Multi-View Contract Graph: A Knowledge Graph-Based Retrieval Framework for Contract Information Extraction

Momojit Biswas
TCS Research, India
momojit.biswas@tcs.com

Preethu Rose Anish
TCS Research, India
preethu.rose@tcs.com

Abstract

Legal contract understanding often requires connecting information distributed across multiple clauses and cross-referenced provisions. While Retrieval-Augmented Generation (RAG) grounds Large Language Models (LLMs) in relevant evidence, conventional retrieval primarily prioritizes direct query relevance and may overlook supporting contractual context. We propose the Multi-View Contract Graph (MCG), a contract-specific knowledge graph that models semantic, entity, keyword, and structural relationships among clauses. By traversing these relationships, MCG retrieves supporting evidence that may not directly match the query but is necessary for accurate interpretation. We evaluate MCG on the Contract Understanding Atticus Dataset (CUAD), covering 15 information categories, including audit rights, governing laws, and liability-related provisions. Across five open-source LLMs, MCG consistently outperforms traditional retrieval, achieving a best F_1 score of 0.460 with Mistral-7B-Instruct-v0.3. Compared with BM25-RAG and ColBERT-RAG, the best MCG configuration improves F_1 by approximately 10.0% and 12.2%, respectively. These results demonstrate that modeling clause-level relationships improves retrieval and information extraction for legal contracts.

1 Introduction

Legal contracts are fundamental to business operations, governing commercial relationships, regulatory compliance, procurement, licensing, employment, and risk management (Junior et al., 2026; Khan, 2022). As organizations manage growing volumes of increasingly complex contracts, automated contract understanding has become essential for contract governance, legal analytics, and contract lifecycle management (Liu et al., 2025). Despite rapid advances in LLMs, contract understanding remains challenging because critical information is often distributed across multiple clauses and sections within a contract (Singh et al., 2025, 2026). Interpreting a contractual provision frequently requires supporting context found elsewhere in the agreement, such as related obligations, defined terms, dates, or notice related requirements (Wang et al., 2025). For example, a termination clause may depend on a notice period defined elsewhere, while a licensing provision may rely on distant definitions or exclusions. Thus, understanding contractual intent often requires reasoning across interconnected clauses rather than analyzing provisions in isolation.

Retrieval-Augmented Generation (RAG) has emerged as a widely adopted approach for grounding LLMs through lexical and semantic retrieval. However, contract understanding remains challenging because relevant evidence is often distributed across non-adjacent clauses, cross-referenced provisions, and structurally related sections (Liu et al., 2025). Conventional retrieval methods primarily optimize query relevance and may overlook supporting provisions needed for accurate interpretation. Such evidence may be linked through definitions, obligations, temporal conditions, cross-references, and recurring contractual concepts. Existing knowledge graph approaches have demonstrated the value of structured relationships for legal retrieval and reasoning, but largely focus on relationships among legal entities, statutes, cases, and legal concepts. In contrast, contract analysis requires modeling heterogeneous dependencies among clauses within the same document. To address this gap, we propose the Multi-View Contract Graph (MCG), a contract-specific knowledge graph that integrates semantic, entity-level, keyword-based, and structural relationships among clauses. Given a query, MCG first identifies relevant anchor clauses and then expands retrieval through graph-connected clauses to recover supporting evidence that may be missed by similarity-based retrieval. The resulting evidence is provided

Proceedings of the 3rd Workshop on KNOwledge GRaphs & Agentic Systems Interplay (NORA 2026), pages 53–62
November 10, 2026 ©2026 Association for Computational Linguistics

to an LLM for contract question answering and information extraction.

We evaluate MCG on contracts from the Contract Understanding Atticus Dataset (CUAD) (Hendrycks et al., 2021), selecting 15 categories from the original 41 annotations to cover diverse contractual information, including dates, parties, governing law, licensing, renewal, audit rights, and liability provisions. Definitions of the selected categories are provided in Table 1. While CUAD formulates contract review as an extractive span-detection task, our setting reformulates it as retrieval-augmented contract information extraction, where relevant and supporting evidence must first be retrieved and then interpreted to answer category-specific questions. This enables evaluation of whether retrieval methods can recover evidence distributed across multiple clauses rather than simply identify a predefined span. We compare MCG with extractive QA, BM25-RAG (Robertson and Zaragoza, 2009), ColBERT-RAG (Khattab and Zaharia, 2020), and Dense-Retrieval-RAG (Lewis et al., 2020) across five open-source LLMs (Mistral-7B-Instruct-v0.2 (Jiang et al., 2023), Mistral-7B-Instruct-v0.3 (Jiang et al., 2023), Llama-3-8B-Instruct (Grattafiori et al., 2024), Saul-7B-Instruct-v1 (Colombo et al., 2024), and Qwen-7B-Chat (Bai et al., 2023)). MCG consistently achieves the strongest performance, with the best configuration attaining an F_1 score of 0.460 using Mistral-7B-Instruct-v0.3, improving over BM25-RAG and ColBERT-RAG by approximately 10.0% and 12.2%, respectively.

The contributions of this paper are threefold: (1) We propose a graph-enhanced retrieval strategy that extends retrieval beyond directly relevant clauses by leveraging weighted inter-clause relationships to recover supporting contractual context. (2) We introduce the Multi-View Contract Graph (MCG), which integrates semantic, entity-level, keyword-based, and structural relationships to model dependencies among contract clauses. (3) We demonstrate through evaluation on CUAD across 15 contract information categories and five open-source LLMs that MCG consistently outperforms extractive QA and multiple retrieval-based baselines.

2 Literature Review

Knowledge Graphs (KGs) have been widely explored for representing legal knowledge and supporting legal information retrieval, reasoning, and question answering. Early efforts focused on constructing structured legal knowledge bases from case law and statutory documents. For example, dAmato et al. (2025) developed an ontology-driven knowledge graph for European Court of Human Rights, while Jain et al. (2022) proposed an RDF-based legal knowledge graph for Indian Supreme Court judgments. Similarly, Sovrano et al. (2020) introduced a knowledge extraction framework that leveraged structured legal representations to support knowledge graph-driven legal question answering. These studies demonstrated the value of explicit knowledge representations for enhancing legal search and reasoning.

Subsequent research investigated the integration of knowledge graphs with legal text processing and semantic retrieval. Filtz et al. (2021) proposed a linked legal data framework connecting legal resources across jurisdictions through standards such as ELI and ECLI. The Lynx platform (Schneider et al., 2022) further combines legal knowledge graphs with multilingual NLP services, including entity linking, temporal analysis, semantic similarity computation, question answering, and summarization. These efforts highlighted the benefits of coupling structured knowledge with natural language processing capabilities, while also exposing practical challenges related to annotation costs, extraction complexity, and the limited availability of large-scale legal evaluation benchmarks.

More recently, knowledge graphs have been integrated with retrieval-augmented generations and large language models. Calvin Barron et al. (2025) proposed a legal KG-RAG architecture that combines vector retrieval, knowledge graphs, and topic modeling to support grounded legal question answering. Chen et al. (2026) introduced the Case-Enhanced Law Article Knowledge Graph (CLAKG), which combines statutory provisions, case law, graph embeddings, and constrained LLM for legal recommendation tasks. Similarly, Li et al. (2024) explored knowledge-enhanced LLMs for legal knowledge extraction, while Pham et al. (2026) demonstrated the effectiveness of graph-based reasoning through explicit modeling of legal concepts and inference rules. Collectively, these approaches illustrate a growing trend toward combining structured knowledge representations, semantic retrieval mechanisms, and LLMs to improve legal-AI systems.

Despite these advances, contract understanding

Table 1: Contract Information Extraction Categories Extracted from CUAD Contracts.

Category	Definition
Agreement Date	The date on which the agreement was executed or signed by the contracting parties.
Audit Rights	A contractual provision granting one party the right to inspect, verify, or review the records, processes, or compliance activities of another party.
Cap on Liability	A clause specifying the maximum amount of financial liability or damages that a party may be required to pay under the agreement.
Effective Date	The date on which the rights, obligations, and terms of the agreement become legally effective.
Exclusivity	A provision granting exclusive rights or restricting parties from entering similar agreements with competitors or third parties.
Expiration Date-Answer	The specific calendar date stated in the agreement on which the contract expires or terminates.
Expiration Date	A clause describing the conditions or timing under which the agreement expires or ceases to remain in effect.
Governing Law	The jurisdiction or legal framework whose laws govern the interpretation and enforcement of the agreement.
Insurance	A contractual requirement that one or more parties maintain specified insurance coverage during the term of the agreement.
License Grant	A provision defining the rights, permissions, and scope of use granted by one party to another over intellectual property, software, or other assets.
Non-Compete	A clause restricting a party from engaging in activities that compete with the business interests of another party during or after the agreement term.
Notice Period to Terminate Renewal	The advance notice required to prevent or terminate an automatic contract renewal.
Parties	The legal entities or individuals entering into the agreement, including any defined names or aliases used within the contract.
Renewal Term	The duration and conditions under which the agreement may be renewed or extended beyond its initial term.
Uncapped Liability	A provision specifying obligations or circumstances for which liability is not subject to any contractual monetary limit.

presents challenges that differ substantially from those encountered in case of law retrieval, statutory reasoning, or legal recommendation. Existing legal knowledge graph systems primarily model relationships among legal entities, cases, statutes, and legal concepts. In contrast, contract analysis often requires identifying dependencies among clauses within the same document, where critical contextual evidence may be distributed across non-adjacent sections connected through definitions, obligations, exceptions, temporal conditions, and structural references. Conventional retrieval methods generally prioritize direct query relevance and may therefore fail to recover supporting contractual context necessary for accurate interpretation. Moreover, existing KG-RAG approaches typically emphasize document-level or concept-level relationships rather than fine-grained clause-level dependencies.

To address this gap, we propose the Multi-View Contract Graph (MCG), a contract-specific retrieval framework that models clause-level relationships using multiple complementary views, including semantic similarity, named entities, contractual keywords, and document structure. By combining these heterogeneous relationship types within a unified graph representation, MCG enables retrieval to extend beyond directly relevant clauses and recover supporting contractual evidence required for downstream contract question answering and information extraction.

3 Methodology

3.1 Overview

We propose MCG, a contract-specific knowledge graph and graph-based retrieval framework for contract question answering and information extraction. The key motivation is that contractual information is often distributed across multiple clauses that may be related through shared entities, contractual concepts, or document structure. Conventional retrieval methods typically rank text chunks according to their direct relevance to a query and may therefore fail to recover supporting clauses required for complete contractual interpretation.

MCG addresses this limitation by representing each contract as a contract-specific knowledge graph that serves as a structured memory representation of contractual knowledge, whose nodes correspond to semantically coherent contract

chunks. Each node is enriched with three complementary views of contractual content: dense semantic embeddings, named entities, and contractual keywords. Relationships between nodes are modeled using semantic similarity, entity overlap, keyword overlap, and positional proximity within the document. At query time, MCG first identifies relevant clauses through semantic matching and subsequently expands retrieval through graph-connected clause expansion to recover supporting context. The retrieved clauses, together with their associated metadata, are then provided to a downstream LLM for answer generation. The framework consists of two stages: (i) contract preprocessing and graph construction, (ii) graph-based retrieval and LLM-based answer generation. An overview of the proposed method is illustrated in Figure 1.

3.2 Multi-View Contract Graph Construction

Given a contract document D, MCG first constructs a weighted graph representation of the document. Text is extracted from PDF files and normalized through Unicode standardization, correction of broken hyphenation, removal of page-level artifacts, whitespace normalization, and filtering of non-standard characters.

Unlike fixed-size chunking approaches, MCG employs semantic chunking to generate semantically coherent contract segments. The document is first divided into individual sentences, and each sentence is encoded using the all-MiniLM-L6-v2 (Wang et al., 2021) Sentence Transformer due to its favorable trade-off between retrieval quality and computational efficiency. The proposed MCG framework is model-agnostic and does not depend on a specific embedding architecture. Alternative embedding models, including domain-specific encoders such as Legal-BERT (Chalkidis et al., 2020) and Contract-BERT (Zheng et al., 2021), as well as retrieval-oriented models such as E5 (Wang et al., 2022) and BGE (Xiao et al., 2024), can be incorporated without modifying the graph construction or retrieval procedures. For two consecutive sentences with embeddings s_i and s_{i+1}, semantic continuity is measured using cosine similarity: $\mathrm{Sim}(s_i, s_{i+1}) = \frac{s_i \cdot s_{i+1}}{\|s_i\| \times \|s_{i+1}\|}$.

A chunk boundary is introduced whenever the cosine similarity between adjacent sentences falls below 0.80, provided that the current chunk satisfies a minimum length of 400 tokens. The similarity threshold was selected empirically to balance semantic coherence and chunk granularity. Lower thresholds tended to merge distinct contractual provisions into larger chunks, whereas higher thresholds often produced fragmented chunks with insufficient context. Chunk sizes are constrained between 400 and 500 tokens to provide sufficient contractual context while remaining within the effective retrieval and context-processing limits of downstream LLMs. In addition, the final two sentences of each chunk are preserved to overlap with the subsequent chunk to maintain contextual continuity across chunk boundaries. The resulting process generates a sequence of semantically coherent contract chunks, $C = \{c_1, c_2, \ldots, c_n\}$, which subsequently serve as the nodes of the graph.

Each chunk is enriched with three complementary representations. First, a dense semantic embedding h_i is generated using the all-MiniLM-L6-v2 Sentence Transformer model. Second, named entities are extracted using the Flair OntoNotes Fast NER model (Akbik et al., 2019). The model identifies diverse entity categories, including persons, organizations, locations, legal references, dates, temporal expressions, and numerical entities. The resulting entity set for chunk c_i is denoted by E_i. Third, contractual keywords are extracted using KeyBERT (Grootendorst, 2020) with unigram phrases (n-gram range = 1). For each chunk, the top-ranked keywords are retained to represent salient contractual concepts. These keywords capture important contractual topics such as confidentiality, indemnification, assignment, renewal, insurance, and limitation of liability. The resulting keyword set is denoted by K_i.

Each graph node is therefore represented as: $v_i = (c_i, h_i, E_i, K_i)$, where c_i denotes the chunk text, h_i its semantic embedding, E_i its entity set, and K_i its keyword set. The contract is modeled as a contract-specific knowledge graph $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_n\}$ represents semantically coherent contract chunks and E captures the semantic, entity-level, conceptual, and structural relationships between them. In the proposed formulation, each chunk is connected to every other chunk within the contract to preserve potential semantic, entity-level, conceptual, and structural dependencies among contractual provisions. This design choice is motivated by the observation that contractually related provisions may be distributed across distant sections of a document and are not necessarily linked through local structure or explicit references. By preserving all pairwise clause relationships, the graph captures the full spectrum

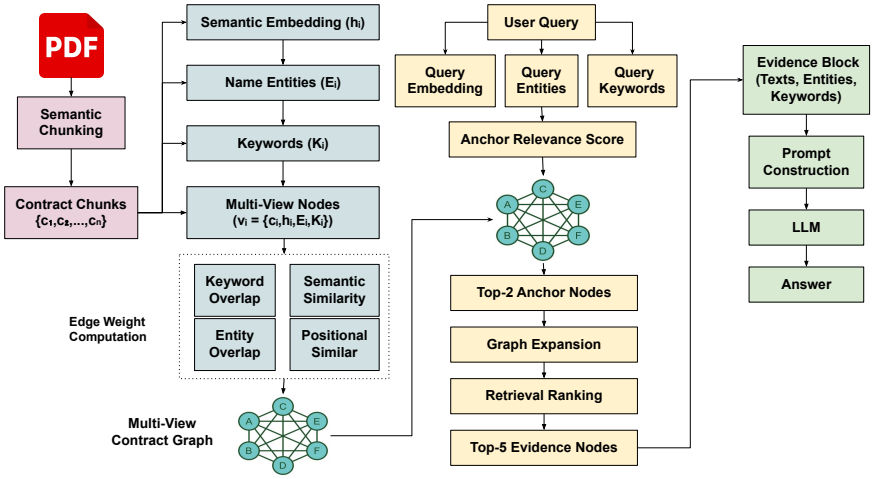


Figure 1: Overview of the proposed Multi-View Contract Graph (MCG).

of potential semantic, entity-level, conceptual, and structural interactions within the contract. Consequently, potentially relevant supporting clauses are not excluded during graph construction, allowing retrieval to leverage both directly relevant clauses and related supporting context. The strength of the relationship between two nodes is determined using four complementary signals. Semantic similarity is computed as: $S_{\text{sem}}(i,j) = \frac{h_i \cdot h_j}{\|h_i\| \times \|h_j\|}$. Entity overlap is measured using Jaccard similarity: $S_{\text{ent}}(i,j) = \frac{|E_i \cap E_j|}{|E_i \cup E_j|}$. Keyword overlap is similarly computed as: $S_{\text{key}}(i,j) = \frac{|K_i \cap K_j|}{|K_i \cup K_j|}$. To capture structural proximity within the document, positional similarity is defined as: $S_{\text{pos}}(i,j) = \frac{1}{|i-j|}$.

Since graph edges are constructed only between distinct nodes ($i \neq j$), self-connections are not included in the graph and the denominator never becomes zero. Consequently, the positional similarity measure remains well-defined for all valid node pairs.

The four signals are combined into a single edge weight: $EW(i,j) = \alpha \cdot S_{\text{sem}}(i,j) + \beta \cdot S_{\text{ent}}(i,j) + \gamma \cdot S_{\text{key}}(i,j) + \delta \cdot S_{\text{pos}}(i,j)$, where α, β, γ, and δ control the relative contributions of semantic similarity, entity overlap, keyword overlap, and positional proximity, respectively. In our implementation, all four coefficients are assigned equal weights ($\alpha = \beta = \gamma = \delta = 0.25$). The resulting graph captures semantic, entity-level, conceptual, and structural relationships among contractual clauses, enabling retrieval beyond direct semantic similarity.

3.3 Graph-Based Retrieval and Answer Generation

Given a user query q, MCG first computes a dense query embedding using the same sentence embedding model employed during graph construction. In addition, lightweight keyword and entity representations are derived from the query text.

To identify relevant entry points into the graph, each node is assigned an anchor relevance score based on its correspondence with the query. The anchor relevance score is computed as: $L(i) = w_1 \times S_q(i) + w_2 \times S_{key}(i) + w_3 \times S_{ent}(i)$, where $S_q(i)$ denotes the cosine similarity between the query embedding and the embedding of node i, while $S_{key}(i)$ and $S_{ent}(i)$ represent the keyword-overlap and entity-overlap scores between the query and node i, respectively. In our implementation, equal weights are assigned to all three components ($w_1 = w_2 = w_3 = 0.33$).

The highest-scoring nodes are selected as anchor nodes, representing the clauses most directly relevant to the user query. Rather than restricting retrieval to these anchor clauses alone, MCG expands retrieval by incorporating graph-connected clauses that exhibit strong relationships with the anchors. For each candidate node i connected to an anchor node a, a retrieval score is computed as: $R(i) = \lambda \times L(a) + (1 - \lambda) \times EW(a,i)$, where $L(a)$ denotes the relevance score of anchor node a, $EW(a,i)$ denotes the edge weight between anchor node a and candidate node i, and λ controls the relative influence of anchor relevance and graph connectivity. In our implementation, equal weighting is adopted ($\lambda = 0.5$). Candidate nodes are

subsequently ranked according to $R(i)$, enabling retrieval to consider both direct query relevance and the multi-view relationships captured within the contract graph. Consequently, clauses that may not be highly similar to the query can still be retrieved when they share strong semantic, entity-level, conceptual, or structural connections with highly relevant anchor clauses.

In our implementation, the two highest-scoring nodes are selected as anchor nodes. Neighboring graph-connected clauses are ranked according to their anchor relevance and edge-weight strength, and the final retrieval context consists of the five highest-ranked nodes obtained from the combined anchor-selection and graph-expansion process.

The retrieved nodes are transformed into evidence blocks containing both the original contract text and their associated metadata, including extracted named entities and contractual keywords. These evidence blocks are then provided to a downstream LLM together with an extraction-oriented prompt that instructs the model to generate answers using only the retrieved evidence. The final answer is therefore grounded in both the retrieved contractual clauses and the contextual relationships captured by the Multi-View Contract Graph, reducing unsupported generation while preserving supporting contractual context that may not be recovered through similarity-based retrieval alone.

4 Dataset and Experimentation Details

4.1 CUAD Dataset

We evaluate the proposed framework using contracts from the Contract Understanding Atticus Dataset (CUAD) (Hendrycks et al., 2021), a benchmark corpus containing over 500 commercial contracts annotated across 41 legal clause categories. The dataset includes diverse agreement types and contractual provisions, making it a representative benchmark for contract understanding and information extraction.

4.2 Task Formulation and Evaluation Setting

Our evaluation differs from the original CUAD benchmark, which formulates contract review as an extractive span-detection task and evaluates performance using metrics such as AUPR and Precision at fixed recall levels. In contrast, we reformulate contract analysis as a retrieval-augmented question-answering task for contract information extraction. Given a natural-language question corresponding to a contractual category, the system retrieves relevant contractual evidence and provides it to an LLM, which generates the final answer. From the original 41 CUAD categories, we evaluate 15 categories (Table 1) covering both straightforward contractual attributes (e.g., parties, dates, and governing law) and more complex provisions (e.g., exclusivity, non-compete, liability, and audit rights). Consequently, our evaluation measures the effectiveness of retrieval-augmented answer generation rather than supervised clause-span classification.

4.3 Baseline Methods

To evaluate the effectiveness of the proposed MCG, we compare against several commonly used retrieval and question-answering baselines: (1) **Extractive QA**: Pre-trained transformer-based question-answering models (Özkurt, 2024) (DistilBERT-QA, RoBERTa-QA, and BERT-QA) without CUAD-specific fine-tuning. This baseline evaluates conventional span-extraction approaches for contract information extraction and provides a retrieval-free benchmark for comparison with retrieval-based methods. (2) **BM25-RAG** (Robertson and Zaragoza, 2009): A retrieval-augmented generation framework that employs BM25 as the retriever. BM25 retrieves contract chunks based on lexical matching, and the retrieved context is supplied to an LLM for answer generation. (3) **Dense-Retrieval-RAG** (Lewis et al., 2020): A conventional retrieval-augmented generation framework that retrieves contract chunks using dense embedding similarity and supplies the retrieved context directly to an LLM for answer generation. (4) **ColBERT-RAG** (Khattab and Zaharia, 2020): A retrieval-augmented generation framework that employs ColBERT as the retriever. Retrieved contract chunks are ranked using late-interaction semantic matching and provided to an LLM for answer generation.

For answer generation, the BM25-RAG, ColBERT-RAG, Dense-Retrieval-RAG, and MCG retrieval frameworks were each evaluated using five open-source LLMs: Mistral-7B-Instruct-v0.2 (Jiang et al., 2023), Mistral-7B-Instruct-v0.3 (Jiang et al., 2023), Llama-3-8B-Instruct (Grattafiori et al., 2024), Saul-7B-Instruct-v1 (Colombo et al., 2024), and Qwen-7B-Chat (Bai et al., 2023). All methods were evaluated using the same set of contractual questions, retrieval settings, and evaluation protocols to ensure a fair comparison.

4.4 Evaluation Metrics

Since our task is formulated as retrieval-augmented answer generation rather than span detection, the evaluation metrics differ from those used in the original CUAD benchmark. For each contractual category, generated answers were compared against reference answers using Precision, Recall, and F_1 Score. To assess semantic similarity, we additionally report BERTScore-P, BERTScore-R, and BERTScore-F_1 (Zhang et al., 2019), which compare generated and reference answers using contextualized language model embeddings. Together, these metrics evaluate both lexical accuracy and semantic faithfulness.

5 Results and Discussion

5.1 Main Results

Table 2 presents the overall performance of Extractive QA, BM25-RAG, ColBERT-RAG, Dense-Retrieval-RAG, and the proposed MCG across five open-source LLMs. The best overall results were obtained by MCG combined with Mistral-7B-Instruct-v0.3, achieving a Precision of 0.458, Recall of 0.471, F_1 score of 0.460, and BERTScore-F_1 of 0.779. Compared with conventional Dense-Retrieval-RAG, MCG improves F_1 by around 17.9%, 12.7%, 13.8%, 16.0%, and 10.8% for Mistral-7B-Instruct-v0.2, Mistral-7B-Instruct-v0.3, Saul-7B-Instruct-v1, Qwen-7B-Chat, and Llama-3-8B-Instruct, respectively. Similar improvements are observed across BERTScore metrics, indicating that graph-enhanced retrieval improves both answer accuracy and semantic faithfulness.

Among the other retrieval-based baselines, BM25-RAG and ColBERT-RAG generally perform competitively and often slightly outperform Dense-Retrieval-RAG. However, MCG achieved the highest F_1 score for four of the five evaluated LLMs and attains the best overall F_1 score of 0.460 with Mistral-7B-Instruct-v0.3. For the same model, BM25-RAG and ColBERT-RAG achieve F_1 scores of 0.418 and 0.410, respectively, while MCG achieves 0.460, corresponding to relative improvements of approximately 10.0% and 12.2%. These results suggest that direct query-to-chunk similarity alone may be insufficient for contract understanding tasks. By incorporating semantic, entity, keyword, and structural relationships between clauses, MCG is able to recover supporting contractual context that is often overlooked by traditional retrieval methods.

Retrieval-based approaches also substantially outperform conventional extractive QA models. The strongest extractive baseline, BERT-QA, achieved an F_1 score of 0.261, whereas the best MCG configuration achieved 0.460, representing an absolute improvement of 19.9%. This result highlights the limitations of retrieval-free span extraction for contract information extraction, where relevant evidence is frequently distributed across multiple clauses and sections of a contract.

5.2 Ablation Study

To assess the contribution of each relationship signal, we perform a leave-one-component-out ablation study using the best-performing configuration (MCG + Mistral-7B-Instruct-v0.3). In each variant, one relationship component is removed from the edge-weight computation while keeping the retrieval procedure unchanged. The *No Semantic Edge*, *No Entity Edge*, *No Keyword Edge*, and *No Positional Edge* variants remove semantic similarity, entity overlap, keyword overlap, and positional proximity, respectively. Results are summarized in Table 3. Removing semantic relationships causes the largest performance drop, reducing F_1 from 0.460 to 0.409 ($\Delta F_1 = -0.051$), followed by entity relationships (0.438, $\Delta F_1 = -0.022$), keyword relationships (0.444, $\Delta F_1 = -0.016$), and positional relationships (0.451, $\Delta F_1 = -0.009$).

Overall, all four relationship signals contribute to retrieval effectiveness, with semantic similarity providing the strongest contribution. The highest performance is achieved when all four signals are combined, supporting the effectiveness of multi-view clause relationship modeling for contract retrieval and information extraction.

6 Limitations and Future Work

Our framework has three main limitations: (1) **Scalability**: MCG involves $O(n^2)$ pairwise relationship computation, which can be costly for large contracts. Future work will explore graph sparsification, approximate nearest-neighbor search, and hierarchical retrieval. (2) **Graph Construction**: MCG relies on manually designed semantic, entity, keyword, and positional relations and the quality of extracted entities and keywords. Future work will investigate learnable graph construction and adaptive edge weighting. (3) **Limited Evaluation**: MCG is evaluated on 15 CUAD categories, which may not capture the diversity of real-world legal

Table 2: Overall Performance Comparison on CUAD.

Method	Model	Precision	Recall	F_1	BERTScore-P	BERTScore-R	BERTScore-F_1
Extractive QA	DistilBERT-QA	0.284	0.187	0.227	0.641	0.582	0.613
Extractive QA	RoBERTa-QA	0.349	0.206	0.253	0.651	0.575	0.610
Extractive QA	BERT-QA	0.344	0.218	0.261	0.677	0.601	0.636
BM25-RAG	Mistral-7B-Instruct-v0.2	0.352	0.413	0.376	0.685	0.704	0.695
BM25-RAG	Mistral-7B-Instruct-v0.3	0.390	0.463	0.418	0.704	0.713	0.707
BM25-RAG	Saul-7B-Instruct-v1	0.355	0.437	0.376	0.694	0.732	0.709
BM25-RAG	Qwen-7B-Chat	0.281	0.523	0.362	0.664	0.727	0.692
BM25-RAG	Llama-3-8B-Instruct	0.262	0.553	0.351	0.613	0.706	0.657
ColBERT-RAG	Mistral-7B-Instruct-v0.2	0.370	0.443	0.397	0.694	0.710	0.702
ColBERT-RAG	Mistral-7B-Instruct-v0.3	0.372	0.469	0.410	0.711	0.723	0.716
ColBERT-RAG	Saul-7B-Instruct-v1	0.319	0.461	0.372	0.683	0.734	0.708
ColBERT-RAG	Qwen-7B-Chat	0.323	0.528	0.395	0.679	0.738	0.705
ColBERT-RAG	Llama-3-8B-Instruct	0.285	0.551	0.372	0.626	0.712	0.668
Dense-Retrieval-RAG	Mistral-7B-Instruct-v0.2	0.341	0.413	0.369	0.681	0.702	0.692
Dense-Retrieval-RAG	Mistral-7B-Instruct-v0.3	0.382	0.454	0.408	0.696	0.709	0.702
Dense-Retrieval-RAG	Saul-7B-Instruct-v1	0.345	0.420	0.377	0.686	0.726	0.707
Dense-Retrieval-RAG	Qwen-7B-Chat	0.284	0.499	0.357	0.663	0.725	0.690
Dense-Retrieval-RAG	Llama-3-8B-Instruct	0.247	0.529	0.334	0.601	0.702	0.648
Proposed Method (MCG)	Mistral-7B-Instruct-v0.2	0.424	0.455	0.435	0.763	0.767	0.765
Proposed Method (MCG)	Mistral-7B-Instruct-v0.3	0.458	0.471	0.460	0.791	0.773	0.779
Proposed Method (MCG)	Saul-7B-Instruct-v1	0.355	0.548	0.429	0.718	0.760	0.741
Proposed Method (MCG)	Qwen-7B-Chat	0.345	0.528	0.414	0.720	0.755	0.736
Proposed Method (MCG)	Llama-3-8B-Instruct	0.301	0.500	0.370	0.711	0.754	0.733

Table 3: Ablation Study of Graph Construction Components.

Variant	F_1	ΔF_1
MCG (Sem + Ent + Key + Pos)	**0.460**	–
No Semantic Edge	0.409	-0.051
No Entity Edge	0.438	-0.022
No Keyword Edge	0.444	-0.016
No Positional Edge	0.451	-0.009

documents. Future work will extend evaluation to larger and more diverse legal datasets.

7 Data and Code Availability

All datasets used in this work are publicly available[1]. The code is provided in a public repository[2].

8 Conclusion

This paper presented the Multi-View Contract Graph (MCG), a contract-specific knowledge graph and graph-based retrieval framework for contract question answering and information extraction. By modeling semantic, entity, keyword, and structural relationships among contract clauses, MCG extends retrieval beyond directly relevant clauses and recovers supporting contractual context that may be overlooked by conventional retrieval methods. Experimental results on the CUAD dataset across 15 contract information categories and five open-source LLMs demonstrate the effectiveness of the proposed approach. MCG achieved the strongest overall performance among the evaluated methods, attaining a best F_1 score of 0.460 and a BERTScore-F_1 of 0.779 with Mistral-7B-Instruct-v0.3. For the best-performing configuration, MCG improved F_1 by approximately 10.0% and 12.2% over BM25-RAG and ColBERT-RAG, respectively. Furthermore, the ablation study showed that all relationship signals contribute to retrieval effectiveness, with semantic relationships providing the largest contribution. Overall, the findings demonstrate that explicitly modeling inter-clause relationships improves evidence of retrieval and downstream contract information extraction. The proposed MCG framework provides a promising foundation for future legal AI systems supporting contract review, legal search, compliance analysis, and contract intelligence applications.

Generative AI Usage Declaration

The authors declare that no Generative AI or Large Language Models were used in the writing, drafting, or content generation of this manuscript.

References

Alan Akbik, Tanja Bergmann, Duncan Blythe, Kashif Rasul, Stefan Schweter, and Roland Vollgraf. 2019.

[1] https://www.atticusprojectai.org/cuad/
[2] https://github.com/mb16biswas/MCG/

Flair: An easy-to-use framework for state-of-the-art nlp. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics (demonstrations)*, pages 54–59.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, et al. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*.

Ryan Calvin Barron, Maksim Eren, Olga Serafimova, Cynthia Matuszek, and Boian Alexandrov. 2025. Bridging legal knowledge and ai: Retrieval-augmented generation with vector stores, knowledge graphs, and hierarchical non-negative matrix factorization. In *Proceedings of the Twentieth International Conference on Artificial Intelligence and Law*, pages 51–60.

Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. Legal-bert: The muppets straight out of law school. In *Findings of the association for computational linguistics: EMNLP 2020*, pages 2898–2904.

Yongming Chen, Miner Chen, Ye Zhu, Juan Pei, Siyu Chen, Yu Zhou, Yi Wang, Yifan Zhou, Hao Li, and Songan Zhang. 2026. Leverage knowledge graph and large language model for law article recommendation: A case study of chinese criminal law. *Journal of Web Semantics*, page 100879.

Pierre Colombo, Telmo Pessoa Pires, Malik Boudiaf, Dominic Culver, Rui Melo, Caio Corro, Andre FT Martins, Fabrizio Esposito, Vera Lúcia Raposo, Sofia Morgado, et al. 2024. Saullm-7b: A pioneering large language model for law. *arXiv preprint arXiv:2403.03883*.

Claudia dAmato, Giuseppe Rubini, Francesco Didio, Donato Francioso, Fatima Zahra Amara, and Nicola Fanizzi. 2025. Automated creation of the legal knowledge graph addressing legislation on violence against women: resource, methodology and lessons learned. *arXiv preprint arXiv:2508.06368*.

Erwin Filtz, Sabrina Kirrane, and Axel Polleres. 2021. The linked legal data landscape: linking legal data across different countries. *Artificial Intelligence and Law*, 29(4):485–539.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Maarten Grootendorst. 2020. Keybert: Minimal keyword extraction with bert.

Dan Hendrycks, Collin Burns, Anya Chen, and Spencer Ball. 2021. Cuad: An expert-annotated nlp dataset for legal contract review. *arXiv preprint arXiv:2103.06268*.

Sarika Jain, Pooja Harde, Nandana Mihindukulasooriya, Sudipto Gosh, Ankush Bisht, and Abhinav Dubey. 2022. Constructing a knowledge graph from indian legal domain corpus. In *TEXT2KG/MK@ ESWC*, pages 80–93.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *Preprint*, arXiv:2310.06825.

Mrs Junior, Barigye Brighton, and Avance University. 2026. Automated contract analysis and the reduction of legal operational costs.

Md Nazrul Islam Khan. 2022. A systematic review of legal technology adoption in contract management, data governance, and compliance monitoring. *American Journal of Interdisciplinary Studies*, 3(01):01–30.

Omar Khattab and Matei Zaharia. 2020. Colbert: Efficient and effective passage search via contextualized late interaction over bert. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, pages 39–48.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33:9459–9474.

Jun Li, Lu Qian, Peifeng Liu, and Taoxiong Liu. 2024. Construction of legal knowledge graph based on knowledge-enhanced large language models. *Information*, 15(11):666.

Shuang Liu, Zelong Li, Ruoyun Ma, Haiyan Zhao, and Mengnan Du. 2025. Contracteval: Benchmarking llms for clause-level legal risk identification in commercial contracts. *arXiv preprint arXiv:2508.03080*.

Cem Özkurt. 2024. Comparative analysis of state-of-the-art q&a models: Bert, roberta, distilbert, and albert on squad v2 dataset. *Chaos and Fractals*, 1(1):19–30.

Vuong T Pham, Dang V Dung, Hung Q Ngo, Binh Nguyen, Nhon V Do, Ali Selamat, and Hien D Nguyen. 2026. Ontology-based knowledge graph approach for legal queries. *Informatica*, 37(1):159–192.

Stephen Robertson and Hugo Zaragoza. 2009. The probabilistic relevance framework: Bm25 and beyond. *Foundations and trends® in information retrieval*, 4(1-2):1–174.

Julián Moreno Schneider, Georg Rehm, Elena Montiel-Ponsoda, Víctor Rodríguez-Doncel, Patricia Martín-Chozas, María Navas-Loro, Martin Kaltenböck, Artem Revenko, Sotirios Karampatakis, Christian Sageder, et al. 2022. Lynx: A knowledge-based ai service platform for content processing, enrichment and analysis for the legal domain. *Information Systems*, 106:101966.

Amrita Singh, Aditya Joshi, Jiaojiao Jiang, and Hye-young Paik. 2025. A survey of classification tasks and approaches for legal contracts. *Artificial Intelligence Review*, 58(12):380.

Amrita Singh, H Suhan Karaca, Aditya Joshi, Hye-young Paik, and Jiaojiao Jiang. 2026. Evaluating customized vs. generalist transformer-based models for legal contract classification. In *Proceedings of the Second Workshop on Customizable NLP: Progress and Challenges in Customizing NLP for a Domain, Application, Group, or Individual (CustomNLP4U)*, pages 44–54.

Francesco Sovrano, Monica Palmirani, and Fabio Vitali. 2020. Legal knowledge extraction for knowledge graph based question-answering. In *Legal Knowledge and Information Systems: JURIX 2020: The Thirty-third Annual Conference, Brno, Czech Republic, December 9–11, 2020*, pages 143–153. SAGE Publications Pvt. Ltd 1 Oliver's Yard, 55 City Road, London, EC1Y 1SP.

Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. 2022. Text embeddings by weakly-supervised contrastive pre-training. *arXiv preprint arXiv:2212.03533*.

Steven H Wang, Maksim Zubkov, Kexin Fan, Sarah Harrell, Yuyang Sun, Wei Chen, Andreas Plesner, and Roger Wattenhofer. 2025. Acord: An expert-annotated retrieval dataset for legal contract drafting. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 24739–24762.

Wenhui Wang, Hangbo Bao, Shaohan Huang, Li Dong, and Furu Wei. 2021. Minilmv2: Multi-head self-attention relation distillation for compressing pretrained transformers. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2140–2151.

Shitao Xiao, Zheng Liu, Peitian Zhang, Niklas Muennighoff, Defu Lian, and Jian-Yun Nie. 2024. C-pack: Packed resources for general chinese embeddings. In *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*, pages 641–649.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

Lucia Zheng, Neel Guha, Brandon R Anderson, Peter Henderson, and Daniel E Ho. 2021. When does pretraining help? assessing self-supervised learning for law and the casehold dataset of 53,000+ legal holdings. In *Proceedings of the eighteenth international conference on artificial intelligence and law*, pages 159–168.