2024.
22. He, Z.; Zhu, J.; Qian, S.; Chai, J.; Koutra, D. LinkGPT: Teaching Large Language Models to Predict Missing Links. *arXiv* **2024**, arXiv:2406.04640. [CrossRef]
23. Barnett, S.; Kurniawan, S.; Thudumu, S.; Brannelly, Z.; Abdelrazek, M. Seven Failure Points When Engineering a Retrieval-Augmented Generation System. In Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering—Software Engineering for AI, Lisbon, Portugal, 14–15 April 2024; pp. 194–199.
24. Knollmeyer, C. Document GraphRAG: A Practical Guide to Building Document Knowledge Graphs for RAG Systems. *Electronics* **2025**, *14*, 2102. [CrossRef]
25. Li, R.; Wang, L.; Jiang, Z.; Hu, Z.; Zhao, M.; Lu, X. Mutually improved dense retriever and GNN-based reader for arbitrary-hop open-domain question answering. *Neural Comput. Appl.* **2022**, *34*, 11831–11851. [CrossRef]
26. Yin, Z.; Shi, L.; Yuan, Y.; Tan, X.; Xu, S. A Study on a Knowledge Graph Construction Method of Safety Reports for Process Industries. *Processes* **2023**, *11*, 146. [CrossRef]
27. Xie, R.; Fan, Z.; Hao, X.; Luo, W.; Li, Y.; Zhao, Y.; Han, J. Prediction Model of Corrosion Rate for Oil and Gas Pipelines Based on Knowledge Graph and Neural Network. *Processes* **2024**, *12*, 2367. [CrossRef]
28. Tang, X.; Feng, Z.; Xiao, Y.; Wang, M.; Ye, T.; Zhou, Y.; Meng, J.; Zhang, B.; Zhang, D. Construction and Application of an Ontology-Based Domain-Specific Knowledge Graph for Petroleum Exploration and Development. *Geosci. Front.* **2023**, *14*, 101426. [CrossRef]
29. Zhou, X.-G.; Gong, R.-B.; Shi, F.-G.; Wang, Z.-F. PetroKG: Construction and Application of Knowledge Graph in Upstream Area of PetroChina. *J. Comput. Sci. Technol.* **2020**, *35*, 368–378. [CrossRef]
30. Zhao, Y.; Zhang, B.; Gao, D. Construction of Petrochemical Knowledge Graph Based on Deep Learning. *J. Loss Prev. Process Ind.* **2022**, *76*, 104736. [CrossRef]
31. Mao, S.; Zhao, Y.; Chen, J.; Wang, B.; Tang, Y. Development of Process Safety Knowledge Graph: A Case Study on Delayed Coking Process. *Comput. Chem. Eng.* **2020**, *143*, 107094. [CrossRef]
32. Yang, Z.; Liu, Z.; Zhou, J.; Song, C.; Xiang, Q.; He, Q.; Hu, J.; Faber, M.H.; Zio, E.; Li, Z.; et al. A Graph Neural Network (GNN) Method for Assigning Gas Calorific Values to Natural Gas Pipeline Networks. *Energy* **2023**, *278*, 127875. [CrossRef]
33. Lee, M.C.; Zhu, Q.; Mavromatis, C.; Han, Z.; Adeshina, S.; Ioannidis, V.N.; Rangwala, H.; Faloutsos, C. HybGRAG: Hybrid Retrieval-Augmented Generation on Textual and Relational Knowledge Bases. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics, Vienna, Austria, 27 July–1 August 2025.
34. Yang, Z.; Zhu, Z.; Zhu, X. CuriousLLM: Elevating Multi-Document QA with LLM-Enhanced Knowledge Graph Reasoning. In Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies, Albuquerque, NM, USA, 29 April–4 May 2025.
35. Jin, Z.; Cao, P.; Chen, Y.; Liu, K.; Jiang, X.; Xu, J.; Li, Q.; Zhao, J. Tug-of-War Between Knowledge: Exploring and Resolving Knowledge Conflicts in Retrieval-Augmented Language Models. *arXiv* **2024**, arXiv:2402.14409.
36. Yu, D.; Zhu, C.; Fang, Y.; Yu, W.; Wang, S.; Xu, Y.; Ren, X.; Yang, Y.; Zeng, M. KG-FiD: Infusing Knowledge Graph in Fusion-in-Decoder for Open-Domain Question Answering. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Dublin, Ireland, 22–27 May 2022.
37. Zhou, Y.; Liu, Y.; Li, X.; Jin, J.; Qian, H.; Liu, Z.; Li, C.; Dou, Z.; Ho, T.Y.; Yu, P.S. Trustworthiness in Retrieval-Augmented Generation Systems: A Survey. *arXiv* **2024**, arXiv:2409.10102. [CrossRef]
38. Yao, Z.; Qi, W.; Pan, L.; Cao, S.; Hu, L.; Liu, W.; Hou, L.; Li, J. SeaKR: Self-Aware Knowledge Retrieval for Adaptive RAG. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics, Vienna, Austria, 27 July–1 August 2025.

39. Xu, W.; Deng, Y.; Zhang, H.; Cai, D.; Lam, W. Exploiting Reasoning Chains for Multi-Hop Science Question Answering. In *Findings of the Association for Computational Linguistics, Proceedings of the EMNLP 2021, Punta Cana, Dominican Republic, 7–11 November 2021*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021.
40. Gusye. Nano-Graphrag. Available online: https://github.com/gusye1234/nano-graphrag (accessed on 16 July 2025).
41. Li, Q.; Wu, J.; Li, Q.; Wang, F.; Cheng, Y. Sediment Instability Caused by Gas Production from Hydrate-Bearing Sediment in Northern South China Sea by Horizontal Wellbore: Sensitivity Analysis. *Nat. Resour. Res.* **2025**, *34*, 1667–1699. [CrossRef]
42. Xiang, Z.; Wu, C.; Zhang, Q.; Chen, S.; Hong, Z.; Huang, X.; Su, J. When to Use Graphs in RAG: A Comprehensive Analysis for Graph Retrieval-Augmented Generation. *arXiv* **2025**, arXiv:2506.05690. [CrossRef]
43. Xia, T.; Dai, Z.; Zhang, Y.; Wang, F.; Zhang, W.; Xu, L.; Zhou, D.; Zhou, J. Construction Method and Practical Application of Oil and Gas Field Surface Engineering Case Database Based on Knowledge Graph. *Processes* **2024**, *12*, 1088. [CrossRef]
44. Cordeiro, F.C.; Silva, P.F.; Gomes, D.S.M.; Souza, R.R.; Coelho, F.C.; Ell, B. Petro KGraph: A methodology for extracting knowledge graph from technical documents—An application in the oil and gas industry. *SSRN* **2024**, preprint. [CrossRef]