

Article

Development of an Automated Construction Contract Review Framework Using Large Language Model and Domain Knowledge

Eu Wang Kim ¹, Yeon Ju Shin ¹, Kyong Ju Kim ² and Sehoon Kwon ^{3,*}

¹ Department of Smart Cities, Chung-Ang University, Seoul 06974, Republic of Korea; kew1013@cau.ac.kr (E.W.K.); shinyeanju@cau.ac.kr (Y.J.S.)

² Department of Civil & Environmental Engineering, Chung-Ang University, Seoul 06974, Republic of Korea; kjkim@cau.ac.kr

³ Samsung C&T, Seoul 04514, Republic of Korea

* Correspondence: archekon@naver.com

Abstract: Construction contract review demands specialized expertise, requiring comprehensive understanding of both technical and legal aspects. While AI advancements offer potential solutions, two problems exist: LLMs lack sufficient domain-specific knowledge to analyze construction contracts; existing RAG approaches do not effectively utilize domain expertise. This study aims to develop an automated contract review system that integrates domain expertise with AI capabilities while ensuring reliable analysis. By transforming expert knowledge into a structured knowledge base aligned with the SCF classification, the proposed structured knowledge-integrated RAG pipeline is expected to enable context-aware contract analysis. This enhanced performance is achieved through three key components: (1) integrating structured domain knowledge with LLMs, (2) implementing filtering combined with hybrid dense-sparse retrieval mechanisms, and (3) employing reference-based answer generation. Validation using Oman's standard contract conditions demonstrated the system's effectiveness in assisting construction professionals with contract analysis. Performance evaluation showed significant improvements, achieving a 52.6% improvement in Context Recall and a 48.3% improvement in Faithfulness compared to basic RAG approaches. This study contributes to enhancing the reliability of construction contract review by applying a structured knowledge-integrated RAG pipeline that enables the accurate retrieval of expert knowledge, thereby addressing the industry's need for precise contract analysis.

Keywords: contract review; automated contract analysis; retrieval-augmented generation; large language models; knowledge base; standard contract form

Academic Editors: John Kamara and Farook Hamzeh

Received: 3 January 2025

Revised: 18 February 2025

Accepted: 12 March 2025

Published: 14 March 2025

Citation: Kim, E.W.; Shin, Y.J.; Kim, K.J.; Kwon, S. Development of an Automated Construction Contract Review Framework Using LLM and Domain Knowledge.

Buildings **2025**, *15*, 923.

<https://doi.org/10.3390/buildings15060923>

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Construction projects often encounter disputes due to inadequate contract management, including ambiguous terms, poor communication, and misinterpretations of obligations. Contract document errors and omissions are primary causes of disputes in major construction markets [1], significantly impacting project success by escalating costs and delaying timelines. Recent industry reports highlight a dramatic rise in construction dispute costs over a short period [2].

Construction contract review is critical as dispute values continue to rise, requiring timely and specialized expertise. Generally, major Korean construction companies have substantial global project expertise. They deploy diverse teams of professionals with 15–20 years of experience in legal, contract, project management, and risk management fields. These teams manage comprehensive project preparation cycles that typically span 4–6 months, including the bidding and contract negotiation phase.

Engineers often struggle with legal aspects, while lawyers may lack understanding of technical details [3] (p. 10). International projects add complexity due to varying legal systems, necessitating the knowledge of diverse contracts and legal systems [4]. Expertise in contract review demands understanding and interpreting complex conditions involving both technical and legal knowledge. This requires deep familiarity with practices that vary according to project type and characteristics and according to country and regional preferences. Professionals must integrate these aspects to identify risks, provide clear interpretations, and align with project objectives. Particularly in the case of FIDIC forms, the complexity is heightened, where highly interconnected clauses necessitate a comprehensive analysis of their interactions [5] (p. 30). Numerous cross-references and complex sentence structures further increase the risk of misunderstandings [6].

The emergence of Artificial Intelligence (AI) and Large Language Models (LLMs) offers promising approaches to provide expert knowledge and increase productivity in contract management. AI-driven contract analysis tools, in particular, hold the potential to streamline contract review processes and strengthen risk management capabilities. By automating repetitive tasks, these technologies enable construction professionals to efficiently manage contracts. However, applying current LLMs in contract management presents limitations. As general-purpose models, they raise concerns about the trustworthiness of their outcomes in precision-critical contractual contexts. The lack of domain-specific knowledge leads to missing subtle meanings in complex clauses, which hinders the models' ability to interpret the contract holistically. This makes it difficult to provide reliable analysis across the entire document, as nuances and interdependencies between clauses may be overlooked or misinterpreted.

To overcome these limitations, this study proposes a structured knowledge-integrated RAG pipeline that enhances the Retrieval-Augmented Generation (RAG) process for construction contract review. The proposed approach includes three key components: (1) a structured knowledge framework that systematically organizes technical and legal expertise; (2) a two-stage retrieval mechanism combining classification-based filtering with dense–sparse retrieval methods for accurate improved knowledge retrieval; and (3) reference-based answer generation that improves response accuracy and reliability.

This research aims to bridge the gap between contract review challenges in practice and the limitations of general purpose LLMs and RAG approaches. The structured knowledge-integrated approach enhances the practical applicability of automated contract review tools in construction projects by addressing key challenges in knowledge retrieval and response generation. This research hypothesizes that the structured knowledge-integrated RAG pipeline provides significantly improved performance in construction contract review compared to existing LLMs and RAG approaches, particularly in terms of retrieval accuracy and response generation reliability.

2. Research Background

2.1. AI and LLM Applications in Contract Management

The application of AI and LLMs in construction contract management has progressed substantially, leading to significant technological advancements as well as various implementation challenges. Early research predominantly employed rule-based Natural Language Processing (NLP) methods for risk detection, compliance checking, and clause

analysis [7–11]. These approaches were later expanded to include uncertainty analysis and automated compliance checking [12,13].

Conventional machine learning methods began with litigation outcome prediction using various classifiers [14]. Subsequent studies implemented Support Vector Machines, Naïve Bayes, and other machine learning approaches for compliance checking and requirement classifications [15–17]. Hybrid approaches combining rule-based and machine learning methods were also explored [18,19].

Deep learning and transformer-based models further enhanced contract analysis capabilities, including Bi-LSTM, Conditional Random Fields, and BERT [20–23]. Recent advancements employ architectures like DeBERTa and pre-trained BERT for risk prediction and dispute classification [24–27]. More recently, LLMs with knowledge-augmented approaches have shown significant potential in contract management [28,29].

However, current AI solutions face significant limitations. LLMs struggle with domain-specific knowledge not widely represented in their training data [30]. Simply increasing the model size does not significantly help in recalling specialized information [31].

Figure 1 examines FIDIC Sub-clause 19.2's 14-day notice requirement and whether it acts as a time-bar provision. This analysis reveals a critical limitation in LLMs' contract analysis capabilities. Although experienced construction professionals can recognize that Clause 19.2 operates as a time-bar, three LLMs (GPT o1, GPT-4 Turbo, and Claude Sonnet) all incorrectly concluded otherwise. Their responses revealed two major oversights: focusing solely on explicit language in one clause and missing the contractual link between Sub-clause 19.2 and 19.4, where timely notice is a prerequisite for Force Majeure relief. This example highlights the challenges that current LLMs face in processing the complex, the interconnected nature of construction contracts. Proper interpretation requires deep domain knowledge and the ability to analyze clauses comprehensively, not in isolation.

Current methods also face substantial challenges, including performance degradation when handling lengthy documents, with accuracy dropping to 55% for contracts exceeding ten paragraphs [32]. These limitations are particularly evident in RAG systems, which face challenges in handling large volumes of unstructured data and generating reliable outputs when processing complex contract conditions.

While AI and LLM-based approaches have made progress—from early rule-based methods to advanced deep learning and knowledge-augmented techniques—several gaps remain. Existing studies often rely on semantic or lexical similarity, making it difficult to capture the contextual interplay of technical and legal details essential for construction contracts. Domain knowledge is often missing from general-purpose LLMs. Moreover, unstructured and lengthy contractual documentation can significantly undermine retrieval performance, which in turn degrades the effectiveness of RAG systems. Hence, a more specialized approach that systematically incorporates expert domain knowledge is required to overcome the fundamental difficulties of complex contractual structures.

Question

Question: The following pertains to FIDIC Clause 19.2. This clause specifies a period of 14 days. Does this constitute a time-bar provision?

Context:

- 19.2 [Notice of Force Majeure]:

... The notice shall be given within 14 days after the Party became aware, or should have become aware, of the relevant event or circumstance constituting Force Majeure ...

- Definition: A time bar is a contractual provision that requires a party (usually the contractor) to submit a notice or claim within a specified ...

- Key Features: Time bars serve as condition precedents in construction contracts, enabling timely communication of potential issues and ...

- Example Clause: From FIDIC Clause 20.1: "If the Contractor fails to give notice of a claim within such period of 28 days, the Time for ...

General-purpose LLM's response

Date: 28 January 2025

GPT o1	GPT 4 Turbo	Claude Sonnet
No, FIDIC Clause 19.2 does not constitute a time-bar provision in the strict sense. While it requires a party to give notice of a Force Majeure event within 14 days of becoming aware (or when it should have become aware) of the event or circumstances constituting the Force Majeure, it does not explicitly state that failure to provide such notice within the specified timeframe results in the loss of the right to claim relief under the Force Majeure provisions. ...	No, the 14-day period specified in FIDIC Clause 19.2 does not constitute a time-bar provision. A time-bar provision in a contract is one that explicitly states that if a party fails to give notice or make a claim within a specified timeframe, they lose their right to claim relief, such as time extensions or additional payments. ...	Clause 19.2's notice requirement likely does not constitute a time-bar provision because: - It lacks explicit consequences for non-compliance - The language is not sufficiently clear to establish it as a condition precedent - The purpose appears informational rather than restrictive - The Force Majeure context suggests a more flexible interpretation ...

Figure 1. Example of LLMs' limitation in interpreting FIDIC clauses.

2.2. Research Problems and Motivation

This study aims to address two critical research problems in automated construction contract review. First, LLMs lack sufficient domain-specific knowledge to analyze construction contracts effectively. The process demands simultaneous understanding of both technical and legal aspects, yet LLMs struggle to integrate and apply this specialized expertise reliably. Second, existing RAG approaches do not effectively utilize domain expertise in their retrieval and answer generation processes. They rely heavily on semantic and lexical similarity, limiting their ability to capture deeper contractual relationships and implications that require domain knowledge.

These challenges necessitate a new approach that can systematically integrate technical and legal expertise while ensuring reliable and consistent automated analysis. By integrating structured domain knowledge into the RAG pipeline, the proposed system aims to overcome current limitations while enhancing the accuracy and consistency of automated contract reviews. This is achieved by implementing a structured knowledge framework that systematically organizes technical and legal expertise, developing an enhanced retrieval mechanism for complex documents, and establishing reliable answer generation processes based on structured reference knowledge.

3. Materials and Methods

3.1. System Requirements

To effectively address these challenges, the proposed construction contract review system must meet several requirements. These requirements focus on integrating expert knowledge, incorporating precise retrieval mechanisms, generating accurate and

traceable answers, and maintaining connections to reference sources, all of which are crucial for processing complex construction contracts.

First, the system needs to integrate technical and legal expertise through a systematically organized knowledge structure. LLMs often struggle with domain-specific knowledge that is not widely represented in their training data—especially in specialized fields like construction contracts. Therefore, integrating structured knowledge into the RAG pipeline is essential for accurately analyzing specific clauses and handling specialized terms, ensuring comprehensive coverage and precise interpretation of contract clauses.

Second, the system must incorporate precise retrieval mechanisms to effectively retrieve relevant clauses and information from extensive text documents. To overcome limitations in handling large volumes of unstructured information and maintaining consistent interpretations, the system should employ classification-based filtering to extract pertinent information, even from lengthy and complex documents with interconnected clauses.

Third, the system must generate answers that are not only accurate but also traceable to their sources. Given the uncertainty inherent in LLM outputs, implementing reference-based answer generation ensures all responses are supported by and linked to specific sections of the knowledge base. This approach enhances the reliability and trustworthiness of the automated analysis, allowing users to verify outputs.

3.2. System Architecture

The system's core functionality comprises three main components, as illustrated in Figure 2. First, the knowledge base contained structured information required for contract review. It stored SCF information and organized expert knowledge, mapping clauses to relevant expert analyses, legal information, and practical considerations. This structured repository provided a clear framework for analyzing various types of construction contracts.

Second, the Retrieval Component employed a two-stage retrieval mechanism to effectively retrieve relevant information. In the first stage, classification-based filtering narrowed down pertinent sections by categorizing clauses based on predefined criteria. In the second stage, dense–sparse retrieval methods extracted precise information from the filtered results. This combination ensures both precision and relevance, enabling the system to identify modified, added, and omitted clauses by comparing the input contract against the SCF, even in lengthy and complex documents with interconnected clauses.

Third, the Generation Component performed differential analysis by comparing the input contracts to the SCF. It utilized contextual understanding to identify deviations and assess their implications. By leveraging the robust knowledge base and the precise retrieval of relevant information, the system generated trustworthy contract analyses. These components worked together to ensure high accuracy and consistency in the generated results.

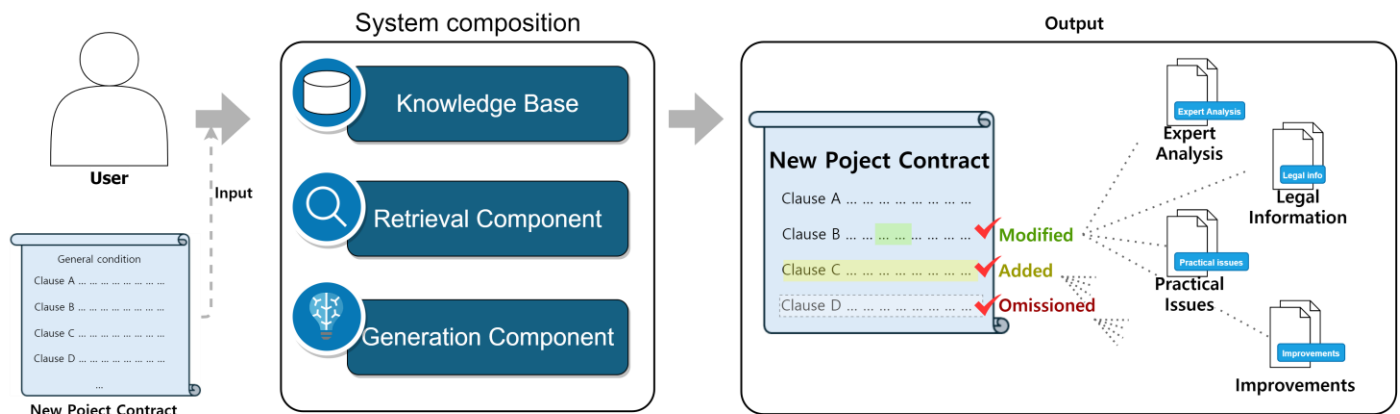


Figure 2. System architecture.

Conventional approaches to contract analysis face significant limitations. Base LLMs rely solely on basic user prompts, leading to hallucination issues and high fine-tuning costs for domain-specific tasks. Search-augmented LLMs, while incorporating web search capabilities, struggle with unreliable information sources and often blend non-expert opinions with expert knowledge. Basic RAG systems, though more advanced, fail to effectively handle complex contractual contexts due to scattered domain information and unstructured expert knowledge.

The proposed method features structured domain knowledge and organized expert insights integrated into the RAG pipeline. Structuring the Knowledge Base according to SCFs provides a clear framework for analyzing construction contracts. This integration enhanced the accuracy of contract clause interpretation while maintaining the efficiency of automated analysis. Consequently, the system delivered transparent and traceable results, effectively addressing the complexities inherent in construction contract review processes. As shown in Table 1, the proposed method offers significant advantages over conventional approaches in information retrieval mechanisms and context processing methods.

Table 1. Comparison of information retrieval and context processing methods.

	Information Retrieval Mechanism	Context Information	Prompt	Limitations
Base LLM	None	None	- Basic user prompt only	- Hallucination issues - High cost requirements (for fine-tuning)
Search-Augmented LLM	Web search	- Unstructured web information - Combined expert and non-expert information - Dynamic online content	- User prompt - Retrieved web context	- Unreliable information sources - Mixes expert/non-expert content
Basic RAG	Corpus-based retrieval	- Scattered domain information - Unstructured expert knowledge	- User prompt - Retrieved context	- Poor information retrieval for complex reasoning - Limited context understanding
Proposed Method	Knowledge-based retrieval	- Structured domain information - Organized expert knowledge	- User prompt - Retrieved context	- High initial costs - Requires ongoing maintenance

3.3. Knowledge Base Component

This section addresses the first requirement: the systematic integration of expert knowledge to support accurate interpretation and analysis of contract clauses. The knowledge base includes expert knowledge organized around the SCF, mapping each clause to relevant expert analyses, legal information, and practical considerations. The knowledge base was developed through multiple expert sources: comprehensive clause-by-clause expert analysis and legal interpretations [3] (Section 4), fundamental practical considerations for contract implementation [33] (Section 2), and region-specific practical issues when applying FIDIC conditions in the United Arab Emirates construction context [34] (pp. 293–385).

Figure 3 illustrates the process of integrating knowledge into the SCF classification within the knowledge base. The workflow encompassed data collection from both SCFs and related guidebooks, highlighting the extraction of relevant information for each clause. Text preprocessing and information processing stages prepared the data by cleaning and standardizing it. The aggregation stage consolidated this information, aligning it with the knowledge base structure through the SCF classification. This classification ensured that each clause is precisely categorized and linked, facilitating comprehensive contextual analysis that considers the context of each clause. By employing the SCF classification, the system systematically organized contract clauses, ensuring that each clause is accurately mapped to its corresponding knowledge.

Table 2 presents the composition of the knowledge base, integrating each SCF clause or sub-clause with three specific types of information. The knowledge was structured around the FIDIC Red Book 1999, providing a detailed framework for each contract clause.

Each sub-clause of the FIDIC Red Book was systematically documented, integrating the original contract text (165 corpuses; averaging 168 words per clause) with three key components: (1) expert analysis (165 corpuses: averaging 1042 tokens each) providing detailed explanations of contractual requirements and obligations; (2) legal information (106 corpuses: averaging 468 tokens each) covering legal requirements and jurisdictional considerations; and (3) practical considerations (123 corpuses: averaging 98 tokens per each) addressing common challenges and historical problems.

The knowledge base requires continuous updates to incorporate new acquired information from project implementations and to reflect occasional changes in SCFs. Within this structured framework, metadata-based management was implemented to maintain consistency and manage updates effectively. Each instance in the SCF classification system included metadata fields such as version number, publication date, and revision history. This metadata supported flexible reflection of updated information including minor changes, SCF revisions, and reprints and acquired knowledge through project lessons learned while maintaining the hierarchical organization of contract clauses.

Table 2. Composition of the knowledge base for FIDIC Red 1999.

Type	Description	Number of Corpus	Average Number of Tokens
Clause/Sub-Clause of SCF	Clause text from standard contract forms, providing the base reference for analysis and comparison	165	168
Expert Analysis	Detailed explanations of contractual requirements and obligations	165	1042
Legal Information	Case law, legal requirements, and jurisdictional considerations	106	468
Practical Considerations	Common challenges and historical problems	123	98

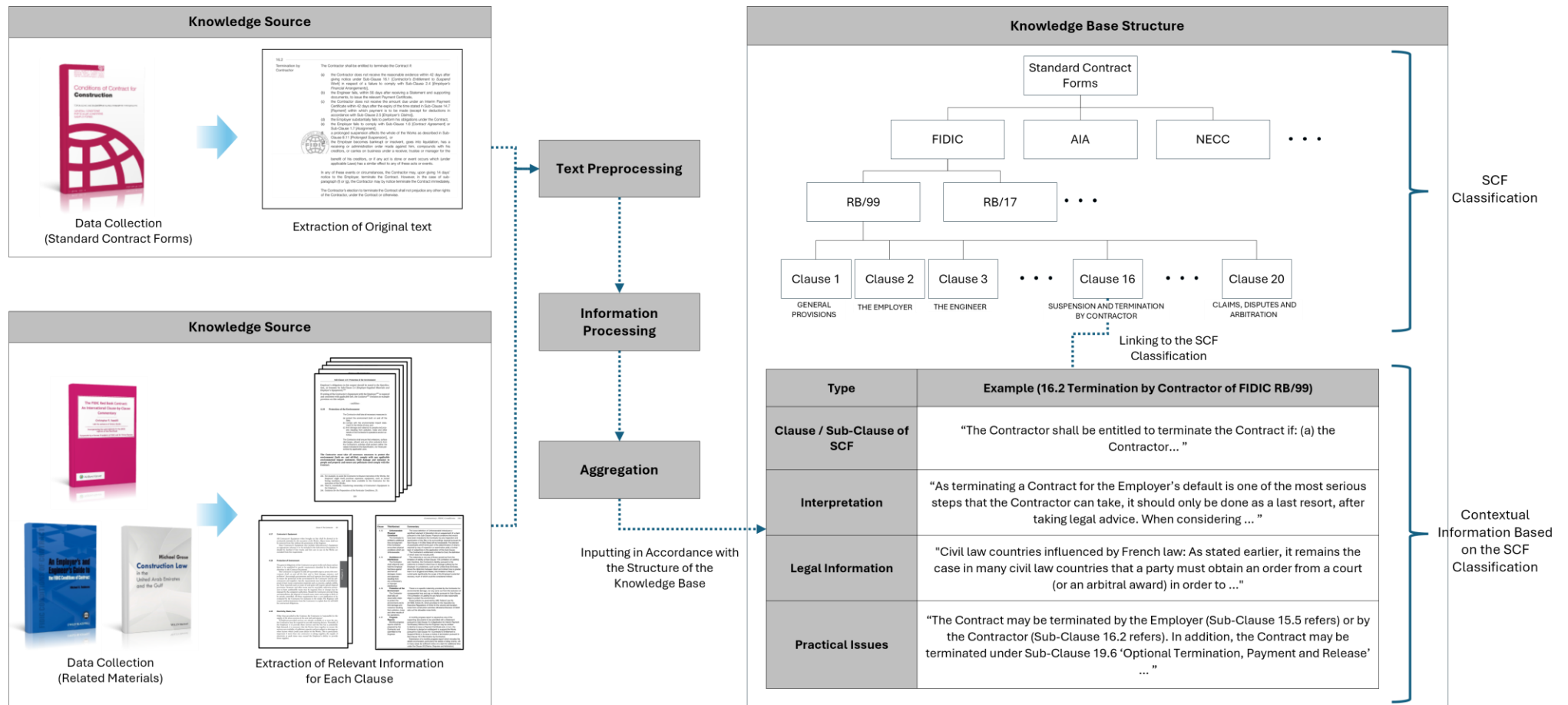


Figure 3. Structured integration process of knowledge into the SCF classification within the knowledge base.

3.4. Retrieval Component

This section covers the second requirement, inclusion of precise retrieval mechanisms. Combining classification-based filtering with hybrid search methods, this component accurately and efficiently retrieved relevant information from complex contract documents.

The retrieval mechanism operates as a two-stage process. First, clause classification based on SCF classification narrowed the search scope within the knowledge base, ensuring that only pertinent clauses are considered, reducing noise and improving efficiency. Second, a hybrid search combined dense vector embeddings and sparse keyword-based methods to precisely identify relevant information, an approach that showed superior performance in text retrieval tasks [35]. Dense retrieval implemented FAISS, a k-NN search library, with the text-embedding-3-small model for semantic similarity searches, mapping textual data into a high-dimensional vector space to capture contextual meaning. Sparse retrieval employed BM25—a well-established TF-IDF-based function that ranks documents according to query terms.

This two-stage retrieval mechanism effectively targeted specific domain knowledge, enhancing the accuracy and reliability of contract analysis. By processing input contracts through this combined classification and hybrid retrieval method—integrating both dense and sparse retrieval techniques with relationship-aware ranking that considers interconnections among contract elements—the system effectively identified relevant clauses and associated expert knowledge.

Furthermore, using SCF classification in the initial filtering reduces the search space and aligns the retrieval process with standardized contract structures. This alignment is crucial for maintaining consistency and ensuring that the retrieved information is contextually relevant. By structuring the retrieval mechanism in this way, the system efficiently handles the complexity and length of construction contracts, providing accurate and reliable analytical results.

3.5. Generation Component

This section addresses the third requirement: generating accurate and traceable answers by maintaining connections to reference sources. The Generation Component is responsible for producing contract analyses based on the information prepared by the Retrieval Component and the structured knowledge base.

After the Retrieval Component identified relevant clauses and associated expert knowledge, the Generation Component performed a differential analysis between the input contract clauses and the corresponding SCF clauses. By comparing the input contract to the standard forms, it identified discrepancies such as modifications, additions, or omissions. Utilizing contextual information from the knowledge base—including expert analyses, legal information, and practical considerations—the component assessed the implications of these discrepancies. It ensured that each conclusion and recommendation is explicitly linked to specific sections of the knowledge base and SCF clauses, allowing users to easily trace the generated analyses back to their underlying sources, thus enhancing trustworthiness.

Figure 4 illustrates the sequence diagram of the proposed method, detailing the interactions between the system components during the contract analysis process. The process begins with the user submitting an input contract to the system. The System Controller received the contract and orchestrated the analysis by coordinating with the Retrieval Component. The Retrieval Component classified the input contract clauses based on the SCF classification to narrow down relevant sections. It then performs a hybrid retrieval

using both dense vector embeddings and sparse keyword-based searches to fetch pertinent information from the knowledge base.

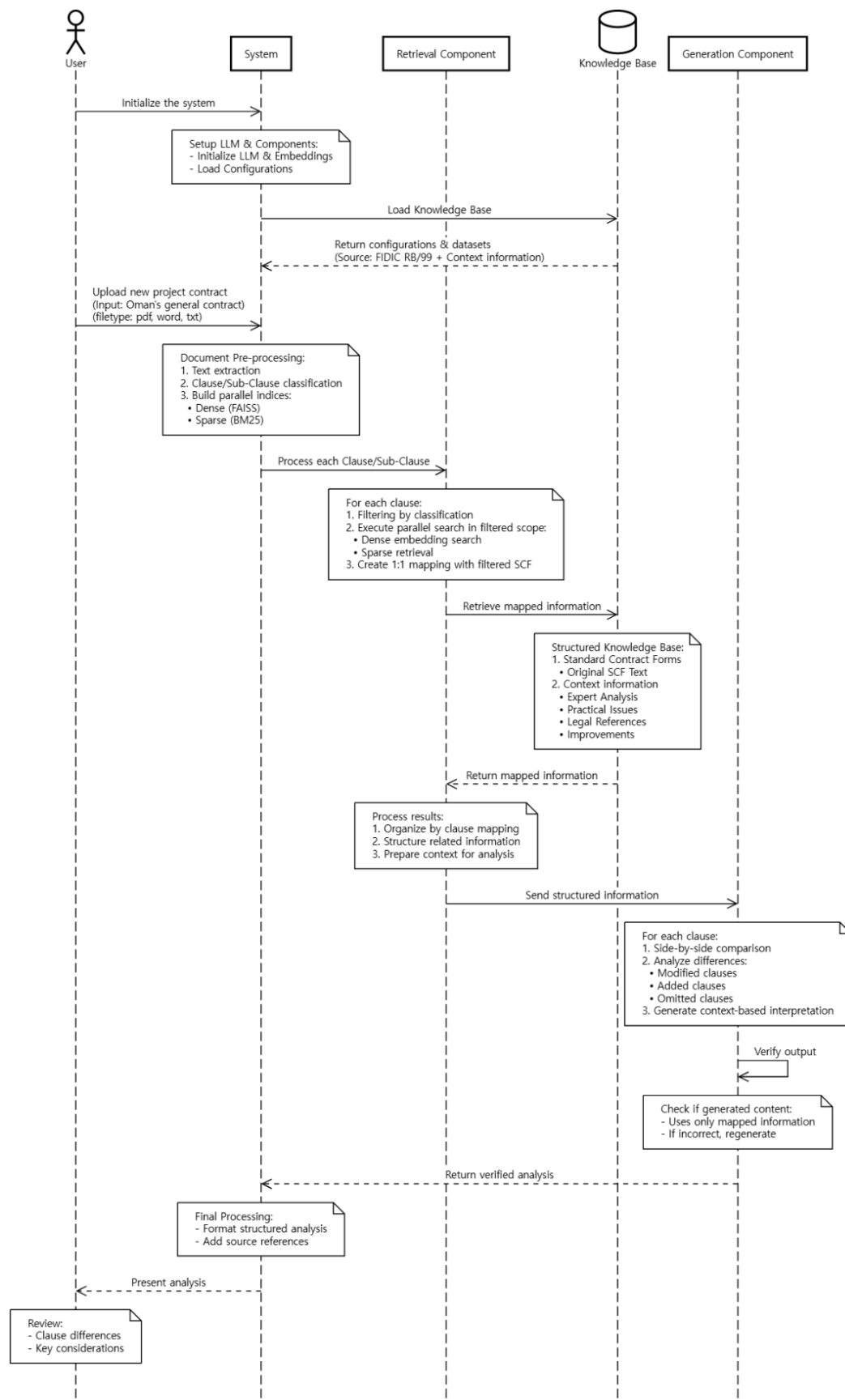


Figure 4. Sequence diagram of proposed method.

Subsequently, the Generation Component compared the input contract clauses with the corresponding SCF clauses to identify differences. It integrated contextual information such as expert analyses and legal considerations to assess the implications of these discrepancies. The system then generated detailed analyses and recommendations, explicitly referencing the relevant sources. Finally, the output was delivered to the user, providing transparent and traceable insights into the deviations from the SCF. By maintaining clear connections to reference sources throughout the analysis, the Generation Component produced outputs that are not only accurate but also verifiable.

3.6. Evaluation Method

In this study, the Retrieval Augmented Generation Assessment System (RAGAS) comprehensively assessed the performance of four approaches: Base LLM, Search-Augmented LLM, Basic RAG, and the proposed method that incorporates structured information. This assessment demonstrates how effectively structured domain knowledge integrates into the RAG for construction contract analysis.

Test cases were designed to evaluate the system's capability to analyze contract clauses by considering multiple aspects, including expert analyses, legal information, and practical considerations. Each test case provides detailed ground truth answers that encompass all these aspects, ensuring that the system's responses can be thoroughly evaluated for completeness and accuracy.

The evaluation method utilized RAGAS metrics, enabling assessment without reliance on ground truth human annotations [36]. These metrics evaluated both retrieval and generation capabilities and were categorized into retrieval performance, response quality, and overall system effectiveness, as detailed in Table 3.

Context Precision measured how relevant the retrieved information is to the query, ensuring that the system focused on pertinent contract clauses without unnecessary content. Context Recall evaluated the completeness of the retrieved information, confirming that all critical clauses needed to answer the query were included.

Answer Relevancy assessed how directly the system's responses address the user's specific queries, ensuring practical usefulness for contract reviewers. Faithfulness measured how accurately the responses reflect information from the SCFs and expert knowledge base, maintaining the integrity of contract interpretation.

Answer Correctness combined accuracy and quality to provide an overall measure of the system's practical utility, determining whether the answers are correct and useful for professional contract review. These metrics assessed the system's ability to retrieve relevant contract information and generate accurate, reliable analyses.

Table 3. Evaluation metrics and description.

Category	Metric	Metric Description	Components	Component Description
Retrieval	Context, Precision	Relevance of retrieved information	Ground Truth, Context	Measures the proportion of relevant context within the retrieved information
	Context Recall	Completeness of retrieved information	Ground Truth, Context	Assesses the proportion of all relevant context successfully retrieved
Generation	Answer Relevancy	How directly the response addresses the query	Answer, Question	Evaluates alignment between the answer and the question for relevance
	Faithfulness	Consistency with retrieved context	Answer, Context	Checks that the answer accurately reflects the retrieved context
Overall	Answer Correctness	Combined accuracy and usefulness of the answer	Answer, Ground Truth	Determines overall answer accuracy and utility against the ground truth

4. Experimental Validation

4.1. Dataset

The test dataset comprises the FIDIC Red Book (1999 version), which is widely used in the industry and Oman’s Standard Contract (2019 version). As a localized adaptation, Oman’s contract was developed based on the FIDIC Red Book 1999. Oman’s contract contains 167 sub-clauses, compared to FIDIC’s 165 sub-clauses. The modifications include the deletion of 10 sub-clauses and the addition of 12 new sub-clauses. Among the 155 maintained sub-clauses, the composition is as follows: 54 major changes, 53 minor modifications, and 46 remained unchanged. The definition section was excluded as they serve as interpretative sources rather than standalone clauses, making them un-suitable for clause-to-clause comparison.

While maintaining a similar structure and conditions to FIDIC Red Book 1999, Oman’s contract includes modifications to reflect local laws and policies, practical requirements, and adjustments to reduce employer’s contractual risks. These modifications include additions or deletions of clauses and changes in sentences, phrases, or terms. Figure 5 illustrates a simple example showing the differences between the FIDIC Red Book and Oman contract, where added conditions in Oman contract are highlighted in green. In the [Labor Laws] clause, Oman’s contract adds conditions for Omanization requirements, stating that the Contractor shall make every effort to employ Omani nationals. Conversely, as an example of deletion, in the [Employer’s Claim] clause, certain conditions regarding claim substantiation and determination procedures were removed.

FIDIC RED(1999)	Oman Standard Contract (2019)
6.4 Labour Laws The Contractor shall comply with all the relevant labour Laws applicable to the Contractor’s Personnel, including Laws relating to their employment, health, safety, welfare, immigration and emigration, and shall allow them all their legal rights. The Contractor shall require his employees to obey all applicable Laws, including those concerning safety at work.	6.4 Labour Laws The Contractor shall comply Added word Laws applicable to the Contractor’s Personnel, including Laws relating to their employment, Omanisation, health, safety, welfare, immigration and emigration, and shall allow them all their legal rights. Added sentence require his employees to obey all applicable Laws, including those concerning The Contractor shall make every effort to employ Omani nationals to replace expatriate employees for the performance of the Works and to apply the Government rules and regulations regarding Omanization / training

Figure 5. Example of modifications in Oman standard contract from FIDIC.

4.2. Experimental Setup

The experimental setup involves comparing four approaches, as detailed in Table 4. A total of 320 test queries were used for each approach for RAGAS evaluation. The Base LLM employs the GPT-4o-mini-2024-07-18 model as a benchmark. The Search-Augmented LLM uses web search using the Perplexity API with the llama-3.1-sonar-large-128k model. The Basic RAG and the proposed method implement a hybrid retrieval system combining dense and sparse retrieval techniques using FAISS and BM25.

All approaches, excluding Perplexity, use the same Base LLM (GPT-4o-mini-2024-07-18) to ensure consistent comparison. This consistent baseline allows for the evaluation of each approach’s effectiveness in contract analysis tasks, demonstrating the advantages of integrating structured domain knowledge into the RAG pipeline. For each query in terms of input information, the Base LLM and Perplexity API only receive the query as input tokens. The Basic RAG receives approximately 4000 tokens per query from unstructured documents, with each corpus containing around 800 tokens. The proposed method uses structured domain knowledge, with input tokens ranging from 1000 to 2000 depending on the varying of knowledge information for each sub-clause.

Table 4. Experimental setup.

Approach	Base Model	Retrieval Configuration
Base LLM	GPT-4o-mini-2024-07-18	N/A
Search-Augmented LLM	Llama-3.1-sonar-large-128k	Web Search (Perplexity API) Top-k: 5
Basic RAG	GPT-4o-mini-2024-07-18	Dense: FAISS, text-embedding-3-small Sparse: BM25, k1 = 1.5, b = 0.75 Hybrid Weights (Dense: 0.5, Sparse: 0.5) Top-k: 5
Proposed Method	GPT-4o-mini-2024-07-18	Dense: FAISS, text-embedding-3-small Sparse: BM25, k1 = 1.5, b = 0.75 Hybrid Weights (Dense: 0.5, Sparse: 0.5) Top-k: 5

4.3. Experimental Results

The experimental results reveal significant performance variations across different model approaches, as Table 5 illustrates. Each approaches shows distinct characteristics in both information retrieval and answer generation capabilities.

Table 5. Performance result of different LLM approaches.

	Evaluation Metrics (Mean $\pm$ SD)					Overall Performance
	Context Metrics		Answer Metrics		Average Score	
Model Type	Context Precision	Context Recall	Answer Relevancy	Faithfulness	Answer Correctness	Average Score
Base LLM (GPT-4o-mini)	N/A	N/A	0.818 $\pm$ 0.312	N/A	0.687 $\pm$ 0.240	0.752 $\pm$ 0.200
Search-Augmented LLM (Perplexity API, llama-3.1-sonar-large)	0.721 $\pm$ 0.297	0.558 $\pm$ 0.392	0.884 $\pm$ 0.093	0.770 $\pm$ 0.252	0.616 $\pm$ 0.205	0.710 $\pm$ 0.134
Basic RAG (GPT-4o-mini, Dense + Sparse Retrieval)	0.587 $\pm$ 0.115	0.635 $\pm$ 0.363	0.895 $\pm$ 0.148	0.602 $\pm$ 0.249	0.697 $\pm$ 0.213	0.683 $\pm$ 0.119
Proposed Method (GPT-4o-mini, Dense + Sparse Retrieval, Structured Information)	0.625 $\pm$ 0.089	0.969 $\pm$ 0.132	0.903 $\pm$ 0.150	0.893 $\pm$ 0.198	0.750 $\pm$ 0.197	0.828 $\pm$ 0.087

4.3.1. Information Retrieval Performance

The Base LLM does not provide context retrieval metrics because it operates without accessing external documents. The Search-Augmented LLM, despite having access to external information, shows moderate retrieval capabilities (Context Precision: 0.721, Recall: 0.558) but with high variability (± 0.297 and ± 0.392 , respectively), indicating stability issues in retrieval performance.

The Basic RAG implementation utilizes a hybrid retrieval approach combining dense and sparse retrieval methods (Context Precision: 0.587 $\pm$ 0.115, Recall: 0.635 $\pm$ 0.363). While this approach enables both the semantic understanding and exact matching of contract terms, its Context Precision is lower than the Search-Augmented LLM, though with improved stability, as shown by lower standard deviations.

The proposed method achieves high Context Recall (0.969 ± 0.132) while exhibiting similar Precision (0.625 ± 0.089) compared to the other methods. This superior performance, particularly in stability as indicated by the lowest standard deviations among all approaches, stems from the structured organization of domain knowledge, enabling the more comprehensive and reliable retrieval of relevant contract information.

4.3.2. Answer Generation Quality

The Base LLM exhibits strong Answer Relevancy (0.818), reflecting its robust language understanding capabilities developed through pre-training. The Search-Augmented LLM shows improved Answer Relevancy (0.884) and Faithfulness (0.770), indicating the effective integration of retrieved information in answer generation. The Basic RAG shows moderate Faithfulness (0.602), indicating the integration of retrieved information in answer generation, though not surpassing the Search-Augmented LLM's performance (0.770).

The proposed method significantly enhances both metrics (Faithfulness: 0.893 ± 0.198 , Answer Relevancy: 0.903 ± 0.150), demonstrating the effective utilization of structured knowledge in generating accurate and relevant responses. Additionally, the comparatively low standard deviations (± 0.198 for Faithfulness and ± 0.150 for Answer Relevancy) indicate consistent response quality across various queries, highlighting the reliability of the proposed method.

Furthermore, the proposed method achieves the highest Answer Correctness of 0.750 ± 0.197 , indicating its superior ability to provide accurate and correct responses essential for construction contract review. The integration of structured information in the knowledge base enhances the quality and organization of the content used for answer generation. This leads to significant improvements in both Faithfulness (from 0.602 to 0.893) and consistency of performance, as shown by lower standard deviations across metrics.

4.3.3. Resource Usage

The resource usage analysis reveals cost differences across the approaches, based on average cost per query. The Base LLM shows the lowest cost at USD 0.0002 per query, while the Search-Augmented LLM using Perplexity API requires the highest at USD 0.0030 per query. The Basic RAG implementation requires USD 0.0005 per query. The proposed method achieves better efficiency than Basic RAG at USD 0.0003 per query. This demonstrates that the integration of structured domain knowledge can enhance cost efficiency by selectively utilizing relevant information.

5. Discussion

5.1. Performance Analysis

By fulfilling the three requirements outlined in Section 3.1, the structured knowledge-integrated RAG pipeline effectively addresses significant challenges in automated contract analysis for construction contracts. Firstly, the systematic integration of expert knowledge through a well-organized knowledge structure enables the system to handle specialized technical and legal tasks with greater accuracy. By incorporating domain-specific expertise into a knowledge base, the system mitigates the limitations of LLMs in dealing with specialized knowledge not widely represented in their training data, ensuring the precise interpretation of contract clauses.

Secondly, the incorporation of precise retrieval mechanisms addresses challenges in handling large volumes of unstructured information and maintaining consistent interpretations. By combining classification-based filtering with hybrid dense-sparse retrieval

techniques, the system mitigates limitations of basic RAG approaches, improving the retrieval of relevant clauses and contextual information.

Thirdly, maintaining connections to reference sources enhances traceability, mitigating the probabilistic nature of LLM outputs and ensuring the reliability. Through reference-based generation, the system allows users to trace conclusions back to original sources, strengthening trustworthiness where accuracy is paramount.

5.2. Performance Enhancements via Structured Information Integration

The proposed method achieved Context Recall of 0.969 and Faithfulness of 0.893, showing significant improvements over Basic RAG's performance (Context Recall: 0.635, Faithfulness: 0.602). This represents a 52.6% improvement in Context Recall and 48.3% improvement in Faithfulness compared to the Basic RAG approach. These substantial improvements demonstrate enhanced reference maintenance and reduced hallucinations through structured knowledge integration.

Figure 6 illustrates the comparative analysis of responses to "What measures must the Contractor take to protect the environment?" across different models. The Base LLM offered only general environmental protection practices without addressing specific contractual requirements. It missed critical obligations like taking "all necessary measures," limiting pollution and noise damage, controlling emissions, and complying with the environmental impact statement.

Both the Search-Augmented LLM and Basic RAG included key contractual obligations but diluted their responses with irrelevant information about local regulations, monitoring systems, and legal consequences. This additional content reduced the precision required in contractual contexts. Moreover, both models exhibited variability in performance metrics. The Search-Augmented LLM demonstrated even more pronounced variability, with a Context Precision standard deviation of ± 0.297 and Context Recall standard deviation of ± 0.392 . This high variability undermines the reliability essential for tasks that require consistent precision, such as contract analysis.

The proposed method often generated responses closely matching the contractual content. It specified that the Contractor must take all necessary measures to protect the environment both on and off site, limit damage from pollution and noise, ensure that emissions do not exceed specified values, and comply with environmental impact statements. This demonstrates the effectiveness of integrating structured information with advanced retrieval techniques for contractual content, enhancing consistency and maintaining high recall.

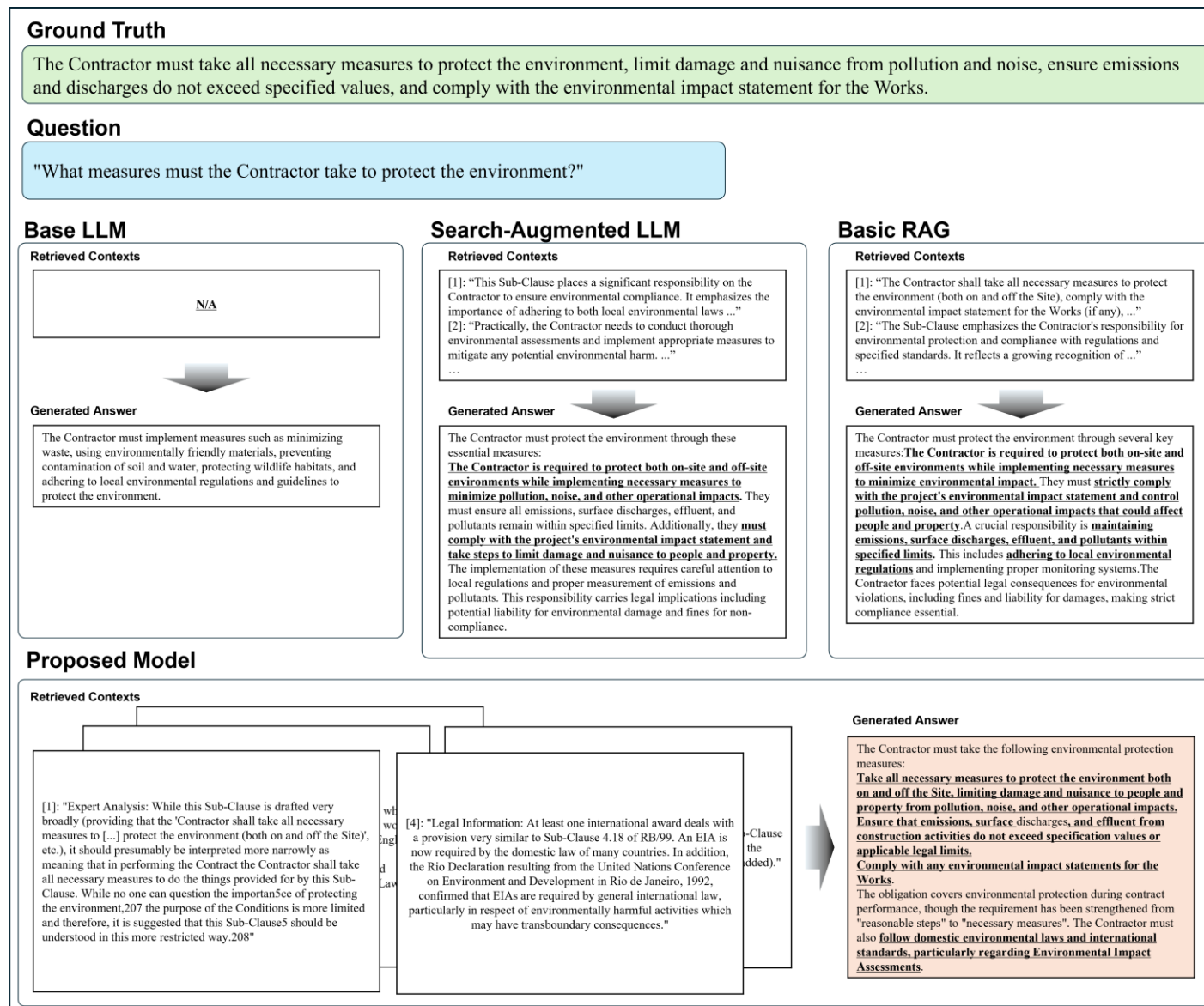


Figure 6. Comparison of retrieved contexts and generated answers by different models.

5.3. Limitations in Answer Generation

The system's answer generation capabilities show fundamental limitations despite high overall performance (Average Score: 0.828 ± 0.087). A critical finding is that these limitations occur during the answer generation phase, even when accurate information exists within the source context. The high standard deviations in Answer Correctness (0.750 ± 0.197) and Faithfulness (0.893 ± 0.198) metrics indicate significant inconsistency in response quality, revealing three issues.

The oversimplification shows a tendency across the analyzed cases. This extends to the generalization of specific numerical values and procedural details, indicating a limitation in maintaining professional domain precision. For example, when the context clearly specified "liquidated damages at a rate of 0.1% per day up to a maximum of 10% of the contract price", the generated answer simply stated "financial penalties for delays" losing critical numerical precision.

Critical information omission manifests through the consistent absence of contractual context and conditional limitations. This reveals the system's difficulty in expert-level judgment of information significance. For instance, while source materials explicitly detailed prerequisite conditions for insurance claims, including notification periods and documentation requirements, the generated response omitted these crucial preconditions.

Professional reasoning failure presents the most significant challenge, particularly in processing complex causal relationships. The system fails to capture the interconnected nature of contractual rights and obligations. For example, in Sub-clause 16.1, the system fails to recognize the logical sequence of Engineer's payment certificate failure → 21-day notice → work suspension → time extension/cost compensation claims, demonstrating inability to comprehend the logical flow and relationships. This failure stems from two key factors: oversimplification reduces the complex contractual procedures to disconnected events, losing the nuanced conditions that link each step, while critical information omission removes key contextual elements about how these rights operate differently.

To address these issues, several fundamental improvements are necessary. First, the development of expert reasoning structures capable of complex context integration is needed. Second, the system must enhance its ability to process multiple context layers simultaneously. Third, mechanisms to preserve professional detail precision need to be implemented, rather than resorting to oversimplification. While simplification is necessary, it should be performed with expert-level judgment to maintain essential professional details. These improvements require fundamental reconsideration of how LLM systems process and integrate professional domain knowledge, beyond simply the expansion of knowledge bases.

5.4. Limitations

The system faces several structural limitations beyond the answer generation issues discussed in Section 5.3. First, the system's effectiveness is fundamentally constrained by its reliance on a specific knowledge base for SCFs, limiting its applicability to contracts outside this scope, particularly for customized or significantly modified contracts.

Second, while RAGAS evaluation provides quantitative insights, it cannot fully capture the nuanced judgments required in professional contract analysis, emphasizing the need for comprehensive human expert evaluations.

Third, while the current study used fixed weights (0.5, 0.5) for dense and sparse retrieval to maintain experimental consistency, future research could explore the following enhancement approaches. Pre-retrieval improvements include query refinement techniques and weight optimization. Post-retrieval document enhancement methods increase the utility of retrieved documents for generation tasks [37].

Lastly, this study focused on clause-by-clause comparison, which does not fully reflect industrial requirements. This limitation arises because comparison target contracts may have significantly different structures from SCFs, and contract analysis requires analyzing cross-references simultaneously. Future research needs to consider the dependencies and logic between clauses more comprehensively.

6. Conclusions

Construction contract review poses significant challenges due to the necessity for specialized technical and legal expertise and the limitations of existing AI applications in handling domain-specific knowledge. This study addressed these challenges through a structured knowledge-integrated RAG pipeline, focusing on the integration of structured domain expertise with LLM. The proposed system demonstrated significant improvements through three components.

First, the implementation of classification-based filtering enhanced retrieval accuracy by narrowing the search scope before executing dense and sparse searches. Second, the structured organization of domain knowledge—encompassing expert analyses, legal information, and practical considerations—enabled comprehensive contract analysis. Third, the reference-based generation process, ensuring content alignment with structured knowledge, produces reliable outputs and enables traceability.

The research hypothesis regarding the structured knowledge-integrated RAG pipeline's superior performance in construction contract review was conclusively validated through experimental results. The practical evaluation using Oman's Standard Conditions demonstrated that the system significantly outperformed basic RAG approaches, achieving a Context Recall of 0.969 and Faithfulness of 0.893—representing substantial improvements of 52.6% and 48.3%, respectively. Furthermore, the system achieved the highest average score with stable performance (0.828 ± 0.087) among all tested models, indicating consistent and reliable outputs. The evaluation through RAGAS confirmed the system's capability in handling complex reasoning and multi-context understanding.

Despite these promising results, several limitations require attention. The system faces challenges in answer generation, including the oversimplification of contractual details, omission of critical information, and failure in professional reasoning. The system's structural limitations include its dependence on SCF-based knowledge, which restricts applicability to non-standard contracts, and the inherent limitations of RAGAS evaluation in capturing nuanced professional judgments. Future work should focus on developing enhanced expert reasoning structures, improving multi-context processing capabilities, and expanding the knowledge base beyond SCFs. Additionally, the optimization of retrieval weights and implementation of pre/post-retrieval enhancement methods could further improve system performance.

Theoretically, it establishes a systematic method to transform expert knowledge into a structured SCF-based format and introduces an enhanced two-stage retrieval mechanism. Practically, it provides an empirically evaluated framework for analyzing modified clauses in localized contracts with traceable results. These innovations improve the reliability and accuracy of automated contract review in the construction industry.

Author Contributions: Conceptualization, E.W.K. and K.J.K.; formal analysis, Y.J.S.; investigation, E.W.K.; methodology, E.W.K.; project administration, S.K.; software, Y.J.S.; supervision, S.K.; visualization, Y.J.S.; writing—original draft, E.W.K.; writing—review and editing, K.J.K.; resources, S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (No. 2019R1A2C2010794).

Data Availability Statement: The datasets presented in this article are not readily available because of copyright restrictions on each book and contract form.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of this study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

AI	Artificial Intelligence
LLM	Large Language Model
NLP	Natural Language Processing
RAG	Retrieval-Augmented Generation
RAGAS	Retrieval-Augmented Generation Assessment System
SCF	Standard Contract Form

References

1. Arcadis. *Global Construction Disputes Report 2022: Adapting to a New Normal*; Arcadis: Amsterdam, The Netherlands, 2022.
2. Arcadis. *13th Annual Construction Disputes Report North America: Embracing Change Moving Forward*; Arcadis: Amsterdam, The Netherlands, 2023.
3. Seppälä, C. *The FIDIC Red Book Contract: An International Clause-by-Clause Commentary*, 1st ed.; Kluwer Law International B.V.: Alphen aan den Rijn, The Netherlands, 2023; Chapter 4.
4. Dmaidi, N.; Dwaikat, M.; Shweiki, I. Construction Contracting Management Obstacles in Palestine. *Int. J. Constr. Eng. Manag.* **2013**, *2*, 15–22.
5. Baker, E.; Mellors, B.; Chalmers, S.; Lavers, A. *FIDIC Contracts: Law and Practice*, 1st ed.; Informa Law from Routledge: London, UK, 2013; p. 30.
6. Rasslan, N.D.; Nassar, A.H. Comparing Suitability of NEC and FIDIC Contracts in Managing Construction Project in Egypt. *Int. J. Eng. Res. Technol.* **2017**, *6*, 531–535.
7. Hamie, J.M.; Abdul-Malak, M.U. Model Language for Specifying the Construction Contract's Order-of-Precedence Clause. *J. Leg. Aff. Dispute Resolut. Eng. Constr.* **2018**, *10*, 04518011. [https://doi.org/10.1061/\(ASCE\)LA.1943-4170.0000260](https://doi.org/10.1061/(ASCE)LA.1943-4170.0000260).
8. Lee, J.; Yi, J.; Son, J. Development of Automatic-Extraction Model of Poisonous Clauses in International Construction Contracts Using Rule-Based NLP. *J. Comput. Civ. Eng.* **2019**, *33*, 04019003. [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000807](https://doi.org/10.1061/(asce)cp.1943-5487.0000807).
9. Serag, E.; Osman, H.; Ghanem, M. Semantic Detection of Risks and Conflicts in Construction Contracts. In Proceedings of the CIB W78 2010: 27th International Conference, Cairo, Egypt, 16–18 November 2010.
10. Xu, X.; Cai, H. *Semantic Frame-Based Information Extraction from Utility Regulatory Documents to Support Compliance Checking*; Springer International Publishing AG: Zurich, Switzerland, 2018; pp. 223–230.
11. Zhang, J.; El-Gohary, N.M. Semantic NLP-Based Information Extraction from Construction Regulatory Documents for Automated Compliance Checking. *J. Comput. Civ. Eng.* **2016**, *30*, 04015014. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000346](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000346).
12. Faraji, A.; Rashidi, M.; Perera, S. Text Mining Risk Assessment-Based Model to Conduct Uncertainty Analysis of the General Conditions of Contract in Housing Construction Projects: Case Study of the NSW GC21. *J. Archit. Eng.* **2021**, *27*, 04021025. [https://doi.org/10.1061/\(ASCE\)AE.1943-5568.0000489](https://doi.org/10.1061/(ASCE)AE.1943-5568.0000489).
13. Guo, D.; Onstein, E.; Rosa, A.D.L. A Semantic Approach for Automated Rule Compliance Checking in Construction Industry. *IEEE Access* **2021**, *9*, 129648–129660. <https://doi.org/10.1109/ACCESS.2021.3108226>.
14. Arditi, D.; Pulket, T. Predicting the Outcome of Construction Litigation Using Boosted Decision Trees. *J. Comput. Civ. Eng.* **2005**, *19*, 387–393. [https://doi.org/10.1061/\(ASCE\)0887-3801\(2005\)19:4\(387\)](https://doi.org/10.1061/(ASCE)0887-3801(2005)19:4(387)).
15. Candaş, A.B.; Tokdemir, O.B. Automating Coordination Efforts for Reviewing Construction Contracts with Multilabel Text Classification. *J. Constr. Eng. Manag.* **2022**, *148*, 04022027. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002275](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002275).

16. Hassan, F.u.; Le, T.; Tran, D. Multi-Class Categorization of Design-Build Contract Requirements Using Text Mining and Natural Language Processing Techniques. In Proceedings of the Construction Research Congress 2020: Project Management and Controls, Materials, and Contract, Tempe, AZ, USA, 8–10 March 2020.
17. Salama, D.M.; El-Gohary, N.M. Semantic Text Classification for Supporting Automated Compliance Checking in Construction. *J. Comput. Civ. Eng.* **2016**, *30*, 04014106. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000301](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000301).
18. Choi, S.; Choi, S.; Kim, J.; Lee, E. AI and Text-Mining Applications for Analyzing Contractor's Risk in Invitation to Bid (ITB) and Contracts for Engineering Procurement and Construction (EPC) Projects. *Energies* **2021**, *14*, 4632. <https://doi.org/10.3390/en14154632>.
19. Hassan, F.u.; Le, T. Automated Requirements Identification from Construction Contract Documents Using Natural Language Processing. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2020**, *12*, 04520009. [https://doi.org/10.1061/\(ASCE\)LA.1943-4170.0000379](https://doi.org/10.1061/(ASCE)LA.1943-4170.0000379).
20. Moon, S.; Chi, S.; Im, S. Automated detection of contractual risk clauses from construction specifications using bidirectional encoder representations from transformers (BERT). *Autom. Constr.* **2022**, *142*, 104465. <https://doi.org/10.1016/j.autcon.2022.104465>.
21. Xue, X.; Hou, Y.; Zhang, J. Automated Construction Contract Summarization Using Natural Language Processing and Deep Learning. In Proceedings of the 39th International Symposium on Automation and Robotics in Construction (ISARC 2022), Bogota, Columbia, 13–15 July 2022.
22. Zheng, Z.; Lu, X.; Chen, K.; Zhou, Y.; Lin, J. Pretrained domain-specific language model for natural language processing tasks in the AEC domain. *Comput. Ind.* **2022**, *142*, 103733. <https://doi.org/10.1016/j.compind.2022.103733>.
23. Zhou, H.; Gao, B.; Tang, S.; Li, B.; Wang, S. Intelligent detection on construction project contract missing clauses based on deep learning and NLP. *Eng. Constr. Archit. Manag.* **2023**, *ahead-of-print*. <https://doi.org/10.1108/ecam-02-2023-0172>.
24. Ko, T.; Lee, J.; Jeong, H.D. Project Requirements Prioritization through NLP-Driven Classification and Adjusted Work Items Analysis. *J. Constr. Eng. Manag.* **2024**, *150*, 04023171. <https://doi.org/10.1061/JCEMD4.COENG-13655>.
25. Pham, H.T.T.L.; Han, S. Natural Language Processing with Multitask Classification for Semantic Prediction of Risk-Handling Actions in Construction Contracts. *J. Comput. Civ. Eng.* **2023**, *37*, 04023027. <https://doi.org/10.1061/jccee5.cpeng-5218>.
26. Qi, X.; Chen, Y.; Lai, J.; Meng, F. Multifunctional Analysis of Construction Contracts Using a Machine Learning Approach. *J. Manag. Eng.* **2024**, *40*, 04024002. <https://doi.org/10.1061/JMENE4.MEENG-5604>.
27. Zhong, B.; Shen, L.; Pan, X.; Zhong, X.; He, W. Dispute Classification and Analysis: Deep Learning-Based Text Mining for Construction Contract Management. *J. Constr. Eng. Manag.* **2024**, *150*, 04023151. <https://doi.org/10.1061/jcemd4.coeng-14080>.
28. Wong, S.; Yang, J.; Zheng, C.; Su, X. Domain Ontology Development Methodology for Construction Contract. In Proceedings of the 27th International Symposium on Advancement of Construction Management and Real Estate, Hong Kong, China, 5–6 December 2022.
29. Zheng, C.; Wong, S.; Su, X.; Tang, Y. A knowledge representation approach for construction contract knowledge modeling. *arXiv* **2023**, arXiv:2309.12132.
30. Mallen, A.; Asai, A.; Zhong, V.; Das, R.; Khashabi, D.; Hajishirzi, H. When Not to Trust Language Models: Investigating Effectiveness of Parametric and Non-Parametric Memories. *arXiv* **2022**, arXiv:2212.10511.
31. Kandpal, N.; Deng, H.; Roberts, A.; Wallace, E.; Raffel, C. Large Language Models Struggle to Learn Long-Tail Knowledge. *arXiv* **2022**, arXiv:2211.08411.
32. Wang, Z.; Zhang, H.; Li, C.; Eisenschlos, J.M.; Perot, V.; Wang, Z.; Miculicich, L.; Fujii, Y.; Shang, J.; Lee, C.; et al. Chain-of-Table: Evolving Tables in the Reasoning Chain for Table Understanding. *arXiv* **2024**, arXiv:2401.04398.
33. Robinson, M.D. *An Employer's and Engineer's Guide to the FIDIC Conditions of Contract*, 1st ed.; Wiley-Blackwell: Newark, NJ, USA, 2013; Chapter 2.
34. Grose, M. *Construction Law in the United Arab Emirates and the Gulf*, 1st ed.; John Wiley & Sons: Hoboken, NJ, USA, 2016; pp. 293–385.
35. Luan, Y.; Eisenstein, J.; Toutanova, K.; Collins, M. Sparse, Dense, and Attentional Representations for Text Retrieval. *Trans. Assoc. Comput. Linguist.* **2021**, *9*, 329–345. https://doi.org/10.1162/tacl_a_00369.

36. Es, S.; James, J.; Espinosa-Anke, L.; Schockaert, S. RAGAS: Automated Evaluation of Retrieval Augmented Generation. *arXiv* **2023**, arXiv:2309.15217.
37. Han, B.; Susnjak, T.; Mathrani, A. Automating Systematic Literature Reviews with Retrieval-Augmented Generation: A Comprehensive Overview. *Appl. Sci.* **2024**, *14*, 9103. <https://doi.org/10.3390/app14199103>.

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.