A Method for Constructing Causal Knowledge Graphs in Oil and Gas Fields that Integrates Data-Driven Approaches and Large Language Models

Qingzeng Cao
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 2771978945@qq.com

Jie Luo
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 3030516025@qq.com

Xin Jing
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: jingxin@xatu.edu.cn

***Abstract*—To address the issues of lacking unified representation for multi-source heterogeneous data in oil and gas field development, poor interpretability of traditional data-driven methods, and the difficulty in structuring expert experience, a causal knowledge graph construction method integrating data-driven and large language models (LLMs) is proposed. This method first learns the causal skeleton from structured data using causal discovery algorithms (PC) and then utilizes LLMs to extract domain knowledge from engineering documents, performing semantic completion for edges with undetermined directions. On this basis, an LLM-based agent is introduced to validate the physical consistency and compliance of the causal graph according to safety constraints and engineering rules, and an incremental update mechanism is designed to adapt to dynamic production environments. Experiments on actual oil field datasets demonstrate that the causal graphs generated by this method are highly consistent with expert annotations, achieving an F1 score of 0.90, significantly outperforming traditional methods in metrics such as true positive rate and precision, effectively realizing the synergy between data patterns and domain knowledge, and providing a new pathway for explainable intelligent analysis in oil and gas fields.**

Keywords-Oil and Gas Field Development; Causal Knowledge Graph; Large Language Models; LLM Evaluation Agent

I. INTRODUCTION

Oil and gas field development is a complex systematic project that covers multiple stages such as geology, drilling, and oil extraction, with each stage tightly coupled. The application of Internet of Things and big data technologies has led to the accumulation of massive amounts of production monitoring data on-site. This data provides a foundation for operational analysis and decision-making optimization. However, this data is scattered across different business systems, lacks a unified knowledge expression structure, and is difficult to form a reusable knowledge system, limiting the deep mining of data value.

Traditional data-driven methods (such as deep learning) have achieved good results in tasks like yield prediction and fault identification, but their decision-making processes lack interpretability, being regarded as 'black box' models that are difficult to reveal causal or mechanistic relationships between variables [1]. Engineering decision-making must be grounded in physical mechanisms, as predictions alone are insufficient. Expert knowledge in the field is mostly present in document form, lacking systematic expression, and manually constructing a knowledge base is costly and hard to adapt to complex working conditions in a timely manner.

In this context, combining data-driven analysis with structured knowledge expression has become an important direction for the intelligence of oil and gas fields. Knowledge graphs can integrate production data and engineering experience and have been preliminarily applied in equipment management and process analysis [2]. However, traditional knowledge graphs primarily focus on static relationships between entities and lack the ability to describe causal links between variables, making it difficult to support mechanism analysis and decision reasoning.

The development of causal reasoning theory has provided a new perspective for the analysis of complex system mechanisms. Causal knowledge graphs (causal graphs) introduce causal direction into traditional knowledge graphs, enabling the explicit expression of causal relationships between variables, which aligns more closely with the cognitive style of engineers. However, in practical construction, purely data-driven causal discovery methods still face limitations: field data often contains noise, missing values, and sampling differences, leading algorithms to typically identify correlations rather than determine causal directions; moreover, relying solely on statistical correlation may produce spurious causality lacking physical meaning.

In recent years, large language models (LLMs) have demonstrated strong capabilities in natural language understanding, enabling them to identify potential domain knowledge and causal descriptions from unstructured texts such as engineering documents and technical reports. If the variable influence relationships embedded in these texts can be effectively extracted, it will help to address the limitations of data-driven methods in causal direction determination. Therefore, integrating the semantic comprehension capabilities of LLMs with data-driven causal discovery methods holds promise for constructing a causal knowledge acquisition pathway that balances data patterns and domain knowledge, thereby enhancing the reliability and rationality of causal relationship identification.

Based on the above background, this paper proposes a method for constructing and refining a causal knowledge graph of oil and gas fields that integrates data-driven approaches with large language models. This method collaboratively utilizes statistical causal discovery results and textual semantic knowledge, introducing domain knowledge to refine the causal structure while ensuring data consistency, thereby constructing a causal knowledge graph that is both data-driven and engineering-wise.

The main contributions of this paper include:

- Design a causal graph generation framework that synergizes with data and semantics, utilizing causal discovery algorithms to learn causal skeletons from production data, and employs LLMs to complete causal direction for edges with undetermined directions.
- Propose an evaluation agent mechanism based on LLM, formalizing oil and gas field safety constraints and engineering rules into evaluation criteria, performing multi-dimensional verification on causal graphs, and enhancing the engineering reliability and interpretability of causal structures.
- Construct an incremental update mechanism based on new data-driven, enabling the causal knowledge graph to dynamically evolve with the accumulation of production data, maintaining consistency between the knowledge structure and actual operating conditions.
- Experiments conducted on multiple representative datasets show that the method can effectively improve the accuracy and rationality of causal structure identification.

II. RELATED WORKS

A. Causal Knowledge Graphs

Knowledge Graphs (KGs) organize knowledge in a graph structure, where nodes represent entities and edges represent semantic relationships, typically expressed in the form of triples (head entity, relationship, tail entity). General-purpose knowledge graphs like ConceptNet and Wikidata are widely used in the field of natural language processing [3]. However, in terms of causal expression, traditional knowledge graphs often only describe causal relationships through simple relations such as 'causes', making it difficult to characterize mediating variables, causal direction, and multi-path mechanisms, thus limiting their application in intervention analysis and counterfactual reasoning [1].

Causal Knowledge Graph (CKG) explicitly introduces causal directions and supports richer causal semantics on the basis of traditional knowledge graphs. For example, the CausalKG [4] framework utilizes the RDF-star extension. This allows causal edges to be annotated with meta-information, such as causal effect values and confidence scores, thereby representing causal relationships with rich meta-attributes. For instance, in an autonomous driving scenario, by constructing a causal graph of 'distracted driving sudden lane change collision' and quantifying different effects, it can provide interpretable causal explanations for accident analysis.

From a theoretical perspective, causal knowledge graphs can be viewed as a graph-structured implementation of structural causal models (SCMs). SCMs utilize directed acyclic graphs (DAGs) to characterize causal dependencies between variables and perform intervention analysis and counterfactual inference through the do-calculus. Building on this, the Causal-aware LLMs [5] framework integrates SCMs with large language models, enhancing the model's decision-making capabilities by modeling the environment through causal graphs.

At the application level, the mental health knowledge graph MDKG [6] integrates causal paths from literature, revealing multiple causal chains related to depression. An example of a causal chain is shown in Fig. 1, demonstrating the value of causal knowledge graphs in studying complex disease mechanisms. The Causal-first Graph-RAG [7] method prioritizes the use of causal relationships during the retrieval phase for path filtering, reducing noisy associations and enhancing reasoning interpretability.

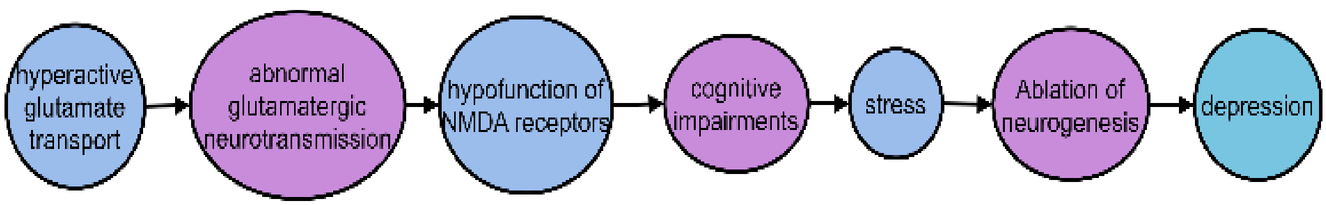


Figure 1. Causal chain of depression

Overall, causal knowledge graphs enable more accurate representation of causal mechanisms between variables by introducing causal direction and rich semantic information, providing an important foundation for interpretable artificial intelligence, intervention analysis, and counterfactual reasoning.

B. Causal Discovery Algorithms

Causal discovery is a key step in learning causal structures among variables from observational data, with existing methods primarily divided into three categories: constraint-based, score-based, and function-based methods.

Constraint-based methods (such as PC, FCI) identify variable associations and determine directions through conditional independence tests. The PC algorithm can output partial directed acyclic graphs (CPDAGs) under the assumptions of causal sufficiency and faithfulness. For example, the CAMA [8] framework utilizes the PC algorithm to construct mathematical causal graphs (MCGs) and identify knowledge point dependencies. However, such methods are sensitive to conditional independence tests, prone to errors in small samples or complex relationships, and the results are often equivalence classes, requiring prior knowledge to determine some directions.

Score-based methods (such as GES, NOTEARS) search for optimal graph structures through scoring functions (such as BIC), where NOTEARS employs continuous optimization to enhance efficiency but requires the assumption of functional relationships [9]. Function-based methods (such as LiNGAM) utilize the non-Gaussianity of noise to identify directions, suitable for linear non-Gaussian data, but have higher distribution requirements.

Despite the progress mentioned above, these methods still face challenges in industrial scenarios such as oil and gas fields. Real-world data often contains non-linear, non-Gaussian noise, and confounding factors, which may violate algorithmic assumptions. Additionally, traditional methods struggle to integrate domain-specific prior knowledge (e.g., physical mechanisms) and tend to learn statistical correlations rather than true causal relationships [10].

Overall, causal discovery provides an important tool for obtaining causal structures, but its effective application depends on algorithmic assumptions, integration of domain knowledge, and validation mechanisms. In complex scenarios such as oil and gas fields, relying solely on data-driven approaches is difficult to obtain reliable causal graphs, requiring a comprehensive analysis that combines expert knowledge and validation.

C. Large Language Models and Domain Knowledge Extraction

Large language models (LLMs), with their strong natural language understanding capabilities and rich pre-trained knowledge, have become an important tool for extracting domain knowledge from unstructured text. Compared to traditional methods, LLMs can quickly adapt to new domains through few-shot prompting or fine-tuning with instructions [11], reducing manual annotation costs while improving knowledge extraction efficiency.

Currently, LLMs are used in multiple fields for automatically extracting structured knowledge. For example, in the field of mental health, MDKG utilizes GPT-4 to extract entities, relationships, and conditional statements from literature abstracts, and assigns reliability scores to triples based on confidence and literature sources; in the field of mathematics, the CAMA framework leverages LLMs to extract atomic knowledge points from problem-solution text for constructing mathematical causal graphs (MCG); in the field of materials science, the KG-FM study employs Qwen2-72B to extract entity relationships from literature abstracts through structured prompts and output them in JSON format, achieving large-scale knowledge graph construction [12]. The above studies indicate that LLMs not only can identify entity relationships but also organize them into structured knowledge, providing effective support for knowledge graph construction.

Although LLMs demonstrate high efficiency in knowledge extraction, they still face several challenges. First, models may generate 'hallucinations', producing content inconsistent with the original text, requiring correction through external knowledge or causal verification mechanisms [13]. Second, causal relationships are often implicit in complex sentence structures, and simple entity extraction may fail to retain intermediate variables in the causal chain, necessitating supplementation through textual structure analysis [14]. Third, in vertical domains, differences in specialized terminology and document structure can affect extraction performance, typically requiring the introduction of domain-specific dictionaries or lightweight fine-tuning [15].

In summary, LLM provides an efficient technical approach for domain knowledge extraction, capable of identifying entity relationships from unstructured text and constructing structured knowledge representations. In the oil and gas field scenario, extraction templates can be designed for engineering documents and technical reports and combined with structured production data to achieve multi-source knowledge fusion and causal modeling.

III. CONSTRUCTING CAUSAL KNOWLEDGE GRAPHS FOR OIL AND GAS FIELDS WITH DATA AND SEMANTIC COLLABORATION

A. Overall framework

The overall framework of the method in this paper is shown in Fig. 2, which includes four core modules: (1) multi-source data preprocessing, which involves cleaning and aligning structured data and unstructured text; (2) data-driven causal skeleton generation, which uses causal discovery algorithms to identify partial causal directions and outputs a causal skeleton containing undirected edges; (3) LLM semantic completion, which targets edges with undetermined directions in the skeleton and invokes the LLM to make directional judgments based on domain knowledge; (4) LLM evaluation agent verification, which conducts consistency evaluation on the complete causal graph based on safety constraints and engineering rules, and outputs a trustworthy causal graph.

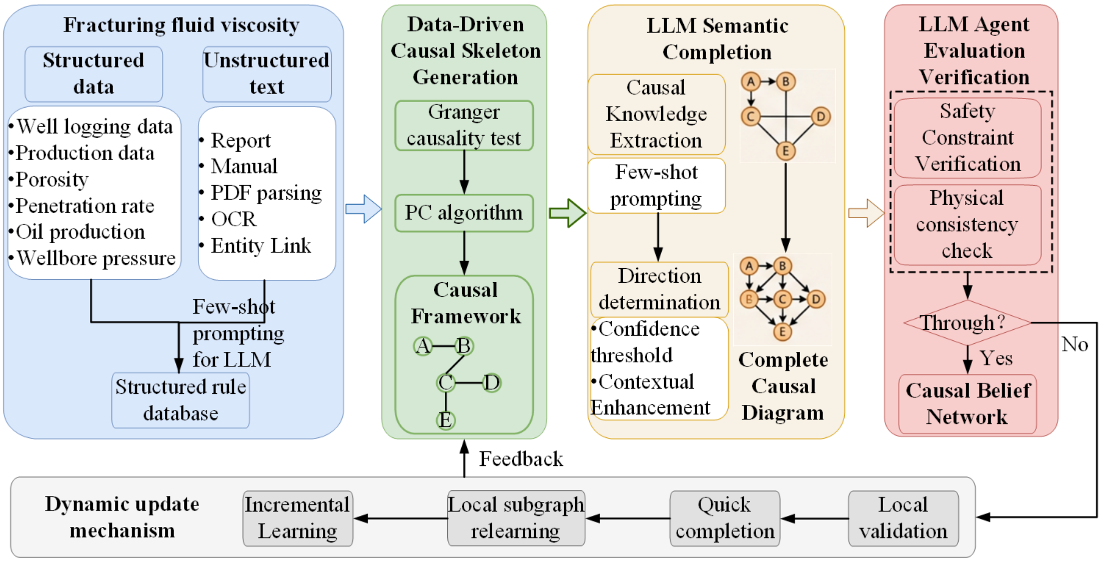


Figure 2. Overall Research Framework

B. Multi-source heterogeneous data processing and fusion

Multi-source heterogeneous data processing and fusion: Structured data (such as well logging, production data) is cleaned, normalized, and processed for feature engineering after which key variables such as Depth, porosity, permeability, oil production are extracted. Examples of the extracted key variables are shown in Table I.

TABLE I. A SUMMARY OF KEY VARIABLES

Number sign	Depth(m)	Porosity(%)	Permeability(mD）
W1	2500	15.2	120
W2	2700	12.8	95
W3	2300	18.5	210

Unstructured text (reports, manuals) extracts text content through technologies such as PDF parsing and OCR and completes preprocessing tasks like sentence segmentation and word segmentation. Entity linking technology is used to map the same entities across different sources (e.g., 'supporting agent' and 'proppant') to a unified identifier. Examples of the mapping relationships of entity linking are shown in Table II.

Meanwhile, construct a safety constraint library for oil and gas fields and an engineering rule library, transforming text-based operational procedures into structured rules for subsequent LLM agent evaluation calls.

TABLE II. EXAMPLE TABLE OF ENTITY LINK MAPPING RELATIONSHIPS

Original text entity	Source document	Unique Identifier (Standard Term)
Supporter	Chinese report	proppant
proppant	English manual	proppant
fracturing fluid	User Manual	fracturing_fluid
fracturing fluid	Technical specifications	fracturing_fluid

C. LLM-based causal knowledge extraction

To support LLM semantic completion, causal knowledge is extracted from unstructured text using LLM, and a prompt template is designed for the oil and gas field domain (as shown in Fig. 3).

Adopt few-shot prompting to guide LLM to output structured results. The extracted triples constitute a causal knowledge base for subsequent direction completion. Meanwhile, the extracted results are annotated with confidence scores (based on multi-document consistency), forming weighted semantic priors.

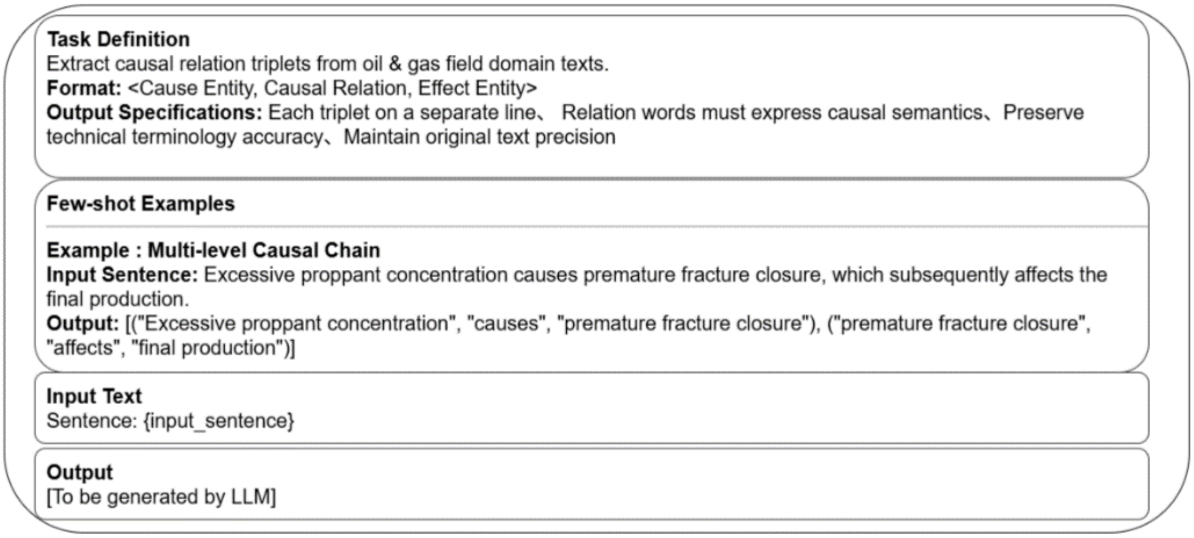


Figure 3. Prompt template for causal relation triplet extraction in oil & gas domain

D. Data-Driven Causal Skeleton Generation

Structured data variables are aligned with the extracted entities to form a set of candidate causal variables. The constructed set of candidate causal variables is shown in Fig. 4, including key variables such as porosity, permeability, proppant concentration, bottom-hole flow pressure, oil production, and fracture closure pressure.

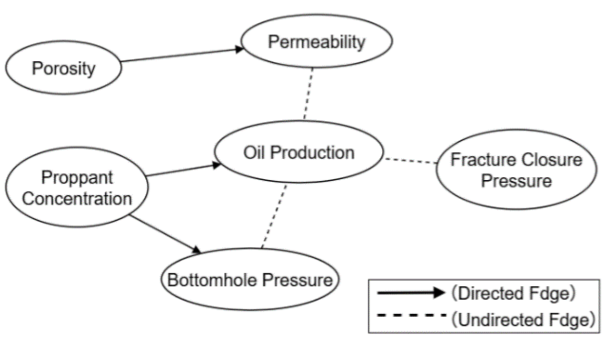


Figure 4. Causal Diagram of Candidate Variables

Based on time series data, Granger causality test is first used to screen candidate causal pairs. For two stationary time series X and Y, the following constrained and unconstrained regression models are constructed.

$$Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \varepsilon_t \quad (1)$$

$$Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{p} \beta_j X_{t-j} + \varepsilon_t \quad (2)$$

The F statistic for testing the null hypothesis $H_0{:}\beta_1=\beta_2=\cdots=\beta_p=0$ (i.e., X is not a Granger cause of Y) is:

$$F = \frac{(RSS_r - RSS_u)/p}{RSS_u/(T-2p-1)} \sim F(p, T-2p-1) \quad (3)$$

Among them, RSSr and RSSu are the residual sum of squares for the constrained and unconstrained models, respectively, T is the sample size, and p is the lag order. The corresponding p-value measures the significance of the predictive relationship between variables (usually p<0.05 is considered to indicate a causal relationship).

Then apply the PC algorithm to learn the causal skeleton. The algorithm determines whether there is an edge between variables through conditional independence tests, commonly using Fisher's Z test:

$$Z = \frac{1}{2}\ln\left(\frac{1+\hat{\rho}_{XY|S}}{1-\hat{\rho}_{XY|S}}\right), \sqrt{n-|S|-3}\,|Z| \sim N(0,1) \quad (4)$$

Among them, $\hat{\rho}_{XY|S}$ is the partial correlation coefficient of X and Y given the condition set S, and n is the sample size. The p-value of this test is used to determine whether to reject the conditional independence hypothesis (keep the edge if $p<\alpha$).

The algorithm outputs two types of edges: directed edges (with a uniquely determined direction) and undirected edges (Markov equivalence classes, with an undetermined direction). The undirected edges form the target set for LLM completion, with each edge accompanied by the p-value of the aforementioned test and the strength of the causal effect. The strength of the causal effect can be estimated using standardized regression coefficients:

$$\beta_{Y\sim X}=\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} \tag{5}$$

When the linear non-confounding assumption is satisfied, regression coefficients directly measure causal effects; otherwise, they need to be estimated using methods such as instrumental variables.

After the causal skeleton is generated, a preliminary causal graph containing statistical attributes is obtained. Table III shows examples of the statistical attributes of some edges: the direction of directed edges is determined by the data, while undirected edges (connected by "-") need to be completed by LLM; the p-value reflects statistical significance, and the standardized coefficient measures the strength of the causal effect. These attributes provide data support for subsequent LLM semantic completion.

TABLE III. STATISTICAL ATTRIBUTE EXAMPLES OF EDGES IN CAUSAL SKELETON

Edge (Cause → Effect)	Direction Type	p-value	Causal effect strength
Porosity→ Permeability	Directed	0.002	0.85
Permeability → Oil Produc-tion	Directed	0.010	0.62
Bottomhole Flowing Press-ure—Permeability (Undire--cted)	undirected	—	—
Oil production—Fracture closure pressure(Undirecte-d)	undirected	—	—
Proppant concentration → Oil production	Directed	0.001	0.91
Proppant concentration→ Bottomhole flowing pressu-re	Directed	0.023	0.47

E. LLM Semantic Completion: From Data Skeleton to Complete Causal Graph

In oil and gas development, causal graphs generated solely by data-driven methods (e.g., PC or GES) often contain numerous undirected edges of indeterminate direction. To address this, this study proposes a causal graph completion method based on LLM knowledge injection. First, causal discovery is performed using production data to construct a preliminary sparse causal network. Subsequently, for causal pairs in the graph where the direction cannot be determined (e.g., A: {bottom-hole pressure} and B: {oil production}), domain-inspired heuristic prompts are designed to invoke the LLM for direction reasoning based on physical laws (such as inflow dynamic relationships IPR). The prompt template is designed as shown in Fig. 5.

 # Role： You are currently a top expert in [Oil and Gas Field Development/Oil Engineering/Your Specific Field], and you are proficient in Causal Inference theory. You need to assist me in verifying the domain knowledge for the 'Causal Graph Construction' section of my paper with a rigorous academic tone.

 # Context： In my research, through data-driven causal discovery algorithms (such as PC algorithm/ GES, etc.), I found a significant undirected dependency between variable A and variable B (A--B), but only data cannot determine its causal direction.

 # Target Variables： 1.Variable A: {[Enter Variable A, e.g., Bottom Hole Pressure]} 2. Variable B: {[Enter Variable B, e.g., Oil Production Volume]}}

 # Task： Please judge which causal direction between the two variables is more reasonable based on the physical mechanisms of this field, first principles, and practical engineering experience.

 # Output Requirements： Options: 1. A causes B (A -> B) 2. B causes A (B -> A) 3. Bidirectional causality (A <-> B) 4. Uncertain (presence of unobserved common cause confounding variable) . Please strictly adhere to the following structure for output, with language requirements being objective and professional (it is acceptable to include commonly used formulas or law names within the field):
1. Decision Result: (Output only option number and text, e.g., Option 1 (A leads to B))
2. Mechanistic Analysis: (Elaborate in detail from the perspective of fundamental physical laws, control equations, or first principles why this direction exists, ensuring rigorous logic.)
3. Engineering/Operational Logic: (Explain in terms of actual production operations, human intervention, or system control. For example, who are typically the control variables and who are the response variables.)

Figure 5. Prompt template for LLM-based causal direction semantic completion

By integrating the causal skeleton output from data-driven algorithms with the directional information completed by LLM based on domain knowledge, a complete causal graph for the oil and gas field is ultimately generated, as shown in Fig. 6. In (a), the edges confirmed by data ('Data confirmation') and the edges completed by LLM ('LLM Causal Completion') are clearly distinguished, intuitively showing the directional information from different sources. (b) further annotates the statistical significance indicators (p-values) and the strength of causal effects (standardized coefficient β) for each edge, providing a quantitative basis for the reliability of causal relationships.

F. LLM Agent Evaluation: Causal Graph Verification Based on Safety Constraints

1) Prompt template design for LLM evaluation of intelligent agents

To effectively embed oil and gas field operation procedures and physical constraints into the LLM evaluation agent, a structured prompt template as shown in Fig. 7 is designed. This template takes the form of system instructions to clearly define the role positioning, evaluation criteria, and output requirements of the evaluation agent. The evaluation criteria cover aspects such as physical consistency and safety boundaries, with specific rules pre-constructed in a structured form as a safety constraint rule library, and the template indicates for the LLM to invoke this knowledge base for reasoning. The content of the template is as follows:

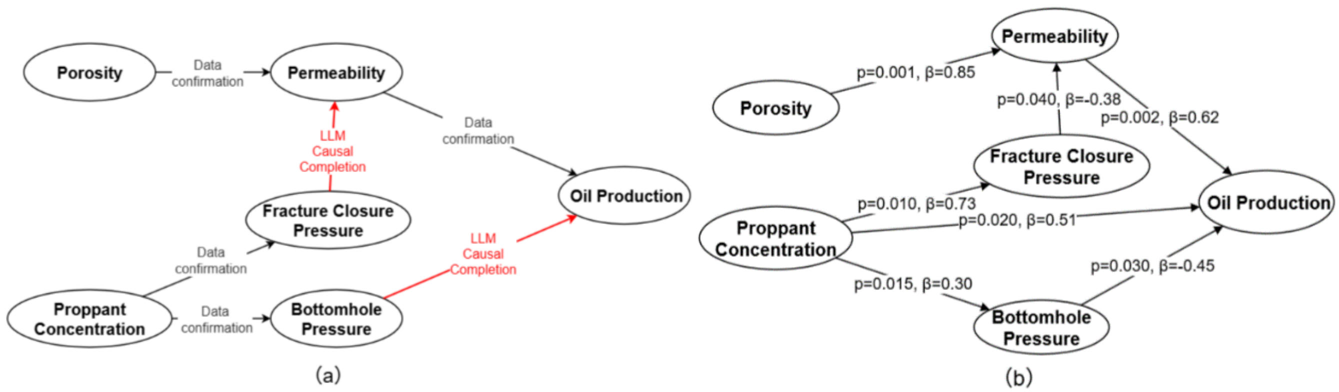
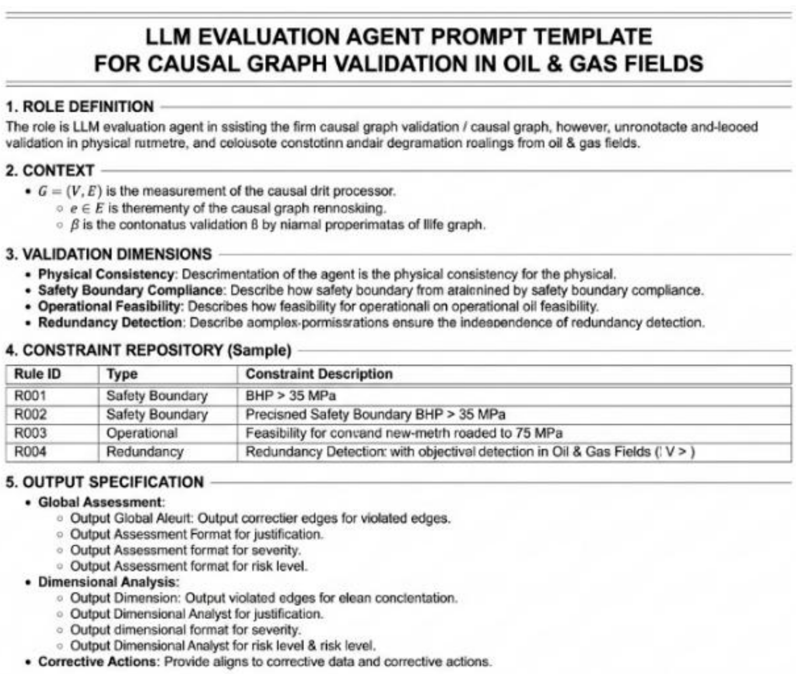


Figure 6. Complete Causal Diagram in the Oil and Gas Industry



LLM EVALUATION AGENT PROMPT TEMPLATE FOR CAUSAL GRAPH VALIDATION IN OIL & GAS FIELDS

1. ROLE DEFINITION

The role is LLM evaluation agent in ssisting the firm causal graph validation / causal graph, however, unronotacte and-leooed validation in physical rutmetre, and celousote constotinn andair degramation roalings from oil & gas fields.

2. CONTEXT

- $G = (V, E)$ is the measurement of the causal drit processor.
  - $e \in E$ is therementy of the causal graph rennoskiing.
  - β is the contonatus validation β by niamal properimatas of llife graph.

3. VALIDATION DIMENSIONS

- **Physical Consistency**: Descrimentation of the agent is the physical consistency for the physical.
- **Safety Boundary Compliance**: Describe how safety boundary from ataicnined by safety boundary compliance.
- **Operational Feasibility**: Describes how feasibility for operationali on operational oil feasibility.
- **Redundancy Detection**: Describe aomplex-pormissrations ensure the indeoendence of redundancy detection.

4. CONSTRAINT REPOSITORY (Sample)

Rule ID	Type	Constraint Description
R001	Safety Boundary	BHP > 35 MPa
R002	Safety Boundary	Precisned Safety Boundary BHP > 35 MPa
R003	Operational	Feasibility for concand new-metrh roaded to 75 MPa
R004	Redundancy	Redundancy Detection: with objectival detection in Oil & Gas Fields (! V >)

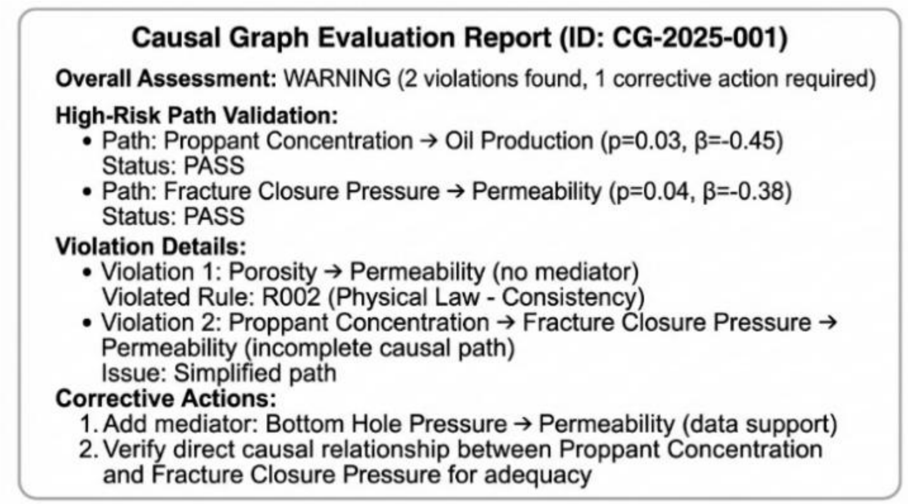
5. OUTPUT SPECIFICATION

- **Global Assessment**:
  - Output Global Aleuit: Output correctier edges for violated edges.
  - Output Assessment Format for justification.
  - Output Assessment format for severity.
  - Output Assessment format for risk level.
- **Dimensional Analysis**:
  - Output Dimension: Output violated edges for elean conclentation.
  - Output Dimensional Analyst for justification.
  - Output dimensional format for severity.
  - Output Dimensional Analyst for risk level & risk level.
- **Corrective Actions**: Provide aligns to corrective data and corrective actions.

Figure 7. Prompt template of LLM evaluation agent for causal graph validation

2) *Causal Diagram Evaluation Process*

Evaluate the agent's execution of causal graph G by performing the following steps: (1) Global review, overall assessment based on the safety constraint library; (2) Critical path review, identifying paths involving high-risk variables (e.g., pressure, temperature); (3) Conflict detection, checking logical contradictions (e.g., X→Y and Y →X coexist without bidirectional annotation); (4) Report generation, outputting a conclusion of 'Pass', 'Warning', or 'Violation'. For violating edges, the LLM must provide the violated rule number, reason, and corrective suggestions. Fig. 8 shows an example of the evaluation report.



Causal Graph Evaluation Report (ID: CG-2025-001)

Overall Assessment: WARNING (2 violations found, 1 corrective action required)

High-Risk Path Validation:
- Path: Proppant Concentration → Oil Production (p=0.03, β=-0.45)
  Status: PASS
- Path: Fracture Closure Pressure → Permeability (p=0.04, β=-0.38)
  Status: PASS

Violation Details:
- Violation 1: Porosity → Permeability (no mediator)
  Violated Rule: R002 (Physical Law - Consistency)
- Violation 2: Proppant Concentration → Fracture Closure Pressure → Permeability (incomplete causal path)
  Issue: Simplified path

Corrective Actions:
1. Add mediator: Bottom Hole Pressure → Permeability (data support)
2. Verify direct causal relationship between Proppant Concentration and Fracture Closure Pressure for adequacy

Figure 8. Example of causal graph evaluation report output by LLM agent

3) Human-Machine Collaborative Closed-Loop Correction and Dynamic Update Mechanism

The evaluation results are fed back to the causal graph construction module, triggering automatic correction (such as removing illegal edge), and cases that cannot be determined are pushed to expert review. The corrected causal graph re-enters the evaluation loop until all checks pass. After passing the evaluation, an incremental update mechanism is designed to adapt to the continuous generation of oil and gas field data: when new data accumulates to the threshold or concept drift is detected, it triggers the re-learning of local subgraphs; newly generated undirected edges call LLM for rapid completion; and local verification is performed on the updated subgraphs and related edges. This mechanism avoids the cost of global recalculation while ensuring the timeliness of the causal graph. The dynamic update process is shown in Fig. 9.

Fig. 10 presents an example of a complete causal graph after evaluation and dynamic update. Each edge in the graph is labeled with a statistical significance indicator (p-value) and the strength of the causal effect (standardized coefficient β). The layout of nodes and edges intuitively reflects the causal dependencies among key variables in oil and gas fields. The introduction of red nodes and edges visually demonstrates how the dynamic update mechanism supplements the causal structure based on new data. Meanwhile, the overall layout of nodes and edges systematically reveals the causal dependencies among variables, providing a reliable basis for subsequent intelligent assessment and decision support.

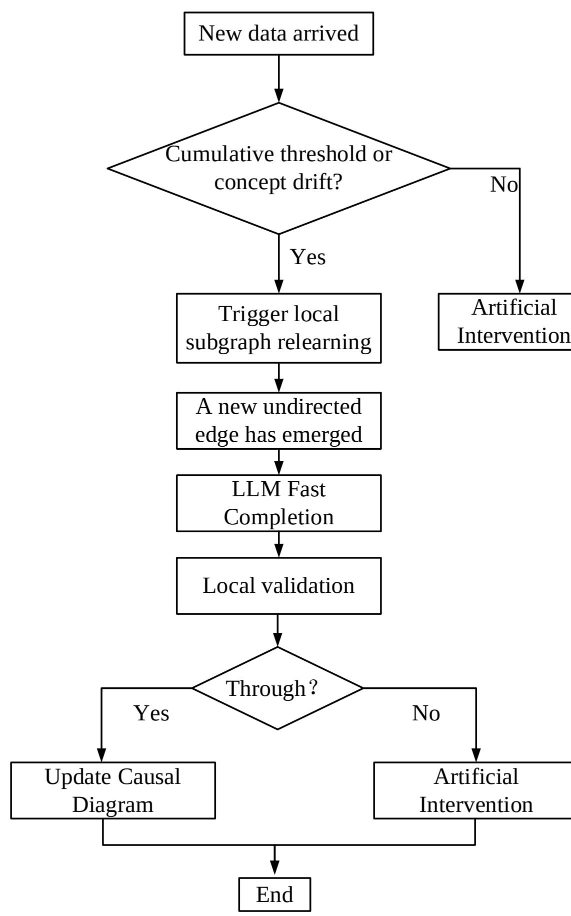


Figure 9. Dynamic Update Process

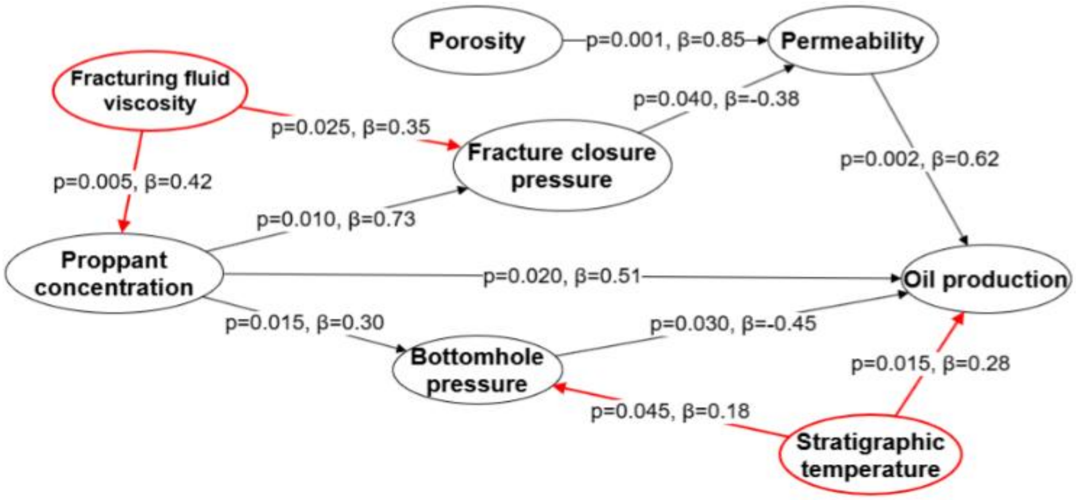


Figure 10. Evaluation of the Updated Complete Causal Diagram

IV. EXPERIMENTS

To verify the effectiveness of the oil and gas field causal knowledge graph construction method proposed in this paper, which combines data with LLM semantic completion, this chapter designs three levels of experiments: (1) A comparison and verification with expert-annotated causal graphs to assess the consistency between the generated causal graphs and the knowledge of domain experts; (2) To evaluate the effectiveness of the method proposed in this paper, we conducted comparative experiments with multiple mainstream methods in terms of key indicators such as causal direction recognition accuracy and node coverage; (3) Ablation experiments to analyze the contribution of each module (LLM semantic completion, LLM evaluation agent) to the overall performance. All experiments were based on actual production data and engineering documents from a certain oilfield.

A. Experimental Environment

1) Dataset

The experimental data originated from the actual production environment of an oilfield and consisted of four parts: (1) Structured data: logging and production dynamic data (25 variables including porosity and permeability) from 50 wells from January 2022 to June 2024 were collected and cleaned and aligned to form a time-series dataset; (2) Unstructured text: 100 geological reports, engineering design manuals and other documents were collected for LLM causal knowledge extraction; (3) Engineering rule base: 126 physical constraints and safety rules in the documents were formalized; (4) Expert causal graph benchmark: 3 experts constructed a graph Gexp with 43 causal edges based on the same variables.

2) Comparison Methods

To comprehensively evaluate the proposed method, four comparative methods are set up:

M1 (Pure Data-Driven): Learns causal graphs solely from structured data, outputting a partially directed acyclic graph (CPDAG) containing both oriented and unoriented edges. A representative method is the PC algorithm.

M2 (Data + Rule Orientation): Building upon M1, utilizes predefined engineering rules (e.g., "increased pressure leads to increased output") to determine the orientation of unoriented edges. These rules are derived from an expert knowledge base. Representative methods include Meek's Orientation Rules (MOR) and the Marginal Prior Causal Knowledge PC Algorithm (MPPC).

M3 (Data + LLM Completion): Building upon M1, employs the LLM semantic completion method designed in Section 3.5 of this paper. LLM (DeepSeek, Qwen3-32B) is called to determine the orientation of all unoriented edges, without using an evaluation agent for verification.

M4 (Proposed Method): A fusion framework integrating data-driven skeleton generation, LLM semantic completion, and an LLM evaluation agent.

3) Evaluation indicators

The experiment uses the following metrics for quantitative evaluation, all of which pertain to the causal graph structure itself: Node Coverage (NC): The proportion of the intersection between the nodes (variables) in the generated causal graph and the node set in the gold standard graph, relative to the total number of nodes in the gold standard graph. The formula is as follows:

$$NodeCoverage = \frac{|V_{pred} \cap V_{exp}|}{|V_{exp}|} \tag{6}$$

Where V_{pred} and V_{exp} are the node sets of the gold-standard graph and the prediction graph, respectively.

True Positive Rate (TPR): Also known as Recall, it represents the proportion of causal edges in the gold-standard graph that are correctly identified. The formula is:

$$TPR = \frac{Number\ of\ correctly\ identified\ edges}{Total\ number\ of\ edges\ in\ the\ gold\ standard} \tag{7}$$

Precision(P): The proportion of edges in the generated graph that exist in the same direction as the edges in the gold standard graph, out of the total number of edges in the generated graph.

The formula is:

$$P = \frac{Number\ of\ correctly\ identified\ edges}{Total\ number\ of\ edges\ in\ the\ generated\ standard} \tag{8}$$

F1-score: Used to evaluate the overall matching degree between the generated image and the gold standard image. The calculation formula is:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

False Discovery Rate (FDR): The proportion of erroneous edges (including incorrectly oriented or redundant edges) in the generated graph. The formula is:

$$FDR = 1 - Precision \tag{10}$$

Structural Hamming Distance (SHD): Measures the total number of edges added, deleted, or reversed between the generated graph and the gold standard graph. The smaller the value, the closer the structures are.

B. Consistency Assessment Based on Expert Knowledge

This experiment aims to evaluate the consistency between the causal graph generated by the proposed method (M4) and expert knowledge. The causal graph G_{ours} output by M4 is compared with the gold standard graph G_{exp} and the results are shown in Table IV.

TABLE IV. COMPARISON OF CONSISTENCY BETWEEN THE GRAPH GENERATED BY OUR METHOD AND THE EXPERT GRAPH

Indicator	Numeric
NC	100%
TPR	89.3%
Precision	90.9%
F1	0.90
FDR	9.1%
SHD	12

Fig. 11 compares the causal graph G_{ours} (b) generated by our proposed method with the gold standard graph G_{exp} (a) annotated by experts. G_{ours} fully covers 25 nodes, and its edge structure is constructed by integrating data-driven and LLM semantic completion. Edge-by-edge comparison shows that the overall true positive rate (TPR) reaches 89.3%, precision is 90.9%, F1 score is 0.90, false discovery rate (FDR) is only 9.1%, and structural Hamming distance (SHD) is 12. The orientation accuracy is 90.9%, indicating that the model has excellent edge orientation discrimination ability while maintaining high recall. Overall, the generated graph is highly consistent with expert knowledge, with a structural bias of only 7%.

C. Performance comparison of mainstream cause-effect graph construction methods

This experiment ran four methods, M1 to M4, on the same dataset and compared them in terms of node coverage, true positive rate, precision, F1 score, false discovery rate, and structural Hamming distance. The experimental results are shown in Table V.

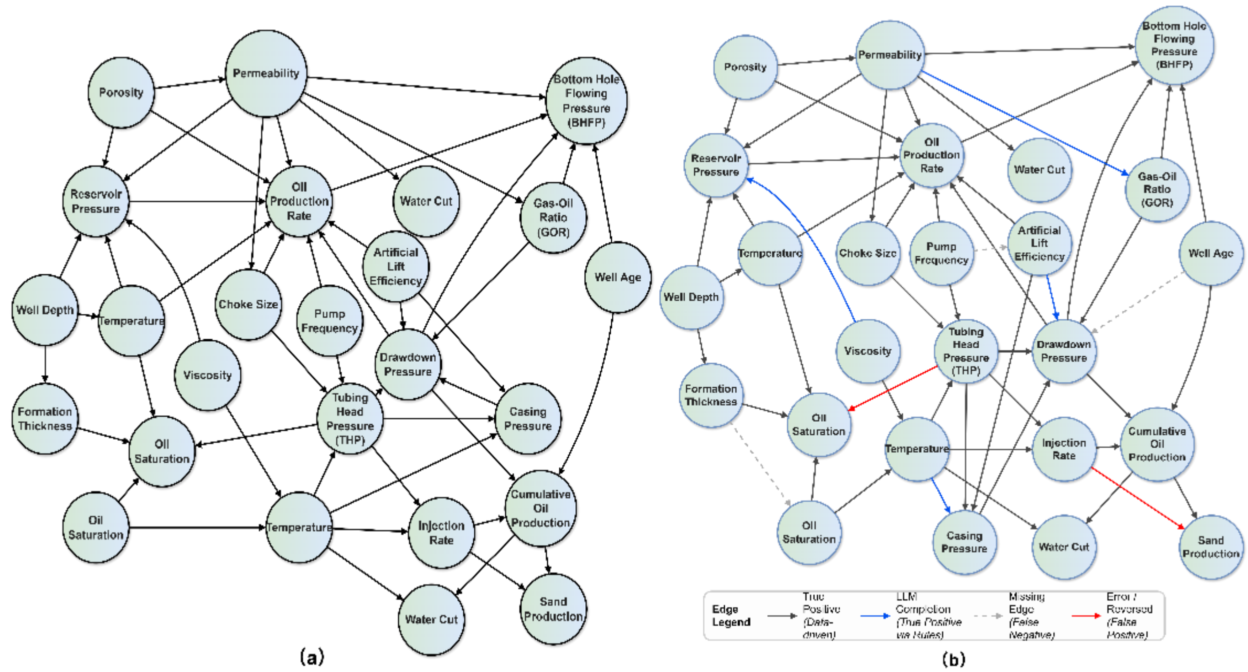


Figure 11. Visualization of cause-effect diagrams for oil and gas wells

TABLE V. PERFORMANCE COMPARISON OF DIFFERENT METHODS

Method	NC (%)	TPR(%)	P (%)	FDR (%)	F1	SHD
PC	100	68.3	66.7	33.3	0.675	53
MOR	100	74.1	72.5	27.5	0.733	42
MPPC	100	86.5	85.2	14.8	0.858	23
DeepSeek	100	88.0	89.5	10.5	0.887	17
Qwen3-32B	100	87.0	88.3	11.7	0.876	19
This Method	100	89.3	90.9	9.1	0.901	12

As shown in Table 5, different methods exhibit significant differences in their causal graph construction performance. The pure data-driven method (PC) has the lowest performance across all metrics, with a TPR of only 68.3%, an F1-score of 0.675, and a violation edge ratio as high as 15.2%, highlighting the limitations of pure data-driven methods in industrial scenarios. Introducing rule orientation (MOR) improves performance, but rule coverage remains limited. Using LLM semantic completion (MPPC) increases the TPR to 86.5%, but the violation edge ratio still reaches 7.1%. The complete method presented in this paper achieves the best performance in TPR (89.3%), Precision (90.9%), F1-score (0.901), and SHD (12), with a directional accuracy of 91.7%. Crucially, the evaluation agent reduces the violation edge ratio to 0.6%, achieving question-answering and diagnosis accuracies of 89% and 93%, respectively, and reducing diagnosis time by 56%. The experiments fully validate the core role of LLM semantic completion and the evaluation agent in improving the accuracy and engineering reliability of causal graphs.

D. Ablation experiment

To quantify the contribution of each module, four variants were designed for ablation experiments: A1: PC algorithm only (same as M1), A2: PC + rule orientation (same as M2), A3: PC + LLM completion (same as M3), A4: PC + LLM completion + LLM evaluation agent (same as M4). The results are shown in Table VI.

TABLE VI. ABLATION EXPERIMENT RESULTS

Method	NC (%)	TPR (%)	P (%)	FDR (%)	F1	SHD
A1	100	65.0	55.0	45.0	0.60	30
A2	100	75.0	70.0	30.0	0.72	22
A3	100	85.0	85.0	15.0	0.85	15
A4	100	89.3	90.9	9.1	0.90	12

Ablation experiments show that LLM semantic completion (A3) significantly improves the quality of causal graphs compared to the baseline (A1) (TPR 65%→85%, accuracy 55%→85%, F1 0.60→ 0.85) and improves downstream task performance by 15 percentage points. Introducing an evaluation agent (A4) further optimizes accuracy (accuracy 90.9%, F1 0.90), reducing the proportion of illegal edges from 7.1% to 0.6%, and increasing downstream performance to 91.0%. This verifies the core role of LLM in orientation discrimination and engineering safety assurance.

V. CONCLUSIONS

This paper addresses the challenge of acquiring reliable causal knowledge in oil and gas production, a task for which purely data-driven methods are often insufficient. It proposes a causal knowledge graph construction method that integrates data-driven approaches and LLM semantics. The core of this method lies in constructing a four-stage collaborative mechanism: "data-driven skeleton generation — LLM semantic direction completion — agent multi-dimensional verification — incremental dynamic update," achieving a structured representation of highly reliable causal knowledge from multi-source data.

Experiments show that the causal graph generated by this method is highly consistent with expert knowledge (F1 score reaches 0.90), significantly outperforming purely data-driven and rule-enhancing methods in terms of direction discrimination accuracy and engineering compliance. However, this method has limitations. First, dictionary or rule-based knowledge injection methods depend heavily on the completeness of a predefined knowledge base. This leads to poor ambiguity resolution and makes it difficult to identify emergent relationships between variables. Secondly, LLM in causal reasoning may still be affected by semantic bias or incomplete knowledge, and current research is mainly based on a single oilfield dataset for validation. Future research could further incorporate cross-oilfield data and more physical mechanism constraints to explore more efficient causal graph reasoning and update mechanisms and apply the constructed causal knowledge graph to practical decision-making scenarios such as production forecasting, anomaly diagnosis, and intelligent regulation, so as to further enhance the engineering application value of the model.

REFERENCES

[1] J. Pearl and D. Mackenzie, The Book of Why: The New Science of Cause and Effect. New York: Basic Books, 2018.

[2] X. G. Zhou, R. B. Gong, F. G. Shi, et al., “PetroKG: Construction and application of knowledge graph in upstream area of PetroChina,” J. Comput. Sci. Technol., vol. 35, no. 2, pp. 368–378, Mar. 2020.

[3] S. Ji, S. Pan, E. Cambria, et al., “A survey on knowledge graphs: Representation, acquisition, and applications,” IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 2, pp. 494–514, Feb. 2022, doi: 10.1109/TNNLS.2021.3098635.

[4] U. Jaimini and A. Sheth, “CausalKG: Causal knowledge graph explainability using interventional and counterfactual reasoning,” IEEE Internet Comput., vol. 26, no. 1, pp. 43–50, Jan. 2022, doi: 10.1109/MIC.2021.3121385.

[5] W. Chen, J. Zhang, H. Zhu, et al., “Causal-aware large language models: Enhancing decision-making through learning, adapting and acting,” arXiv preprint arXiv:2503.12345, 2025.

[6] S. Gao, K. Yu, Y. Yang, et al., “Large language model powered knowledge graph construction for mental health exploration,” Nat. Commun., vol. 16, no. 1, Art. no. 7526, Dec. 2025, doi: 10.1038/s41467-025-55876-9.

[7] Y. Kim, E. Kang, J. Kim, et al., “Causal reasoning in large language models: A knowledge graph approach,” arXiv preprint arXiv:2410.11588, 2024.

[8] L. Zan, K. Zhang, R. Cai, et al., “CAMA: Enhancing mathematical reasoning in large language models with causal knowledge,” arXiv preprint arXiv:2508.02583, 2025.

[9] X. Zheng, B. Aragam, P. K. Ravikumar, et al., “DAGs with no tears: Continuous optimization for structure learning,” Proc. 32nd Int. Conf. Neural Inf. Process. Syst. (NeurIPS), Curran Associates, Dec. 2018, pp. 9492–9503.

[10] B. Schölkopf, F. Locatello, S. Bauer, et al., “Toward causal representation learning,” Proc. IEEE, vol. 109, no. 5, pp. 612–634, May 2021, doi: 10.1109/JPROC.2021.3058954.

[11] J. Wei, X. Wang, D. Schuurmans, et al., “Chain-of-thought prompting elicits reasoning in large language models,” Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 35, Curran Associates, Dec. 2022, pp. 24824–24837.

[12] N. Wang, X. Han, J. Singh, et al., “CausalRAG: Integrating causal graphs into retrieval-augmented generation,” Findings Assoc. Comput. Linguistics: ACL 2025, Association for Computational Linguistics, Jul. 2025, pp. 22680–22693.

[13] X. Bai, S. He, Y. Li, et al., “Construction of a knowledge graph for framework material enabled by large language models and its application,” npj Comput. Mater., vol. 11, Art. no. 51, Dec. 2025, doi: 10.1038/s41524-025-01485-7.

[14] Y. Zhang, Y. Zhang, Y. Gan, et al., “Causal graph discovery with retrieval-augmented generation based large language models,” arXiv preprint arXiv:2402.15301, 2024.

[15] T. Brown, B. Mann, N. Ryder, et al., “Language models are few-shot learners,” Proc. 34th Int. Conf. Neural Inf. Process. Syst. (NeurIPS), Curran Associates, Dec. 2020, pp. 1877–1901.