

Article

Claim Knowledge Graph Construction and GraphRAG-Based Question-Answering System

Xinxue Wang * and Jun Fang

School of Civil Engineering and Architecture, Wuhan University of Technology, Wuhan 430062, China;
fangjun@whut.edu.cn

* Correspondence: wangxinxue@whut.edu.cn; Tel.: +86-15827011634

Abstract

Traditional claim management relies heavily on manual analysis and expert judgment, resulting in inefficiencies, information omissions, and heightened risks of disputes. To address these challenges, this paper constructs a domain-specific ontology for construction engineering claims through a five-step process, organizing the relevant knowledge into five unified core classes. Based on this ontology, a knowledge graph is built and stored in Neo4j. The resulting knowledge graph-enhanced LLM question-answering system, evaluated using BLEU-4, BERT-Cosine similarity, ROUGE-1, and ROUGE-L metrics, demonstrates superior performance compared to both the base LLM and Vector RAG approaches. The results indicate that the proposed ontology effectively serves the purpose of knowledge sharing and reuse while providing practical support for construction claim management.

Keywords: construction claim management; knowledge graph; ontology; large language model; retrieval-augmented generation

1. Introduction

Construction claims play a pivotal role in the construction industry, exerting significant influence on a project's overall performance, success, or failure [1]. These claims directly affect key project outcomes, including schedules, costs, and profitability. When not managed effectively, they not only erode financial returns but also amplify risks for all stakeholders involved, potentially leading to substantial delays, cost overruns, and protracted disputes.

The causes of construction claims are multifaceted and diverse, frequently stemming from ambiguous contract clauses, frequent design changes, scope creep, unforeseen site conditions, delayed payments, and disruptions beyond the parties' control [2]. From the owner's perspective, ineffective claim management can substantially undermine overall project profitability through escalated indirect costs, prolonged disputes, and diminished returns. Conversely, from the contractor's viewpoint, robust claim management is essential for safeguarding legitimate rights, securing fair economic compensation, and promoting standardized, transparent operations that foster long-term industry credibility.

Traditional claim management approaches rely heavily on manual document review, expert judgment, systematic documentation, negotiation, and process optimization to identify entitlements, substantiate claims, and resolve disputes [3,4]. While these methods have served the industry for decades, they are increasingly inadequate in the face of growing project complexity, voluminous unstructured data, and the demand for timely, evidence-based decision making. This reliance often leads to information omissions, insufficient



Academic Editor: Pramen P. Shrestha

Received: 6 January 2026

Revised: 9 February 2026

Accepted: 17 February 2026

Published: 19 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

evidentiary support, subjective biases, and low processing efficiency, thereby exacerbating risks of delays, cost overruns, and protracted litigation.

In parallel, the construction sector is undergoing a profound digital transformation, driven by imperatives of sustainability, intelligence, and operational efficiency. Emerging technologies—such as artificial intelligence (AI), digital twins, big data analytics, and knowledge graphs—are increasingly recognized as critical enablers of this upgrade [1,5]. Among these, knowledge graph-based techniques stand out for their ability to structure domain knowledge, capture complex interrelationships, and support precise retrieval in knowledge-intensive tasks.

To address the aforementioned limitations, this paper proposes the construction of a domain-specific ontology for construction engineering claims, encompassing key entities such as parties, contractual provisions, claim events, evidentiary materials, and legal bases. Leveraging this ontology, knowledge is extracted from relevant textual sources to populate a knowledge graph (KG). The resulting KG is then integrated with large language models (LLMs) through Retrieval-Augmented Generation (RAG), creating an intelligent question-answering system that grounds responses in structured, verifiable domain knowledge. This approach aims to enhance factual accuracy, reduce hallucinations, improve traceability, and facilitate efficient knowledge management, thereby advancing intelligent practices in construction claim handling.

The remainder of this paper is organized as follows: Section 2 reviews related work in construction claim management, ontology engineering, knowledge graphs, and Retrieval-Augmented Generation; Section 3 details the research framework and methodology; Section 4 presents the experimental setup and results; Section 5 discusses the findings, implications, and limitations; and Section 6 concludes this paper and outlines directions for future research.

2. The Literature Review

2.1. Claim Management

A substantial body of research has investigated claim management in the construction industry, with a primary focus on identifying the types, causes, frequency, severity, and impacts of claims across various project contexts and regions.

For instance, ref. [6] quantitatively examines the relationship between rework causes and contractual claims, categorizing rework factors into process-related, management and planning, material and equipment, and human resource dimensions. Ref. [7] evaluates the causes, frequency, and severity of claims in Indian highway projects, distilling 22 key causes and proposing mitigation strategies. Ref. [8] employs machine learning methods to quantify the effects of three major claim categories—Delay, Extra Work, and Differing Site Conditions—on overall project performance.

Studies specific to certain regions include [9], which identifies variations, contractor delays, and inadequate site investigations as the most frequent causes in UAE road projects. Ref. [10] analyzes root causes and interaction mechanisms of contractors' claims in international EPC projects from the perspective of Chinese contractors, using industry surveys, structural equation modeling, and case studies. Ref. [11] systematically investigates how claim events contribute to project delays, quantifying the influence of key causes through factor analysis and structural equation modeling. Ref. [12] uses questionnaire surveys among UAE construction practitioners to analyze claim types, causes, frequency, and severity, offering preventive recommendations. Ref. [13] applies complex network theory and interpretive structural modeling to map cause networks, critical paths, and hierarchical structures in hydropower claims, aiming to improve management effectiveness.

Additional regional studies encompass [14], which quantitatively examines major issues and root causes faced by contractors in Thai public building projects via surveys; ref. [15], which analyzes archival data from completed projects to explore claim types, frequency, magnitude, and their relation to original contract amounts; and [16], which surveys contractors in the Gaza Strip to assess claim causes and their relative importance, providing management recommendations from the contractors' viewpoint.

Despite these advancements, the majority of existing research concentrates on the types and causes of claim events, while other critical elements of claim management—such as evidentiary materials, legal bases—remain underexplored. Consequently, there is a lack of a comprehensive knowledge framework that holistically describes the full spectrum of construction claim management processes.

To address this gap, this paper constructs a domain-specific ontology for construction claims, providing a more comprehensive and structured representation of claim management. This ontology serves as the foundation for subsequent knowledge extraction, knowledge graph construction, and intelligent applications in the field.

2.2. Retrieval-Augmented Generation with Knowledge Graph

With the latest advancements in LLMs, their synergistic integration with KGs has garnered increasing attention. The structured advantages of KGs provide robust support for multi-hop reasoning [17,18] and complex logical queries, effectively addressing the shortcomings of traditional RAG approaches that rely solely on vector similarity retrieval. This integration significantly enhances LLMs' reasoning capabilities [19–21] and global contextual understanding [22]. Moreover, by explicitly storing entities and relationships, KGs confer greater interpretability [23] and factual accuracy [24] to the system. Compared to conventional databases, KGs also offer superior capabilities for dynamic knowledge updating, as exemplified by dynamic data-driven methods for railway bridge construction KG updates. KG-enhanced LLMs are primarily applied in question-answering systems [25–27] and recommendation systems [28–31], demonstrating substantial value across various industries. For example, ref. [32] introduced ChatDiet, an LLM-powered framework for nutrition-oriented food recommendations, achieving a 92% effectiveness rate in food recommendation tests. Ref. [33] exploited the structured knowledge grounding benefits of KGs to mitigate hallucinations in LLM-based sentiment extraction and reduce noise from heterogeneous information fusion, thereby improving recommendation accuracy and diversity. Ref. [34] constructed a pan-cancer KG to develop a pan-cancer question-answering benchmark, promoting the discovery of novel drug repurposing opportunities through potential drug–cancer associations while minimizing factual errors in reasoning. Ref. [35] built a KG from historical customer service tickets, leveraging the graph's structural preservation and relational modeling strengths to improve retrieval precision, preserve intra- and inter-issue structures and relationships, as well as to alleviate information loss caused by text chunking, ultimately enhancing the efficiency and quality of customer service question answering.

In construction domain, KGs have been extensively applied by researchers in sub-domains such as safety management and contract management. Ref. [36] proposed CONSTRAG, an LLM-driven multi-agent framework that integrates LLMs with KG reasoning to generate personalized safety training materials for construction workers. Ref. [37] integrated LLMs with a Nested Contract Knowledge Graph (NCKG) through a Graph Retrieval-Augmented Generation (GraphRAG) framework, thereby improving retrieval and reasoning performance in contract review tasks. Ref. [38] developed a KG-enhanced LLM framework for question answering in hydraulic structure safety management.

Numerous studies have explored effective strategies for integrating knowledge graphs with large language models to enhance their reasoning and generation capabilities. These efforts have validated the effectiveness of GraphRAG and demonstrated its advantages in improving factual accuracy, reducing hallucinations, and supporting complex multi-hop reasoning. Collectively, they offer valuable insights into knowledge retrieval mechanisms, graph representation learning, and structured reasoning over knowledge graphs, thereby providing a solid theoretical and methodological foundation for the present work.

Building on these advancements, this study proposes a RAG framework based on a domain-specific knowledge graph for construction claims, applied to large language models, with the aim of advancing the intelligent management of construction claims.

2.3. Research Gaps

Despite these valuable contributions, several critical gaps remain in the current body of knowledge:

1. Lack of holistic integration of domain knowledge

The majority of studies concentrate narrowly on the causes of claims, while other essential elements of claim management are rarely addressed in a systematic and integrated manner. There is an absence of a comprehensive knowledge framework that encompasses all relevant aspects of the entire claim management lifecycle.

2. Lack of efficient and accurate decision-support methods, traditional claim

Management still relies heavily on manual document analysis, expert judgment, and experience-based reasoning. This approach is time-consuming, labor-intensive, prone to information omission, and limited in processing efficiency, especially when dealing with large volumes of unstructured textual data. Consequently, it struggles to provide timely, accurate, and consistent support for complex decision making in claim identification, substantiation, and resolution.

3. Risk of overlooking uncommon or complex claim scenarios

Due to the heavy dependence on practitioners' personal experience and manual processes, less frequent or highly context-specific claim scenarios are easily overlooked or incompletely analyzed. Without a structured, exhaustive knowledge representation, it is difficult to ensure comprehensive coverage of all possible claim events, legal provisions, and evidentiary requirements, which increases the risk of incomplete claims, inequitable outcomes, and prolonged disputes.

To bridge these gaps, this study proposes the construction of a domain-specific ontology for construction claims, followed by knowledge extraction to build a knowledge graph, and the integration of this graph with large language models through RAG. This approach aims to establish a comprehensive, structured, and intelligent knowledge management system that supports efficient, accurate, and holistic decision making in construction claim management.

3. Methodology

This section outlines the overall methodological framework adopted in this study to develop an intelligent question-answering system for construction claims. The proposed approach integrates domain-specific knowledge representation with modern RAG techniques, aiming to overcome the limitations of traditional manual claim management and conventional RAG methods. The overall research framework is illustrated in Figure 1, and the methodology comprises four main sequential steps:

Step 1: Domain data collection. A diverse corpus of construction claim-related textual data was collected to serve as the primary knowledge source. This includes: laws, regulations, and normative documents governing construction activities and claims; judg-

ments sourced from China Judgments Online and construction claim reports from actual construction projects; and academic research papers.

Step 2: Construction of the construction claim domain ontology. A formal domain ontology for construction claims was developed following the seven-step ontology development methodology. The ontology defines core classes, object properties, and datatype properties. This structured schema captures the essential concepts, relationships, and constraints inherent to claim management processes, providing a reusable and semantically consistent foundation for subsequent knowledge extraction and graph population.

Step 3: Knowledge graph construction. Utilizing the predefined ontology as a schema, a knowledge graph was automatically populated from the collected textual corpus. Large language models were employed to assist in entity recognition and relation extraction, identifying candidate instances of predefined classes and relations, with all extracted results subject to manual verification and correction by domain experts. The resulting KG was stored in Neo4j, enabling efficient traversal, multi-hop reasoning, and structured querying over complex claim-related knowledge.

Step 4: GraphRAG-based question-answering system and evaluation. The populated knowledge graph was integrated into a GraphRAG pipeline, where retrieval was performed via graph traversal rather than purely vector-based similarity search. Retrieved subgraphs or relevant triples were injected as context into the LLM prompt, enabling the generation of accurate, traceable, and domain-grounded answers to claim-related queries.

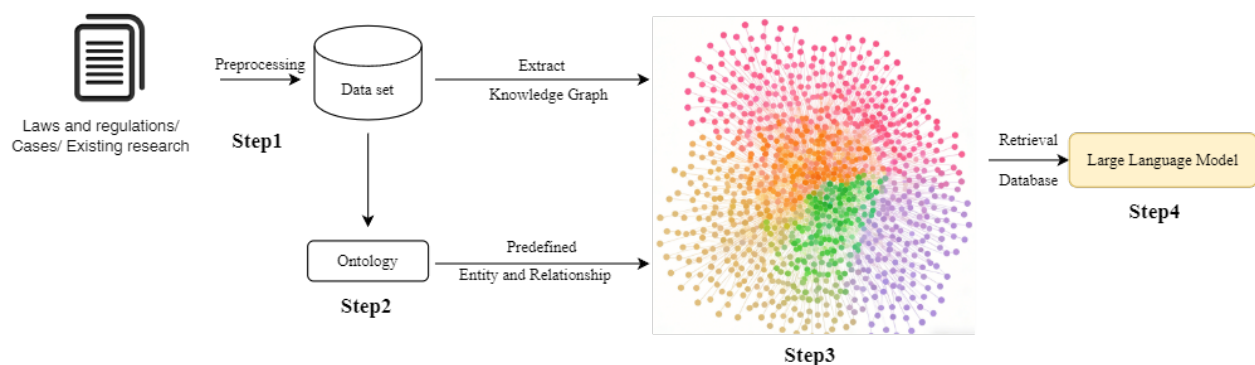


Figure 1. Flowchart of the methodology.

3.1. Ontology Design

This study employs the seven-step ontology development methodology to construct a domain ontology for construction claims, thereby achieving standardized knowledge representation and structured organization. Seven-step method is an ontology development method proposed by the Stanford University [39]. This method consists of seven explicit steps: 1. determine the domain and scope of the ontology; 2. consider the reuse of existing ontologies; 3. enumerate important terms in the ontology; 4. define the classes and the class hierarchy; 5. define the properties of classes—slots; 6. define the facts of the slots; and 7. create instances. To the best of our knowledge, there is no previous research on the ontology construction in construction claim domain, so we skipped step 2 in this research. A development method with five steps was summarized based on the seven-step-method, as shown in Figure 2.

Step 1: Determine the domain and scope of the ontology

The ontology constructed in this study defines classes and properties that are designed to be generally applicable to construction claims across all jurisdiction. However, the population of instances is limited to Chinese construction projects. This restriction arises because laws, regulations, normative standards, and contractual practices vary significantly across countries and regions, and the classification criteria for claim items—such as cost

breakdowns and valuation methods—are not fully consistent internationally. Therefore, this research is primarily conducted based on China’s prevailing management standards and practices and industry practices. Furthermore, with respect to the involved parties, the present study focuses exclusively on claims occurring between the employer and the general contractor.

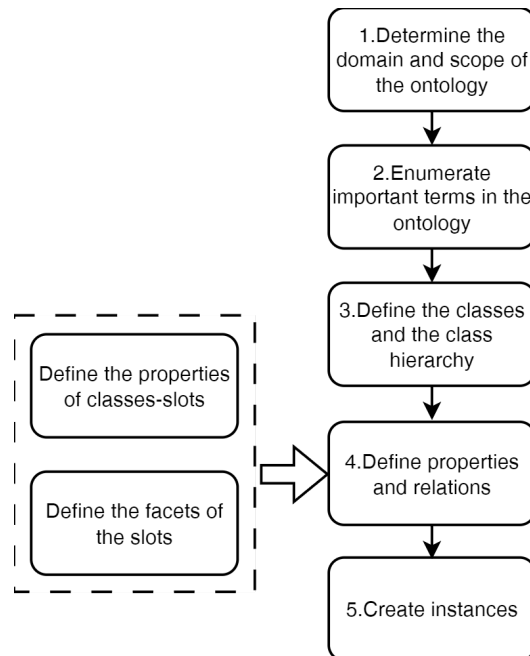


Figure 2. Overview of the development method with five steps.

Step 2: Enumerate important terms in the ontology

In this study, the primary knowledge sources for constructing the domain ontology are relevant laws and regulations, judicial cases, project-related claim documentation, and existing academic research. These four main categories are outlined as follows:

1. **Laws and regulations:** Laws and regulations specify the types of construction events that qualify for cost and schedule claims, identify the responsible party for each type of claim event, and provide the specific legal or contractual clause references as the basis for entitlement. In addition, certain clauses explicitly list the items that should be claimed. In this study, the following key Chinese legal and normative documents were referenced, with clauses related to construction claims selectively extracted and incorporated into the ontology: (a) Civil Code of the People’s Republic of China, (b) Construction Project Construction Contract (Model Text) (version GF-2017-0201), (c) Standard for bills of quantities and valuation for construction works (version GB/T 50500-2024), and (d) Interpretation (I) by the Supreme People’s Court on Issues Concerning the Application of Law in the Trial of Cases Involving Disputes over Construction Contracts for Construction Projects.

2. **Cases:** The case corpus comprises 31 judgments sourced from China Judgments Online and 13 construction claim reports from actual construction projects. From judgments sourced from China Judgments Online, adjudication rulings specifically pertaining to construction contract disputes between the employer and the contractor were selected. Typically, a single adjudication document addresses multiple issues, such as contract validity, determination of the actual constructor, and the starting date for interest accrual. Non-claim-related content and irrelevant information were subsequently removed. During the screening process, a strict distinction was maintained between construction claims and liquidated damages, with compensation for liquidated damages excluded from the scope of construction claims.

These materials provide concrete factual instances, judicial interpretations, evidentiary evaluations, legal applications, and practical outcomes, serving as critical sources for instantiating entities, relationships, and attributes in the knowledge graph.

3. Existing research

Existing research including [6–14] are referred to in this study. Existing research has extensively investigated the causes of construction claims and the causal relationships among these causes. Typically, ref. [13] summarizes 58 causal factors of construction claims and identifies 140 claim causation chains.

Step 3: Define the classes and the class hierarchy

Starting from the specific concepts identified in Step 2, a bottom-up abstraction process was adopted to progressively generalize them into higher-level, more universal parent classes. Throughout this process, consultations were conducted with professors specializing in the teaching of construction claims and construction management research, as well as with practitioners responsible for cost management from both owner and contractor organizations, to ensure the scientific validity, practicality, and domain appropriateness of the proposed class hierarchy. Through iterative refinement and expert feedback, the ontology was ultimately distilled into five primary top-level classes: Claim Event, Party, Claim, Evidence Material, and Contract/Law/Regulation. These core classes serve as the foundational structure for representing the essential elements and relationships involved in construction claim management.

Claim Event refers to any event occurring during the performance of a construction contract that results in a deviation between the actual circumstances and the contractual provisions, thereby causing changes to the project duration and/or costs. Common claim event types are shown in Figure 3. Relevant laws, regulations, and standard contract templates typically describe such events by assigning responsibility to one party for potential time or cost implications, or by delineating the obligations of the employer and the contractor that give rise to claim entitlements. For example, Article 805 of the Civil Code of the People's Republic of China provides that, if the employer alters the plan or supplies inaccurate data leading to rework in survey or design, additional costs shall be paid by the employer. In this clause, the employer's plan alteration or provision of inaccurate data constitutes a claim event, with the employer bearing liability for the resulting cost losses. When a causal relationship exists between claim events, they are linked via the "cause" relation—for instance, a design drawing change may cause project delay.

Contract/Law/Regulation denotes specific contractual or legal clauses that serve as the basis for a claim event or associated loss. The main laws and regulations in China regarding provisions for construction/engineering claims are shown in Figure 4. An illustrative example is Clause 7.5.1 of the General Terms in the Construction Project Construction Contract (Model Text) (version GF-2017-0201), which provides grounds for project delays attributable to the employer. In the construction industry in China, the primary laws, regulations, and normative documents related to construction claims mainly include Civil Code of the People's Republic of China, Construction Project Construction Contract (Model Text) (version GF-2017-0201), Standard for bills of quantities and valuation for construction works (version GB/T 50500-2024) and Interpretation (I) by the Supreme People's Court on Issues Concerning the Application of Law in the Trial of Cases Involving Disputes over Construction Contracts for Construction Projects.

The parties to a construction claim consist of the claimant and the respondent, as shown in Figure 5, which are typically the employer and the contractor. These entities are linked to the claim event via the relations "as claimant" and "as respondent", respectively.

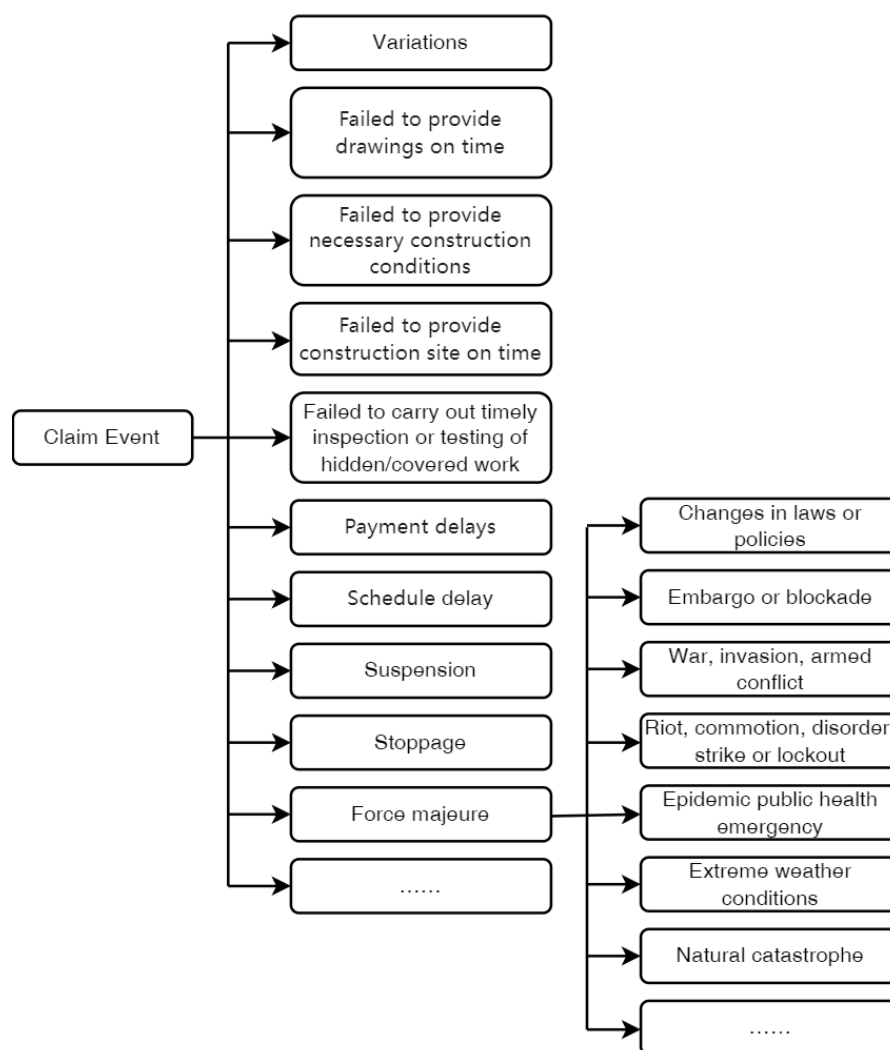


Figure 3. Excerpt of the claim event taxonomy.

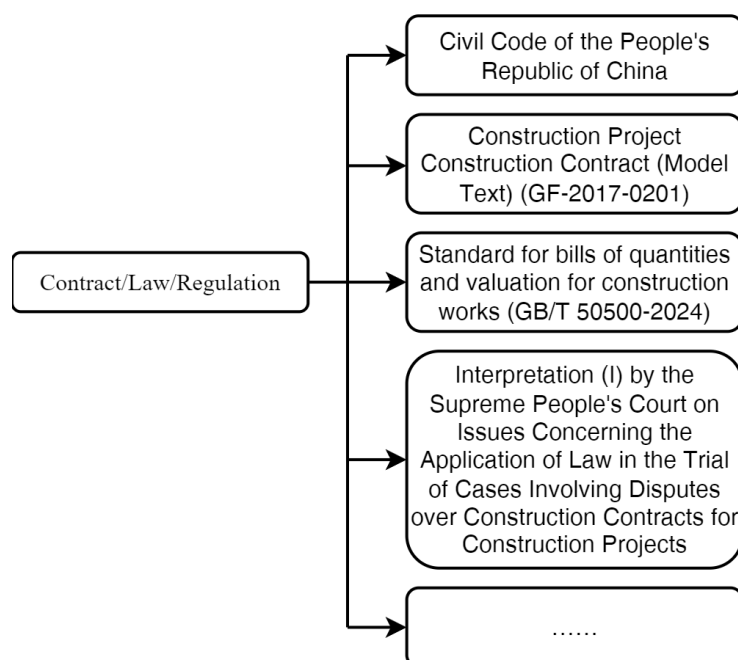


Figure 4. Excerpt of the contract/law/regulation taxonomy.

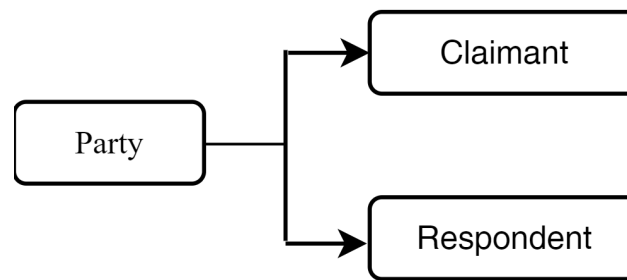


Figure 5. Party taxonomy.

Evidence Material encompasses all documents produced during project construction that can corroborate a claim event or loss, including meeting minutes, correspondence letters, and visa records from the construction process. The common types of evidence materials are shown in Figure 6. In adjudication documents of construction claim cases, evidence submitted by the claimant is not invariably accepted as valid; materials must satisfy specific conditions—such as complete signatures or seals, or explicit owner agreement—to be admissible. Based on judicial evaluations of validity in the cases, an “isValid” attribute is assigned to evidence nodes to indicate whether the material serves as effective proof. When evidence is invalid, an additional “reasonForInvalid” attribute is included to specify the grounds for rejection.

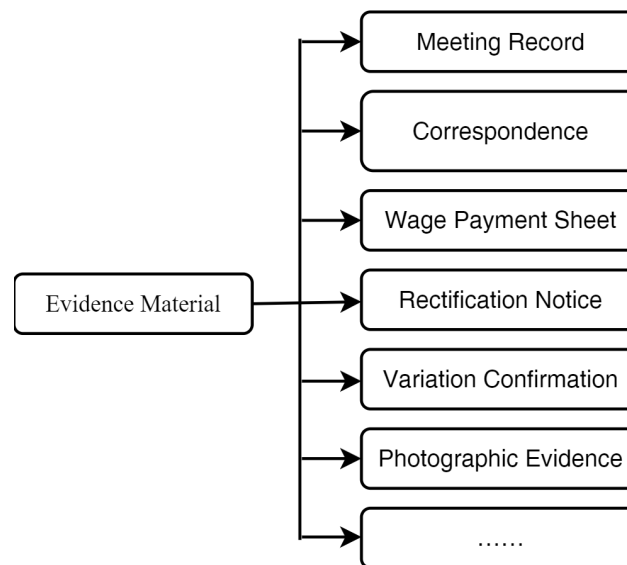


Figure 6. Excerpt of the evidence material taxonomy.

Claim primarily covers claims for extension of time and additional costs. Common types of claims are shown in Figure 7. In adjudication documents, losses claimed by the claiming party are ultimately determined as substantiated or unsubstantiated. Accordingly, an “isSubstantiated” attribute is added to loss nodes to specify whether the claim for that particular loss was upheld. If unsubstantiated, a “reasonForUnsubstantiated” attribute is included to explain the grounds for denial.

Step 4: Define properties and relations

The class taxonomies alone cannot provide sufficient information to fully describe a complete construction claim event. Therefore, the required additional details are captured as properties of the classes.

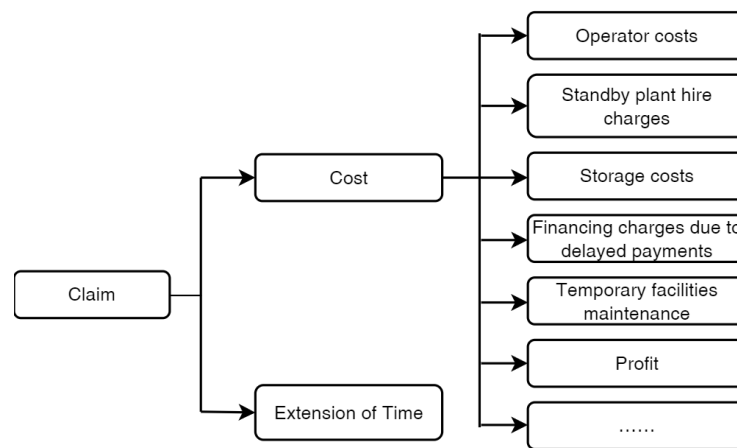


Figure 7. Excerpt of the claim taxonomy.

The properties of classes include Value Properties and Object Properties. Value Properties represent intrinsic characteristics or attribute values of the class itself. For example, in the judgment document of a construction claim event, a value property might indicate whether a particular evidence material is valid, along with the specific reasons for invalid. Object Properties define relationships between the current class and other classes in the domain. For example, a relationship such as ClaimEvent → hasClaimant → Party links a claim event to the party asserting the claim.

Figure 8 presents the key properties defined for the primary classes, together with their respective data types.

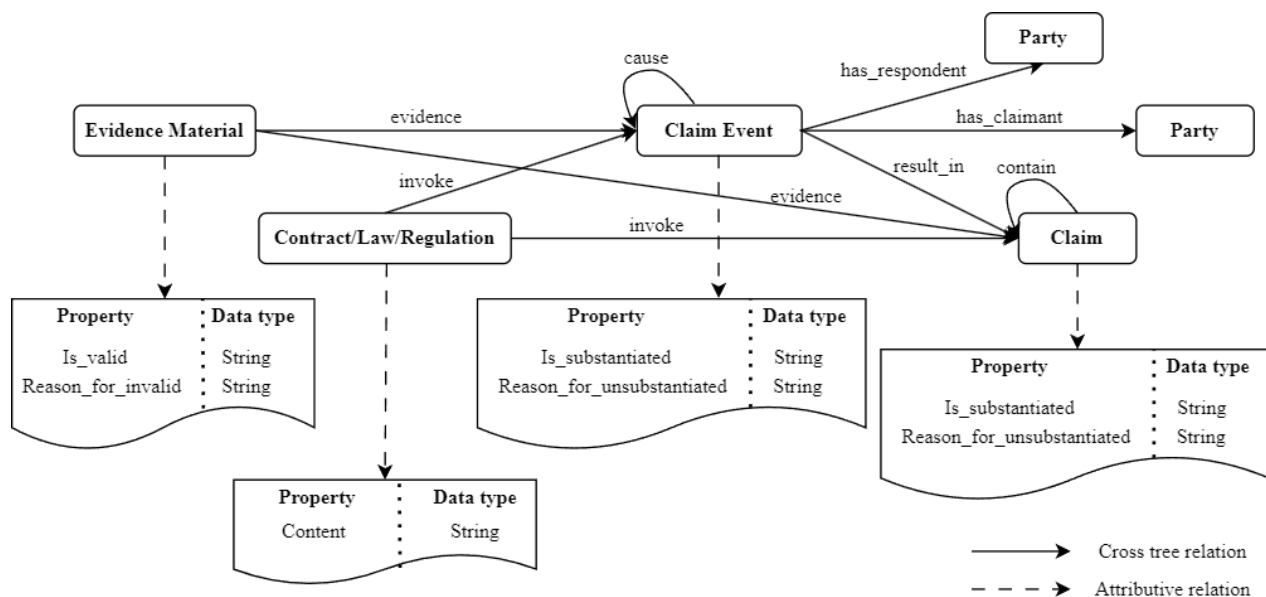


Figure 8. Properties and relations defined in ontology.

Step 5: Create the instances

All instances are stored and managed in a Neo4j graph database, which facilitates efficient storage, querying, and traversal of the resulting knowledge graph. Neo4j's query flexibility arises from its Cypher query language—a declarative, intuitive, and highly optimized mechanism for complex relationship traversals—enabling developers to express queries concisely without requiring extensive procedural expertise. The official neo4j-graphrag Python package delivers comprehensive end-to-end GraphRAG workflows, with seamless integration to the Neo4j driver for knowledge graph construction, hybrid retrieval,

The visualization of the knowledge graph equips practitioners with abundant domain knowledge, thereby facilitating effective claim management. For example, Claim event nodes provides owners and contractors with a comprehensive overview of potential claim events and their precursor risks early in the project lifecycle, including less common events that are prone to oversight in daily operations. Evidence material nodes allows contractors to recognize claim submission risks during the initial project stages and gain a thorough understanding of the evidentiary requirements for particular claim events.

3.2. Application and Evaluation

The constructed KG is then integrated with LLMs via RAG to form a specialized question-answering system capable of addressing claim-related queries. The performance of three configurations—(a) Base LLM, (b) Vector RAG, and (c) GraphRAG—is evaluated by comparing the generated answers against high-quality reference answers.

The input questions are carefully designed to align directly with real-world claim management tasks. For each question, responses generated under the three scenarios are quantitatively evaluated against the corresponding, reference answer using the widely adopted automatic metrics BLEU-4 [40], BERT-Cosine [41], ROUGE-1 and ROUGE-L [42].

BLEU-4 measures the surface-level n-gram overlap (up to 4 grams) between the generated text and the reference text, with a focus on precision. It is particularly suitable for tasks requiring strict adherence to standardized phrasing or template-like outputs such as machine translation or formal document generation. However, BLEU-4 tends to penalize valid paraphrasing, synonym substitution, and natural linguistic variation, which limits its sensitivity to semantic equivalence in more open-ended generation settings.

BERT-Cosine computes the cosine similarity between sentence embeddings produced by a BERT-based model for the generated and reference texts. By operating at the semantic level, this metric effectively captures meaning-preserving reformulations and is robust to differences in lexical choice. It is especially well-suited for evaluating question-answering and dialogue systems where semantic fidelity and conceptual alignment are more important than exact word matching.

ROUGE-1 measures unigram overlap between candidate and reference texts. It is defined in Equation (1)

$$\text{ROUGE-1} = \frac{\text{Number of overlapping unigrams}}{\text{Total unigrams in reference}} \quad (1)$$

This is a recall-oriented metric. To compute Precision, Recall, and F1, refer to Equations (2)–(4):

$$P_1 = \frac{\text{Overlapping unigrams}}{\text{Unigrams in candidate}} \quad (2)$$

$$R_1 = \frac{\text{Overlapping unigrams}}{\text{Unigrams in reference}} \quad (3)$$

$$F1_1 = \frac{2 \cdot P_1 \cdot R_1}{P_1 + R_1} \quad (4)$$

ROUGE-L uses the Longest Common Subsequence (LCS) to measure sentence-level structural similarity. Its precision, recall, and F1 are defined in Equations (5)–(7):

$$P_l = \frac{\text{LCS}(X, Y)}{|X|} \quad (5)$$

$$R_l = \frac{\text{LCS}(X, Y)}{|Y|} \quad (6)$$

$$F1_L = \frac{(1 + \beta^2) \cdot P_L \cdot R_L}{R_L + \beta^2 \cdot P_L} \quad (7)$$

The performance of the three LLM methods (GraphRAG, Vector RAG, and Base LLM) is objectively and scientifically compared by following these steps to determine whether the observed differences in their scores across various evaluation metrics are statistically significant.

1. Aggregation of Scores Within the Domain

For each evaluation metric, we collect the scores of all models (GraphRAG, Vector RAG, and Base LLM) across the evaluated instances.

2. Kruskal–Wallis Overall Test

We first apply the Kruskal–Wallis test to determine whether there are statistically significant differences in the score distributions among the three models. The null hypothesis is that all models have identical score distributions. If the p -value is less than 0.05, we reject the null hypothesis and proceed to pairwise comparisons.

3. Mann–Whitney U Pairwise Comparisons

For post hoc analysis, we conduct pairwise Mann–Whitney U tests on all possible combinations of the three models (resulting in 3 pairs). For each pair, we compute the U statistic, the corresponding p -value, and assess whether the score distributions differ significantly between the two models.

$$U = R_1 - \frac{n_1(n_1 + 1)}{2} \quad (8)$$

Here, R denotes the sum of ranks for the first group, and n represents the sample size of that group.

4. Significance Level and Marking

Statistical significance is assessed at $\alpha = 0.05$. Differences with $p < 0.05$ are marked with *, $p < 0.01$ with **, and $p < 0.001$ with ***.

4. Experiment

This section employs the claim event “work stoppage” as a representative case to evaluate the effectiveness of the proposed GraphRAG framework in a question-answering system for construction claims. Three baseline scenarios are compared: (a) Base LLM, (b) Vector RAG, and (c) GraphRAG. Under each scenario, the models are tasked with answering the following four domain-specific questions: (a) what events can lead to the claim event “work stoppage”, (b) what compensable items are available for claims arising from the “work stoppage” event, (c) what evidentiary materials can support claims triggered by a “work stoppage” in a construction project and how can these materials be ensured to meet legal validity requirements, and (d) which laws, regulations, or contractual clauses are primarily implicated in claims resulting from a “work stoppage” event in a construction project.

In GraphRAG, the retrieval process begins by identifying key entities and concepts from the input query. Relevant nodes and their associated relationships are then extracted from the knowledge graph based on these query elements. The retrieved sub-graph—comprising the selected nodes, edges, and related attributes—is converted into structured text and incorporated as contextual prompt information for the LLM. Leveraging the LLM’s powerful generation capabilities, the model synthesizes a coherent, accurate, and contextually grounded answer based on this precisely retrieved domain knowledge.

In Vector RAG, the retrieval process relies on semantic similarity: the embedding model computes vector representations of both the input query and the pre-split text chunks; then, it ranks and selects the top- k most semantically similar chunks from the

corpus. These retrieved chunks are subsequently formatted and provided as contextual prompt information to the LLM, which leverages this retrieved content to generate the final answer.

In all three experimental scenarios, we adopt DeepSeek-Chat as the base large language model to ensure fair comparison. For the Vector RAG configuration, we employ OpenAI's text-embedding-3-large as the embedding model to generate dense vector representations, combined with jina-reranker-v3 as the reranker. The reranker re-orders the initial candidate documents according to their semantic relevance to the user query, thereby refining and improving the quality of the final retrieved context beyond pure vector similarity. The retrieval pipeline uses a parent-document chunk size of 1024 tokens and a child-document chunk size of 512 tokens, with an overlap of 100 tokens between consecutive child chunks. During inference, we retrieve the top-10 most relevant child chunks after embedding-based retrieval and reranking, which are then passed to the LLM for generation.

For each of the four questions, three independent responses are generated under each scenario (Base LLM, Vector RAG, and GraphRAG), resulting in a total of 36 generated answers. These responses are provided in Supplementary Text S1. Each generated answer is then quantitatively compared against a high-quality reference text using multiple evaluation metrics, including BLEU-4, BERT-Cosine similarity, ROUGE-1, and ROUGE-L. For every question, the two scores per scenario are averaged to obtain a per-question mean score per metric. The mean values and statistical significance levels for GraphRAG, Vector RAG, and Base LLM on the BLEU-4 score are presented in Figure 10. Similarly, the mean values and significance levels for these three approaches on BERT-Cosine similarity, ROUGE-1, and ROUGE-L scores are shown in Figure 11, with complete individual scores and comparisons provided in Supplementary Table S1.

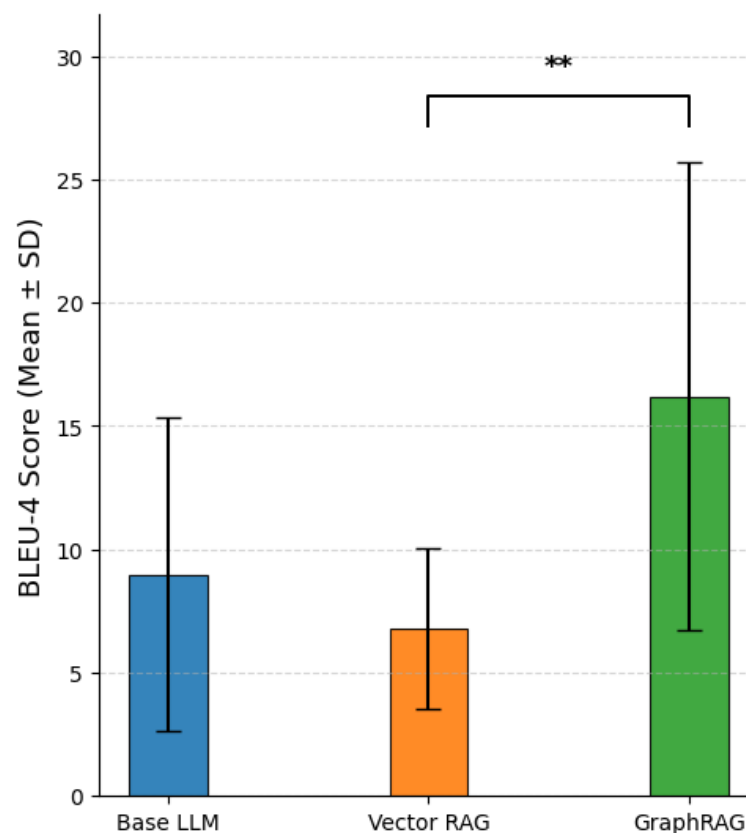


Figure 10. Performance comparison of different RAG architectures on BLEU-4 score. (Difference with $p < 0.01$ is marked with **).

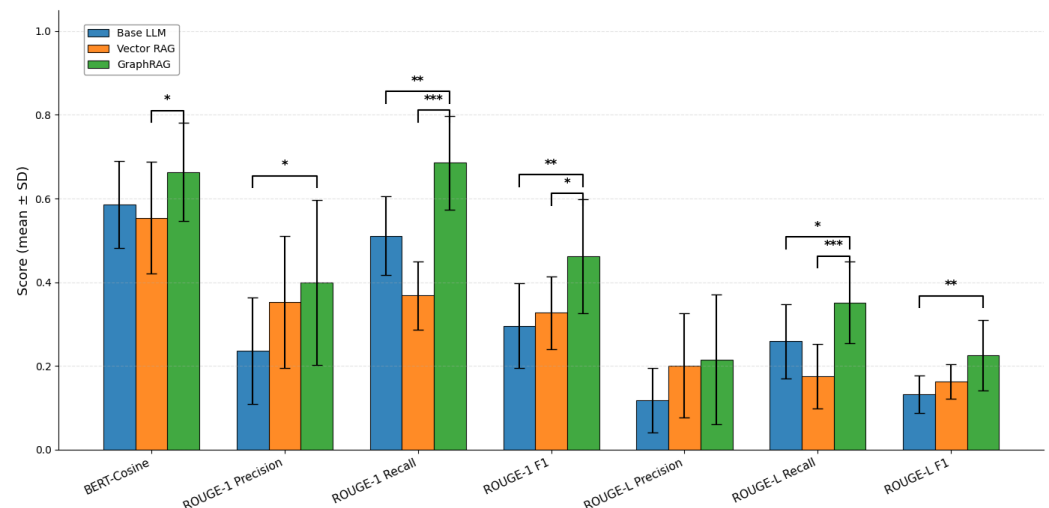


Figure 11. Performance comparison across retrieval and generation metrics: base LLM, vector-based RAG, and GraphRAG. (Differences with $p < 0.05$ are marked with *, $p < 0.01$ with **, and $p < 0.001$ with ***).

GraphRAG significantly outperforms both Vector RAG and the baseline LLM across the majority of automatic evaluation metrics. Its most pronounced statistical advantages appear in recall-oriented and composite F1 metrics, with the strongest gains observed in recall of key facts. Specifically, GraphRAG achieves substantially higher ROUGE-1 Recall and ROUGE-L Recall scores compared to Vector RAG (with p -values < 0.001 , indicating very strong statistical significance) and also shows clear superiority over the baseline LLM ($p < 0.01$ for ROUGE-1 Recall and $p < 0.05$ for ROUGE-L Recall). This demonstrates that GraphRAG's structured retrieval mechanism enables far more comprehensive coverage and retrieval of relevant knowledge points from the knowledge base—representing its primary and most significant strength. Interestingly, on recall metrics, the baseline LLM actually exhibits higher average performance than Vector RAG. In Vector RAG, although query-relevant chunks are retrieved, important knowledge points are often scattered across many chunks. Due to the fixed top- k retrieval limit, this frequently results in redundant information among the selected chunks, incomplete coverage (as some critical details fall outside the top- k results), and the introduction of noise or inaccuracies stemming from purely semantic similarity-based matching. In terms of precision, both GraphRAG and Vector RAG show noticeable advantages over the baseline LLM. By grounding generation in retrieved content from a verifiable knowledge base, RAG-based approaches substantially improve answer fidelity, reduce hallucination rates, and ensure outputs remain closely aligned with the provided source material.

5. Discussion

Our empirical evaluation of the proposed GraphRAG framework—built upon a knowledge graph constructed from a domain-specific ontology tailored to the construction claims domain—demonstrates superior performance compared to Base LLM and Vector RAG approaches. This superiority is clearly shown across multiple evaluation metrics, including BLEU-4, BERT-Cosine similarity, ROUGE-1, and ROUGE-L. By incorporating domain-specific knowledge extracted from industry texts, our method significantly improves the precision and recall of the LLM's retrieval process through more accurate and structured retrieval mechanisms. This integration results in answers that are both more precise and more comprehensive, effectively addressing the limitations of conventional chunk-based retrieval, such as insufficient diversity, as well as potential inaccuracies caused by se-

semantic similarity mismatches. In contrast, Vector RAG often retrieves chunks that fail to capture complete or non-redundant information due to top-k limitations, frequently introducing noise or omitting critical details. GraphRAG, however, leverages structured entity-relationship traversals within the knowledge graph to achieve more accurate and holistic knowledge coverage, ultimately producing responses that are semantically closer to the reference answers while exhibiting higher levels of precision and recall.

Furthermore, the KG-based architecture facilitates dynamic knowledge updates, enabling the seamless and real-time incorporation of newly emerging claim elements—such as updated contractual clauses, revised legal precedents, or novel evidentiary requirements—without necessitating retraining of the underlying LLM. This capability allows practitioners to promptly query and leverage the most current domain knowledge, ensuring that responses remain aligned with the latest developments in the field. This real-time adaptability is complemented by enhanced traceability, as each retrieved knowledge element can be directly linked back to its source ontology or textual origin, fostering transparency and verifiability in model outputs. In terms of interpretability, GraphRAG demystifies the Retrieval-Augmented Generation process by exposing the traversed graph paths and entity relations, allowing users to scrutinize the logical chain from query to response, which contrasts with the opaque nature of vector embeddings.

Query efficiency is another key benefit, as graph queries can be optimized through indexing and traversal algorithms, potentially reducing latency compared to exhaustive vector similarity searches over large corpora, especially in scenarios with complex, multi-hop relations. The ontology-driven knowledge extraction process itself represents a pivotal advancement, distilling focal domain knowledge from unstructured texts into a structured KG representation that faithfully captures intricate interdependencies among claim-related entities. This KG-augmented RAG paradigm elevates LLM efficacy by providing contextually precise knowledge injections, resulting in more accurate depictions of domain-specific nuances while bolstering interpretability and user confidence in the reliability of generated outputs. Overall, these attributes position GraphRAG as a robust solution for domain-adapted language modeling, with implications for scalable applications in knowledge-intensive industries.

This study holds substantial theoretical and practical implications for advancing knowledge-intensive applications in construction claims.

Methodologically, the primary contribution of this study lies in the development of a comprehensive domain-specific ontology for the construction claims domain. This ontology systematically formalizes the core concepts, entities, relationships, and attributes involved in claim management processes. Building upon this ontological foundation, a structured knowledge graph is constructed from domain texts and subsequently integrated with large language models through RAG, enabling contextually grounded, and factually reliable question answering in the field of construction claims. This end-to-end methodological pipeline—from ontology design and knowledge graph population to GraphRAG implementation—offers a replicable and transferable framework that can be adapted to other knowledge-intensive domains requiring structured domain knowledge infusion and enhanced reasoning capabilities in large language models.

On the practical front, the proposed GraphRAG framework empowers practitioners in construction project by delivering accurate, traceable, and interpretable claim-related insights, reducing decision-making uncertainties in dispute resolution and contract management. By mitigating hallucinations through KG-grounded retrieval, it supports real-time knowledge updates and queries, allowing professionals to access up-to-date information on evolving regulations and precedents without extensive manual research. This not only streamlines workflows in legal and project management contexts but also fosters more

reliable AI-assisted tools for industry stakeholders, potentially lowering litigation costs and improving compliance efficiency in dynamic construction environments.

However, this research presents some limitations.

First, the scope of the ontology and populated knowledge graph remains geographically and jurisdictionally constrained. Although the ontology is designed with sufficient generality to accommodate construction claims worldwide, the instantiated knowledge graph primarily incorporates Chinese legal frameworks, regulations, precedents, and case examples due to resource and data availability constraints. This limits the framework's immediate applicability to global, cross-border projects where local laws, cultural dispute resolution practices, or jurisdiction-specific precedents play a dominant role.

Second, the current approach focuses predominantly on qualitative knowledge integration and reasoning without extending to the quantitative dimensions of claims. While the present framework effectively supports qualitative reasoning and factual grounding to reduce hallucinations in textual responses, the absence of integrated quantitative computation limits its ability to address this core aspect of claims practice.

6. Conclusions

This study has successfully demonstrated the effectiveness of a knowledge graph-augmented Retrieval-Augmented Generation framework tailored to the domain of construction engineering claims. By first constructing a domain-specific ontology that systematically organizes key entities and then populating a structured knowledge graph from industry texts, the proposed methodology addresses longstanding limitations in traditional claim management practices.

Empirical evaluation using the representative claim event “work stoppage” showed that the GraphRAG approach consistently outperforms both the Base LLM and Vector RAG across multiple metrics: BLEU-4, BERT-Cosine similarity, ROUGE-1, and ROUGE-L. These results confirm that integrating a structured KG enables more precise, comprehensive, and semantically faithful retrieval of domain knowledge, leading to generated responses that exhibit higher factual accuracy, improved coverage of relevant knowledge points, and closer alignment with reference answers. In particular, GraphRAG mitigates the key shortcomings of Vector RAG—such as incomplete coverage due to chunk dispersion, redundancy, and noise introduction—while providing advantages in traceability, interpretability, and dynamic knowledge updates.

Despite these advances, limitations remain, including the current geographic focus of instantiated knowledge, the qualitative emphasis without integrated quantitative computation for time and cost quantification, and the need for broader multilingual and cross-jurisdictional expansion. Future research directions include incorporating quantitative delay and cost analysis modules and expanding the KG with international precedents and multilingual sources.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/buildings16040845/s1>, Text S1: LLM-Generated answers; Table S1: Complete Scores for Each LLM-Generated Answer Across Metrics.

Author Contributions: Conceptualization, X.W.; methodology, X.W.; software, X.W.; validation, X.W.; formal analysis, X.W.; investigation, X.W.; resources, X.W. and J.F.; data curation, X.W.; writing—original draft preparation, X.W.; writing—review and editing, X.W. and J.F.; visualization, X.W.; supervision, J.F. 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 presented in this study are available upon request from the corresponding author.

Acknowledgments: We thank the editors and reviewers of Buildings for their careful review and helpful comments on this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LLM	Large Language Model
RAG	Retrieval-Augmented Generation
KG	Knowledge Graph

References

1. Kalogeraki, M.; Antoniou, F. Claim management and dispute resolution in the construction industry: Current research trends using novel technologies. *Buildings* **2024**, *14*, 967. [\[CrossRef\]](#)
2. Alrasheed, K.; Bahman, M.; Soliman, E.; Albader, H. Enhancing Claims Management in the Construction Industry Using Analytic Hierarchy Process. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2025**, *17*, 04525038. [\[CrossRef\]](#)
3. Narayan, S.; Tan, H.C.; Jack, L.B. Claims management: A review of challenges faced. *Int. J. Constr. Manag.* **2024**, *24*, 1789–1795. [\[CrossRef\]](#)
4. Ali, B.; Zahoor, H.; Nasir, A.R.; Maqsoom, A.; Khan, R.W.A.; Mazher, K.M. BIM-based claims management system: A centralized information repository for extension of time claims. *Autom. Constr.* **2020**, *110*, 102937. [\[CrossRef\]](#)
5. Nyokum, T.; Tamut, Y. Artificial intelligence in civil engineering: Emerging applications and opportunities. *Front. Built Environ.* **2025**, *11*, 1622873. [\[CrossRef\]](#)
6. Asadi, R.; Rotimi, J.O.B.; Wilkinson, S. Analyzing underlying factors of rework in generating contractual claims in construction projects. *J. Constr. Eng. Manag.* **2023**, *149*, 04023036. [\[CrossRef\]](#)
7. Ghosh, B.; Karmakar, S. Assessing the causes of claims in highway construction projects with a case study: A construction practitioner's perspective. *Asian J. Civ. Eng.* **2024**, *25*, 3035–3048. [\[CrossRef\]](#)
8. Hasan, H.M.; Khodeir, L.; Yassa, N. Assessing the impact of claims on construction project performance using machine learning techniques. *Asian J. Civ. Eng.* **2024**, *25*, 5765–5779. [\[CrossRef\]](#)
9. Mishmish, M.; El-Sayegh, S.M. Causes of claims in road construction projects in the UAE. *Int. J. Constr. Manag.* **2018**, *18*, 26–33. [\[CrossRef\]](#)
10. Shen, W.; Tang, W.; Yu, W.; Duffield, C.F.; Hui, F.K.P.; Wei, Y.; Fang, J. Causes of contractors' claims in international engineering-procurement-construction projects. *J. Civ. Eng. Manag.* **2017**, *23*, 727–739. [\[CrossRef\]](#)
11. Do, S.T.; Nguyen, V.T.; Tran, C.N.; Aung, Z.M. Identifying and evaluating the key claim causes leading to construction delays. *Int. J. Constr. Manag.* **2023**, *23*, 1999–2011. [\[CrossRef\]](#)
12. Zanelidin, E.K. Investigating the types, causes and severity of claims in construction projects in the UAE. *Int. J. Constr. Manag.* **2020**, *20*, 385–401. [\[CrossRef\]](#)
13. Liao, L.; Wei, N.; Zheng, C.; Ye, Y.; Chen, Y. Understanding causes for construction claims in hydropower projects. *J. Manag. Eng.* **2023**, *39*, 04023032. [\[CrossRef\]](#)
14. Techapeeraparnich, W.; Oo, N.M.S.; Wantanakorn, D.; Limsawasd, C. Problems and Causes of Contractors' Claims in Thai Construction Projects. In *Proceedings of the International Conference on Civil Engineering and Architecture*; Springer: Singapore, 2023; pp. 315–330.
15. Akinradewo, F.O.; Oladinrin, O.T. A quantitative study on the magnitude of construction claims in construction projects in Nigeria. *J. Constr. Proj. Manag. Innov.* **2018**, *8*, 1872–1885. Available online: <https://journals.co.za/doi/abs/10.10520/EJC-13434ddbcc> (accessed on 1 January 2026).
16. Enshassi, A.; Choudhry, R.M.; El-Ghandour, S. Contractors' perception towards causes of claims in construction projects. *Int. J. Constr. Manag.* **2009**, *9*, 79–92. [\[CrossRef\]](#)
17. Saxena, A.; Tripathi, A.; Talukdar, P. Improving multi-hop question answering over knowledge graphs using knowledge base embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 4498–4507.
18. Jiang, J.; Zhou, K.; Zhao, W.X.; Wen, J.R. Uniqgqa: Unified retrieval and reasoning for solving multi-hop question answering over knowledge graph. *arXiv* **2022**, arXiv:2212.00959.

19. Jiang, J.; Zhou, K.; Zhao, W.X.; Song, Y.; Zhu, C.; Zhu, H.; Wen, J.R. Kg-agent: An efficient autonomous agent framework for complex reasoning over knowledge graph. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2025; pp. 9505–9523.
20. Sun, J.; Xu, C.; Tang, L.; Wang, S.; Lin, C.; Gong, Y.; Ni, L.M.; Shum, H.Y.; Guo, J. Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph. *arXiv* **2023**, arXiv:2307.07697.
21. Procko, T.T.; Ochoa, O. Graph retrieval-augmented generation for large language models: A survey. In *Proceedings of the 2024 Conference on AI, Science, Engineering, and Technology (AIxSET)*; IEEE: Piscataway, NJ, USA, 2024; pp. 166–169.
22. Edge, D.; Trinh, H.; Cheng, N.; Bradley, J.; Chao, A.; Mody, A.; Truitt, S.; Metropolitan, D.; Ness, R.O.; Larson, J. From local to global: A graph rag approach to query-focused summarization. *arXiv* **2024**, arXiv:2404.16130. [\[CrossRef\]](#)
23. Luo, L.; Li, Y.F.; Haffari, G.; Pan, S. Reasoning on graphs: Faithful and interpretable large language model reasoning. *arXiv* **2023**, arXiv:2310.01061.
24. Kim, J.; Park, S.; Kwon, Y.; Jo, Y.; Thorne, J.; Choi, E. FactKG: Fact verification via reasoning on knowledge graphs. *arXiv* **2023**, arXiv:2305.06590. [\[CrossRef\]](#)
25. Wang, S.; Yang, H.; Liu, W. Research on the construction and application of retrieval enhanced generation (RAG) model based on knowledge graph. *Sci. Rep.* **2025**, *15*, 40425. [\[CrossRef\]](#)
26. Linders, J.; Tomczak, J.M. Knowledge graph-extended retrieval augmented generation for question answering. *arXiv* **2025**, arXiv:2504.08893. [\[CrossRef\]](#)
27. Xu, Y.; He, S.; Chen, J.; Wang, Z.; Song, Y.; Tong, H.; Liu, G.; Liu, K.; Zhao, J. Generate-on-graph: Treat llm as both agent and kg in incomplete knowledge graph question answering. *arXiv* **2024**, arXiv:2404.14741.
28. Wei, W.; Ren, X.; Tang, J.; Wang, Q.; Su, L.; Cheng, S.; Wang, J.; Yin, D.; Huang, C. Llmrec: Large language models with graph augmentation for recommendation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*; Association for Computing Machinery: New York, NY, USA, 2024; pp. 806–815.
29. Cui, Z.; Weng, Y.; Tang, X.; Lyu, F.; Liu, D.; He, X.; Ma, C. Comprehending knowledge graphs with large language models for recommender systems. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*; Association for Computing Machinery: New York, NY, USA, 2025; pp. 1229–1239.
30. Wang, S.; Fan, W.; Feng, Y.; Shanru, L.; Ma, X.; Wang, S.; Yin, D. Knowledge graph retrieval-augmented generation for llm-based recommendation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2025; pp. 27152–27168.
31. Guo, T.; Liu, C.; Wang, H.; Mannam, V.; Wang, F.; Chen, X.; Zhang, X.; Reddy, C.K. Knowledge graph enhanced language agents for recommendation. *arXiv* **2024**, arXiv:2410.19627. [\[CrossRef\]](#)
32. Yang, Z.; Khatibi, E.; Nagesh, N.; Abbasian, M.; Azimi, I.; Jain, R.; Rahmani, A.M. ChatDiet: Empowering personalized nutrition-oriented food recommender chatbots through an LLM-augmented framework. *Smart Health* **2024**, *32*, 100465. [\[CrossRef\]](#)
33. Cui, Y.; Wang, K.; Yu, H.; Guo, X.; Cao, H. KLLMs4Rec: Knowledge graph-enhanced LLMs sentiment extraction for personalized recommendations. *Expert Syst. Appl.* **2025**, *282*, 127430. [\[CrossRef\]](#)
34. Feng, Y.; Zhou, L.; Ma, C.; Zheng, Y.; He, R.; Li, Y. Knowledge graph-based thought: A knowledge graph-enhanced LLM framework for pan-cancer question answering. *GigaScience* **2025**, *14*, giae082. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Xu, Z.; Cruz, M.J.; Guevara, M.; Wang, T.; Deshpande, M.; Wang, X.; Li, Z. Retrieval-augmented generation with knowledge graphs for customer service question answering. In *Proceedings of the 47th international ACM SIGIR Conference on Research and Development in Information Retrieval*; Association for Computing Machinery: New York, NY, USA, 2024; pp. 2905–2909.
36. Chen, Q.; Yin, X.; Yuan, B.; Chen, Q. Personalized safety training for construction workers: A large language model-driven multi-agent framework integrated with knowledge graph reasoning. *Comput. Ind.* **2026**, *174*, 104399. [\[CrossRef\]](#)
37. Zheng, C.; Wong, S.; Su, X.; Tang, Y.; Nawaz, A.; Kassem, M. Automating construction contract review using knowledge graph-enhanced large language models. *Autom. Constr.* **2025**, *175*, 106179. [\[CrossRef\]](#)
38. Zhang, D.; Ma, G.; Qu, T.; Wang, X.; Zhou, W.; Wang, X. A knowledge graph-enhanced large language model for question answering of hydraulic structure safety management. *Adv. Eng. Inform.* **2025**, *66*, 103468. [\[CrossRef\]](#)
39. Noy, N.F.; McGuinness, D.L. Ontology Development 101: A Guide to Creating Your First Ontology, Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880. 2001. Available online: https://corais.org/sites/default/files/ontology_development_101_aguide_to_creating_your_first_ontology.pdf (accessed on 1 January 2026).
40. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.J. Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2002; pp. 311–318.

41. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 4171–4186.
42. Lin, C.Y. ROUGE: A Package for Automatic Evaluation of Summaries, Text Summarizat. Branches Out 74–81. 2004. Available online: <https://aclanthology.org/W04-1013.pdf> (accessed on 1 January 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.